Chapter 11

Distributions and the Fourier Transform

A distribution is a continuous linear functional on a space of test functions. Distributions provide a simple and elegant extension of functions that clarifies many aspects of analysis. For example, the delta function may be interpreted as a distribution. An advantage of distributions is that every distribution is differentiable, and differentiation is a continuous operation on spaces of distributions. Moreover, every tempered distribution has a Fourier transform, and a function whose Fourier transform is not defined as a function may nevertheless have a distributional transform. One limitation on the use of distributions is that there is no product of distributions that preserves the usual properties of the pointwise product of functions. Therefore, when studying nonlinear problems involving distributions, one must make sure that any products of distributions that appear are well defined.

11.1 The Schwartz space

In this section, we define a space of test functions on $\mathbb{R}^n$ called the *Schwartz space* that consists of smooth, rapidly decreasing functions.

We begin by introducing a concise notation for partial derivatives. Let

$$\mathbb{Z}_+ = \{n \in \mathbb{Z} \mid n \geq 0\}$$

denote the nonnegative integers. A *multi-index*

$$\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_+^n$$

is an n-tuple of nonnegative integers $\alpha_i \geq 0$. For multi-indices $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_1, \ldots, \beta_n)$, we define

$$\begin{aligned}
&|\alpha| = \sum_{i=1}^n \alpha_i, \qquad \alpha! = \prod_{i=1}^n \alpha_i!, \\
&\alpha + \beta = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n), \\
&\alpha \geq \beta \quad \text{if and only if } \alpha_i \geq \beta_i \text{ for } i = 1, \ldots, n.
\end{aligned}$$

If $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_+^n$, then we define

$$\partial^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n},$$
$$x^\alpha = \prod_{i=1}^n x_i^{\alpha_i},$$
$$|x| = \sqrt{x_1^2 + \ldots + x_n^2}.$$

We use the notation $x^\alpha f$ to denote the function whose value at x is $x^\alpha f(x)$.

The Taylor remainder theorem for $f \in C^k(\mathbb{R}^n)$ may be written as

$$f(x) = \sum_{|\alpha| \le k} \frac{1}{\alpha!} \partial^\alpha f(x_0)(x - x_0)^\alpha + r_k(x), \tag{11.1}$$

where the remainder term r_k satisfies

$$\lim_{x \to x_0} \frac{r_k(x)}{|x - x_0|^k} = 0.$$

The Leibnitz rule for the derivative of the product of $f, g \in C^k(\mathbb{R}^n)$ may be written as

$$\partial^\alpha (fg) = \sum_{\beta + \gamma = \alpha} \frac{\alpha!}{\beta!\gamma!} \left(\partial^\beta f\right) \left(\partial^\gamma g\right). \tag{11.2}$$

For multi-indices $\alpha, \beta \in \mathbb{Z}_+^n$, and $\varphi \in C^\infty(\mathbb{R}^n)$, we define

$$p_{\alpha,\beta}(\varphi) = \sup_{x \in \mathbb{R}^n} \left| x^\alpha \partial^\beta \varphi(x) \right|. \tag{11.3}$$

We also write $p_{\alpha,\beta}(\varphi)$ as $\|\varphi\|_{\alpha,\beta}$.

Definition 11.1 (Schwartz space) The *Schwartz space* $\mathcal{S}(\mathbb{R}^n)$, or $\mathcal{S}$ for short, consists of all functions $\varphi \in C^\infty(\mathbb{R}^n)$ such that $p_{\alpha,\beta}(\varphi)$ in (11.3) is finite for every pair of multi-indices $\alpha, \beta \in \mathbb{Z}_+^n$.

If $\varphi \in \mathcal{S}$, then for every $d \in \mathbb{N}$ and $\alpha \in \mathbb{Z}_+^n$ there is a constant $C_{d,\alpha}$ such that

$$|\partial^\alpha \varphi(x)| \le \frac{C_{d,\alpha}}{\left(1 + |x|^2\right)^{d/2}} \qquad \text{for all } x \in \mathbb{R}^n.$$

Thus, an element of $\mathcal{S}$ is a smooth function such that the function and all of its derivatives decay faster than any polynomial as $|x| \to \infty$. Elements of $\mathcal{S}$ are called *Schwartz functions*, or *test functions*. There are many functions in $\mathcal{S}$. For example, every function of the form

$$q(x) e^{-c|x - x_0|^2},$$

where $c > 0$, $x_0 \in \mathbb{R}^n$, and

$$q(x) = \sum_{|\alpha| \leq d} c_\alpha x^\alpha$$

is a polynomial function on $\mathbb{R}^n$, is a Schwartz function.

In order to define a notion of the convergence of test functions, we want to put a topology on $\mathcal{S}$. As we will see, the appropriate topology is not derived from a norm, but instead from the countable family $\{p_{\alpha,\beta}\}$ of seminorms. We therefore first discuss topologies defined by seminorms in more generality.

Definition 11.2 Suppose that X is a real or complex linear space. A function $p : X \to \mathbb{R}$ is a *seminorm* on X if it has the following properties:

(a) $p(x) \geq 0$ for all $x \in X$;
(b) $p(x + y) \leq p(x) + p(y)$ for all $x, y \in X$;
(c) $p(\lambda x) = |\lambda| p(x)$ for every $x \in X$ and $\lambda \in \mathbb{C}$.

A seminorm p has the same properties as a norm, except that $p(x) = 0$ need not imply $x = 0$. Suppose that $\{p_\alpha\}_{\alpha \in \mathcal{A}}$ is a countable or uncountable family of seminorms, indexed by a set $\mathcal{A}$, defined on a linear space X. Then X is a topological linear space with the following base $\mathcal{N}$ of open neighborhoods:

$$\mathcal{N} = \{N_{x;\alpha_1,\ldots,\alpha_n;\epsilon} \mid x \in X,\ \alpha_1 \ldots, \alpha_n \in \mathcal{A}, \text{ and } \epsilon > 0\},$$
$$N_{x;\alpha_1,\ldots,\alpha_n;\epsilon} = \{y \in X \mid p_{\alpha_i}(x - y) < \epsilon \text{ for } i = 1, \ldots, n\}.$$

A sequence (x_n) converges to $x \in X$ in this topology if and only if $p_\alpha(x - x_n) \to 0$ as $n \to \infty$ for each $\alpha \in \mathcal{A}$.

We say that a family $\{p_\alpha\}_{\alpha \in \mathcal{A}}$ of seminorms *separates points* if $p_\alpha(x) = 0$ for every $\alpha \in \mathcal{A}$ implies that $x = 0$. In that case, the associated topology is Hausdorff. A topological linear space whose topology may be derived from a family of seminorms that separates points is called a *locally convex space.*

If the family of seminorms $\{p_1, \ldots, p_n\}$ is finite and separates points, then

$$\|x\| = p_1(x) + \ldots + p_n(x)$$

defines a norm on X. Thus, there is no additional generality in using a finite family of seminorms instead of a norm. The main case of interest to us here is that of a locally convex space X whose topology is generated by a countably infinite family of seminorms $\{p_n \mid n \in \mathbb{N}\}$. In that case, the topology is metrizable because

$$d(x, y) = \sum_{n \in \mathbb{N}} \frac{1}{2^n} \frac{p_n(x - y)}{1 + p_n(x - y)} \tag{11.4}$$

defines a metric on X with the same collection of open sets as those generated by the family of seminorms (see Exercise 11.2). A metrizable, locally convex space that is complete as a metric space is called a *Fréchet space.*

The function $p_{\alpha,\beta}$ in (11.3) is a seminorm on $\mathcal{S}$. We equip $\mathcal{S}$ with the topology generated by the countable family of seminorms

$$\{p_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{Z}_+^n\}. \tag{11.5}$$

This family separates points, since $p_{0,0}$ is just the sup-norm. The following proposition shows that $\mathcal{S}$ is a Fréchet space.

Proposition 11.3 The Schwartz space $\mathcal{S}$ with the metrizable topology generated by the countable family of seminorms (11.5), where $p_{\alpha,\beta}$ is given by (11.3), is complete.

Proof. Let (φ_n) be a Cauchy sequence in $\mathcal{S}$. We have to prove that (φ_n) converges in the topology of $\mathcal{S}$ to a function $\varphi \in \mathcal{S}$. The sequence (φ_n) is Cauchy with respect to the sup-norm $p_{0,0}$. Since the space of bounded continuous functions on $\mathbb{R}$ with the supremum norm is complete, there is a bounded continuous function φ such that $\varphi_n \to \varphi$ uniformly. For each multi-index α, the sequence $\partial^\alpha \varphi_n$ is Cauchy with respect to the sup-norm, and hence converges uniformly to a bounded continuous function ψ_α. We claim that

$$\psi_\alpha = \partial^\alpha \varphi \qquad \text{for every multi-index } \alpha. \tag{11.6}$$

We prove (11.6) by induction on $|\alpha|$. The equation holds for $|\alpha| = 0$. Suppose we have proved (11.6) for every α with $|\alpha| \leq m$. Then, if $|\beta| = m + 1$, there exists an $\alpha \in \mathbb{Z}_+^n$ such that $|\alpha| = m$ and $\beta = \alpha + e_j$ for some j, where e_j is the jth standard basis vector of $\mathbb{Z}^n$. The fundamental theorem of calculus implies that

$$\partial^\alpha \varphi_n(x + te_j) - \partial^\alpha \varphi_n(x) = \int_0^t \partial^{e_j} \partial^\alpha \varphi_n(x + se_j)\, ds.$$

Clearly, $\partial^{e_j} \partial^\alpha = \partial^\beta$. Letting $n \to \infty$, we obtain that

$$\partial^\alpha \varphi(x + te_j) - \partial^\alpha \varphi(x) = \int_0^t \psi_\beta(x + se_j)\, ds.$$

We divide this expression by t and take the limit of the resulting expression as $t \to 0^+$. Using the definition of derivative and the continuity of ψ_β, we find that

$$\partial^\beta \varphi(x) = \psi_\beta(x).$$

Finally, for every pair of multi-indices (α, β), the sequence $(x^\alpha \partial^\beta \varphi_n)$ is Cauchy with respect to the uniform norm, so it converges uniformly. The uniform limit is equal to the pointwise limit $x^\alpha \partial^\beta \varphi$, so $p_{\alpha,\beta}(\varphi_n - \varphi) \to 0$ for all multi-indices, and therefore (φ_n) converges in $\mathcal{S}$. □

One main motivation for the use of this topology on $\mathcal{S}$ is that differentiation is a continuous operation.

Proposition 11.4 For each $\alpha \in \mathbb{Z}_+^n$, the partial differentiation operator $\partial^\alpha : \mathcal{S} \to \mathcal{S}$ is a continuous linear operator on $\mathcal{S}$.

Proof. The fact that ∂^α is a linear map of $\mathcal{S}$ into $\mathcal{S}$ is obvious. To prove the continuity, suppose that $\varphi_n \to \varphi$ in $\mathcal{S}$. Then $p_{\beta,\gamma}(\varphi_n - \varphi) \to 0$ as $n \to \infty$ for all $\beta, \gamma \in \mathbb{Z}_+^n$. Therefore,

$$p_{\beta,\gamma}(\partial^\alpha \varphi_n - \partial^\alpha \varphi) = p_{\beta,\alpha+\gamma}(\varphi_n - \varphi) \to 0$$

as $n \to \infty$ for all $\beta, \gamma \in \mathbb{Z}_+^n$, so $\partial^\alpha \varphi_n \to \partial^\alpha \varphi$ in $\mathcal{S}$. □

The Schwartz space is not the only possible space of test functions. Another useful choice is the smaller space $\mathcal{D} = C_c^\infty(\mathbb{R}^n)$ of smooth functions with compact support. The appropriate topology on $\mathcal{D}$ is, however, harder to describe than the topology on $\mathcal{S}$ because it is not metrizable.

11.2 Tempered distributions

The topological dual space of $\mathcal{S}$, denoted by $\mathcal{S}^*$ or $\mathcal{S}'$, is the space of continuous linear functionals $T : \mathcal{S} \to \mathbb{C}$. Elements of $\mathcal{S}^*$ are called *tempered distributions.* The space $\mathcal{S}^*$ is a linear space under the pointwise addition and scalar multiplication of functionals.

Since $\mathcal{S}$ is a metric space, a functional $T : \mathcal{S} \to \mathbb{C}$ is continuous if and only if for every convergent sequence $\varphi_n \to \varphi$ in $\mathcal{S}$, we have

$$\lim_{n\to\infty} T(\varphi_n) = T(\varphi).$$

The continuity of a linear functional T is implied by an estimate of the form

$$|T(\varphi)| \le \sum_{|\alpha|,|\beta|\le d} c_{\alpha,\beta} \|\varphi\|_{\alpha,\beta}$$

for some $d \in \mathbb{Z}_+$ and constants $c_{\alpha,\beta} \ge 0$. Conversely, one can show that if T is continuous, then such an estimate holds for some d and $c_{\alpha,\beta}$.

Example 11.5 The fundamental example of a distribution is the *delta function.* The name "delta function" is a misnomer because it is not a function on $\mathbb{R}^n$, but a functional on $\mathcal{S}$. We define $\delta : \mathcal{S} \to \mathbb{C}$ by evaluation at 0:

$$\delta(\varphi) = \varphi(0).$$

The linearity of δ is trivial. To show the continuity, suppose that $\varphi_n \to \varphi$ in $\mathcal{S}$. Then $\varphi_n \to \varphi$ uniformly, and therefore $\varphi_n(0) \to \varphi(0)$. Hence, $\delta(\varphi_n) \to \delta(\varphi)$, so $\delta \in \mathcal{S}^*$ is a continuous linear functional. Similarly, for each $x_0 \in \mathbb{R}^n$, we define the delta function supported at x_0 by evaluation at x_0:

$$\delta_{x_0}(\varphi) = \varphi(x_0).$$

Example 11.6 Suppose that f is a continuous, or Lebesgue measurable, function on $\mathbb{R}^n$ such that

$$|f(x)| \leq g(x)\left(1+|x|^2\right)^{d/2}$$

a.e. in $\mathbb{R}^n$ for a nonnegative integer $d \geq 0$ and a nonnegative, integrable function $g : \mathbb{R}^n \to \mathbb{R}$. Then

$$T_f(\varphi) = \int_{\mathbb{R}^n} f(x)\varphi(x)\,dx \tag{11.7}$$

defines a tempered distribution, as follows from the estimate:

$$\begin{aligned} |T_f(\varphi)| &\leq \int_{\mathbb{R}^n} g(x)\left(1+|x|^2\right)^{d/2}|\varphi(x)|\,dx \\ &\leq \left[\int_{\mathbb{R}^n} g(x)\,dx\right] \sup_{x\in\mathbb{R}^n}\left[\left(1+|x|^2\right)^{d/2}|\varphi(x)|\right]. \end{aligned}$$

Moreover, the function f is uniquely determined, up to pointwise-a.e. equivalence, by the distribution T_f. To see this, let $\{\varphi_\epsilon \mid \epsilon > 0\}$ be an approximate identity in $\mathcal{S}(\mathbb{R}^n)$, for example the Gaussian approximate identity,

$$\varphi_\epsilon(x) = \frac{1}{(2\pi\epsilon)^{n/2}}\exp\left(-\frac{|x|^2}{2\epsilon}\right).$$

Then for each $\epsilon > 0$ and $x \in \mathbb{R}^n$, the function $\varphi_{\epsilon,x}$ defined by

$$\varphi_{\epsilon,x}(y) = \varphi_\epsilon(x-y)$$

is an element of $\mathcal{S}(\mathbb{R}^n)$, and

$$T_f\left(\varphi_{\epsilon,x}\right) = \left(\varphi_\epsilon * f\right)(x).$$

Since we can recover f pointwise-a.e. from its convolutions with an approximate identity, we see that f is determined by T_f.

Distributions of the form (11.7) that are given by the integration of a test function φ against a function f are called *regular distributions*, and distributions, such as the delta function, that are not of this form are called *singular distributions*. Thus, we may regard tempered distributions as a generalization of locally integrable functions with polynomial growth.

A function that has a nonintegrable singularity, or a function that grows faster than a polynomial (such as $e^{c|x|^2}$ where $c > 0$), does not define a regular tempered distribution since its integral against a Schwartz function need not be finite.

Example 11.7 The function $(1/x) : \mathbb{R} \setminus 0 \to \mathbb{R}$ has a nonintegrable singularity at $x = 0$, so it does not define a regular distribution. We can, however, use a limiting procedure to define a singular distribution called a *principal value distribution*, denoted by p.v.$(1/x)$. We define its action on a test function $\varphi \in \mathcal{S}(\mathbb{R})$ by

$$\text{p.v.}\frac{1}{x}(\varphi) = \lim_{\epsilon \to 0^+} \int_{|x|>\epsilon} \frac{\varphi(x)}{x}\, dx.$$

The limit is finite because of a cancellation between the nonintegrable contributions of $1/x$ for $x < 0$ and $x > 0$:

$$\text{p.v.}\frac{1}{x}(\varphi) = \lim_{\epsilon \to 0^+} \int_\epsilon^\infty \frac{\varphi(x) - \varphi(-x)}{x}\, dx = \int_0^\infty \frac{\varphi(x) - \varphi(-x)}{x}\, dx.$$

The integrand is bounded at $x = 0$ since φ is smooth. For $x > 0$, we have

$$\left|\frac{\varphi(x) - \varphi(-x)}{x}\right| \leq \frac{1}{x}\int_{-x}^{x} |\varphi'(t)|\; dt \leq 2\|\varphi'\|_\infty,$$

so the continuity of p.v.$(1/x)$ on $\mathcal{S}$ follows from the estimate

$$\begin{aligned} \left|\text{p.v.}\frac{1}{x}(\varphi)\right| &\leq \int_0^1 \left|\frac{\varphi(x) - \varphi(-x)}{x}\right| dx + \int_1^\infty \left|\frac{x\left[\varphi(x) - \varphi(-x)\right]}{x^2}\right| dx \\ &\leq 2\|\varphi'\|_\infty + 2\|x\varphi\|_\infty \\ &= 2\left(\|\varphi\|_{0,1} + \|\varphi\|_{1,0}\right). \end{aligned}$$

Example 11.8 The function $1/|x|^2 : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ has an integrable singularity at the origin when $n \geq 3$ since the radial integral

$$\int_0^1 r^{-2}\, r^{n-1} dr$$

is finite. In that case

$$\frac{1}{|x|^2}(\varphi) = \int_{\mathbb{R}^n} \frac{\varphi(x)}{|x|^2}\, dx$$

defines a regular distribution in $\mathcal{S}^*(\mathbb{R}^n)$. If $n = 2$, the function is not integrable, but we can define an associated singular distribution, called a *finite part distribution*, denoted by f.p.$(1/|x|^2)$:

$$\text{f.p.}\frac{1}{|x|^2}(\varphi) = \int_{|x|<1} \frac{\varphi(x) - \varphi(0)}{|x|^2}\, dx + \int_{|x|>1} \frac{\varphi(x)}{|x|^2}\, dx.$$

The action of the elements of the dual space $\mathcal{S}^*$ on $\mathcal{S}$ may be represented by a *duality pairing*, which resembles an inner product:

$$\langle \cdot, \cdot \rangle : \mathcal{S}^* \times \mathcal{S} \to \mathbb{C}.$$

We write the action of a distribution T on a test function φ as

$$T(\varphi) = \langle T, \varphi \rangle.$$

If $\mathcal{H}$ is a Hilbert space, then the duality pairing on $\mathcal{H}^* \times \mathcal{H}$ can be identified with the inner product on $\mathcal{H}$ by the Riesz representation theorem. Note, however, that in the case of an inner product on a Hilbert space, the duality pairing is antilinear in one of the variables, whereas the duality pairing on $\mathcal{S}^* \times \mathcal{S}$ is linear in both variables.

Another notation for the action of $T \in \mathcal{S}^*$ on $\varphi \in \mathcal{S}$ is

$$T(\varphi) = \int T(x)\varphi(x)\, dx,$$

as if $\mathcal{S}^*$ were a function space. If T_f is the regular distribution defined in (11.7), then this notation amounts to the identification of T_f with f. The action of the distribution δ_{x_0} is then written as

$$\delta_{x_0}(\varphi) = \int \delta(x - x_0)\varphi(x)\, dx = \varphi(x_0).$$

Since the pairing on $\mathcal{S}^* \times \mathcal{S}$ shares a number of properties with inner products defined through an integral, this notation is often convenient in computations, provided one remembers that it is just a way to write continuous linear functionals.

The tempered distributions are a subspace of the space $\mathcal{D}^*$ of *distributions* that are continuous linear functionals on the space $\mathcal{D}$ of smooth, compactly supported test functions. Unlike tempered distributions, distributions in $\mathcal{D}^*$ can grow faster than any polynomial at infinity. The Fourier transform of a distribution in $\mathcal{D}^*$ need not belong to $\mathcal{D}^*$, however, whereas we will see that every distribution in $\mathcal{S}^*$ has a Fourier transform that is also in $\mathcal{S}^*$. To be specific, we therefore restrict our discussion to tempered distributions, although similar ideas apply to distributions defined on other spaces of test functions.

11.3 Operations on distributions

We say that a continuous function $f : \mathbb{R}^n \to \mathbb{C}$ is of *polynomial growth* if there is an integer d and a constant C such that

$$|f(x)| \le C \left(1 + |x|^2\right)^{d/2} \qquad \text{for all } x \in \mathbb{R}^n.$$

If $T \in \mathcal{S}^*$ and $f \in C^\infty(\mathbb{R}^n)$ is such that f and $\partial^\alpha f$ have polynomial growth for every $\alpha \in \mathbb{Z}^n_+$, then we define the product $fT \in S^*$ by

$$\langle fT, \varphi\rangle = \langle T, f\varphi\rangle \qquad \text{for all } \varphi \in \mathcal{S}.$$

This definition makes sense because $f\varphi \in \mathcal{S}$ when $\varphi \in \mathcal{S}$. It is straightforward to check that fT is a continuous linear map on $\mathcal{S}$ if T is.

Example 11.9 If $T = \delta$ is the delta function, then

$$\langle f\delta, \varphi\rangle = \langle \delta, f\varphi\rangle = f(0)\varphi(0) = \langle f(0)\delta, \varphi\rangle.$$

Hence, $f\delta = f(0)\delta$.

The definition of products may be extended further; for example, the product $f\delta = f(0)\delta$ makes sense for any continuous function f. It is not possible, however, to define a product $ST \in S^*$ for general distributions $S, T \in S^*$ with the same algebraic properties as the pointwise product of functions (see Exercise 11.7).

Next, we define the derivative of a distribution. To motivate the definition, we first consider the regular distribution T_f associated with a Schwartz function f. Integrating by parts, we find that the action of the regular distribution $T_{\partial^\alpha f}$, associated with the αth partial derivative of f, on a test function φ is given by

$$\langle T_{\partial^\alpha f}, \varphi\rangle = \int (\partial^\alpha f)\, \varphi\, dx = (-1)^{|\alpha|} \int f\, (\partial^\alpha \varphi)\, dx = (-1)^{|\alpha|}\langle T_f, \partial^\alpha \varphi\rangle.$$

The following definition extends the differentiation of functions to the differentiation of distributions.

Definition 11.10 Suppose that T is a tempered distribution and α is a multi-index. The αth *distributional derivative* of T is the tempered distribution $\partial^\alpha T$ defined by

$$\langle \partial^\alpha T, \varphi\rangle = (-1)^{|\alpha|}\langle T, \partial^\alpha \varphi\rangle \qquad \text{for all } \varphi \in \mathcal{S}. \tag{11.8}$$

Equation (11.8) does define a distribution. The linearity of the map $\partial^\alpha T : \mathcal{S} \to \mathbb{C}$ is obvious. The continuity of $\partial^\alpha T$ follows from the continuity of T and ∂^α on $\mathcal{S}$. If $\varphi_n \to \varphi$ in $\mathcal{S}$, then $\partial^\alpha \varphi_n \to \partial^\alpha \varphi$ in $\mathcal{S}$, so

$$\langle \partial^\alpha T, \varphi_n\rangle = (-1)^{|\alpha|}\langle T, \partial^\alpha \varphi_n\rangle \to (-1)^{|\alpha|}\langle T, \partial^\alpha \varphi\rangle = \langle \partial^\alpha T, \varphi\rangle.$$

Thus, every tempered distribution is differentiable. The space of distributions is therefore an extension of the space of functions that is closed under differentiation. The following structure theorem, whose proof we omit, shows that $\mathcal{S}$ is a minimal extension of the space of functions of polynomial growth that is closed under differentiation.

Theorem 11.11 For every $T \in \mathcal{S}^*$ there is a continuous function $f : \mathbb{R}^n \to \mathbb{C}$ of polynomial growth and a multi-index $\alpha \in \mathbb{Z}_+^n$ such that $T = \partial^\alpha f$.

If T_f is a regular distribution whose distributional derivative is also a regular distribution T_g, then

$$\int_{\mathbb{R}^n} g\varphi\, dx = (-1)^{|\alpha|} \int_{\mathbb{R}^n} f \partial^\alpha \varphi\, dx \qquad \text{for all } \varphi \in \mathcal{S}.$$

In this case, we say that the function g is the *weak* or *distributional* derivative of the function f, and we write $g = \partial^\alpha f$. Thus, the weak L^2-derivatives considered in Section 10.4 were a special case of the distributional derivative. If f does not have a weak derivative g, then the distributional derivative of T_f still exists, but it is a singular distribution not associated with a function.

Example 11.12 Let $f : \mathbb{R} \to \mathbb{R}$ be the function

$$f(x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } x \geq 0. \end{cases}$$

Then f is Lipschitz continuous on $\mathbb{R}$, but it is not differentiable pointwise at $x = 0$, where its graph has a corner. Integrating by parts, and using the rapid decrease of a test function, we find that the action of the distributional derivative of f on a test function φ is given by

$$\langle f', \varphi \rangle = -\langle f, \varphi' \rangle = -\int_0^\infty x\varphi' \, dx = \int_0^\infty \varphi \, dx = \langle H, \varphi \rangle,$$

where H is the step function,

$$H(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x > 0. \end{cases}$$

Thus, f is weakly differentiable, and its weak derivative is the step function H.

Example 11.13 The distributional derivative of the step function is given by

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

Hence, the step function is not weakly differentiable. Its distributional derivative is the delta function, as stated in (10.12).

Example 11.14 The derivative of the one-dimensional delta function δ is given by

$$\langle \delta', \varphi \rangle = -\langle \delta, \varphi' \rangle = -\varphi'(0).$$

More generally, the kth distributional derivative of δ is given by

$$\langle \delta^{(k)}, \varphi \rangle = (-1)^k \varphi^{(k)}(0).$$

Example 11.15 The pointwise derivative of the Cantor function F, defined in Exercise 1.19, exists a.e. and is equal to zero except on the Cantor set. The function is not constant, however, and its distributional derivative is not zero. One can show that the distributional derivative of F is the Lebesgue-Stieltjes measure μ_F associated with the Cantor function, described in Example 12.15, meaning that

$$\langle F', \varphi \rangle = \int_{-\infty}^\infty \varphi(x) \, d\mu_F(x).$$

The use of duality to extend differentiation from test functions to distributions may be applied to other operations. Suppose that $K, K' : \mathcal{S} \to \mathcal{S}$ are continuous linear transformations on $\mathcal{S}$ such that

$$\int_{\mathbb{R}^n} (Kf)\varphi \, dx = \int_{\mathbb{R}^n} f(K'\varphi) \, dx \qquad \text{for all } f, \varphi \in \mathcal{S}. \tag{11.9}$$

We call K' the transpose of K. The transpose K' differs from the L^2-Hilbert space adjoint K^* of K because, unlike the L^2-inner product, we do not use a complex-conjugate in the duality pairing. If T is a tempered distribution, then we define the tempered distribution KT by

$$\langle KT, \varphi \rangle = \langle T, K'\varphi \rangle \qquad \text{for all } \varphi \in \mathcal{S}.$$

If T_f is the regular distribution associated with a test function $f \in \mathcal{S}$, then we have $KT_f = T_{Kf}$, so this definition is consistent with the definition for test functions.

Example 11.16 For each $h \in \mathbb{R}^n$, we define the *translation operator* $\tau_h : \mathcal{S} \to \mathcal{S}$ by

$$\tau_h f(x) = f(x - h).$$

We then have that

$$\int_{\mathbb{R}^n} (\tau_h f)\varphi \, dx = \int_{\mathbb{R}^n} f(\tau_{-h}\varphi) \, dx \qquad \text{for all } f, \varphi \in \mathcal{S}.$$

We therefore define the translation $\tau_h T$ of a distribution T by

$$\langle \tau_h T, \varphi \rangle = \langle T, \tau_{-h}\varphi \rangle \qquad \text{for all } \varphi \in \mathcal{S}.$$

For instance, we have $\delta_{x_0} = \tau_{x_0}\delta$.

Example 11.17 Let $R : \mathcal{S} \to \mathcal{S}$ be the *reflection operator*,

$$Rf(x) = f(-x).$$

Then

$$\int_{\mathbb{R}^n} (Rf)\varphi \, dx = \int_{\mathbb{R}^n} f(R\varphi) \, dx \qquad \text{for all } f, \varphi \in \mathcal{S}.$$

Thus, for $T \in \mathcal{S}^*$, we define the reflection $RT \in \mathcal{S}^*$ by

$$\langle RT, \varphi \rangle = \langle T, R\varphi \rangle \qquad \text{for all } \varphi \in \mathcal{S}.$$

We end this section by defining the convolution of a test function and a distribution. The convolution $\varphi * \psi$ of two test functions $\varphi, \psi \in \mathcal{S}$ is defined by

$$(\varphi * \psi)(x) = \int_{\mathbb{R}^n} \varphi(x - y)\psi(y) \, dy. \tag{11.10}$$

The following properties of the convolution are straightforward to prove.

Proposition 11.18 For any $\varphi, \psi, \omega \in \mathcal{S}$, we have:

(a) $\varphi * \psi = \psi * \varphi$;
(b) $(\varphi * \psi) * \omega = \varphi * (\psi * \omega)$,
(c) $\tau_h(\varphi * \psi) = (\tau_h\varphi) * \psi = \varphi * (\tau_h\psi)$ for every $h \in \mathbb{R}^n$.

It is clear from (11.10) that the definition of convolution can be extended from test functions to more general functions provided that the integral converges. For example, the convolution of a continuous function with compact support and an arbitrary continuous function exists, and the convolution of two L^1-functions exists and belongs to L^1. On the other hand, the convolution of two functions neither of which decays at infinity need not be well defined.

Using the translation and reflection operators defined in Example 11.16 and Example 11.17, we may write the convolution in (11.10) as

$$(\varphi * \psi)(x) = \int_{\mathbb{R}^n} (R\tau_x\varphi)(y)\psi(y)\,dy.$$

We therefore define the convolution $\varphi * T : \mathbb{R}^n \to \mathbb{C}$ of a test function $\varphi \in \mathcal{S}$ and a tempered distribution $T \in \mathcal{S}^*$ by

$$(\varphi * T)(x) = \langle T, R\tau_x\varphi\rangle.$$

One can prove that $\varphi * T \in C^\infty(\mathbb{R}^n)$, and is of at most polynomial growth.

Example 11.19 The convolution of a test function with the delta function is given by

$$(\varphi * \delta)(x) = \langle \delta, R\tau_x\varphi\rangle = (R\tau_x\varphi)(0) = (R\varphi)(-x) = \varphi(x),$$

meaning that $\varphi * \delta = \varphi$. This fact provides a distributional interpretation of the formula

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

Similarly, the convolution with a derivative of the delta function is

$$(\varphi * \partial^\alpha\delta)(x) = (-1)^{|\alpha|}\langle \delta, \partial^\alpha R\tau_x\varphi\rangle = \partial^\alpha\varphi(x).$$

More general convolutions of distributions may also be defined (for example, $\partial^\alpha\delta * T = \partial^\alpha T$ for any $T \in \mathcal{S}^*$), but we will not give a detailed description here.

11.4 The convergence of distributions

Let (T_n) be a sequence in $\mathcal{S}^*$. We say that (T_n) converges to T in $\mathcal{S}^*$ if and only if

$$\lim_{n\to\infty}\langle T_n, \varphi\rangle = \langle T, \varphi\rangle \qquad \text{for every } \varphi \in \mathcal{S}. \tag{11.11}$$

We denote convergence in the space of distributions by

$$T_n \rightharpoonup T \qquad \text{as } n \to \infty.$$

Example 11.20 Let T_n be the distribution in $\mathcal{S}(\mathbb{R})$ defined by

$$\langle T_n, \varphi \rangle = n^3 \int_{\mathbb{R}} e^{inx} \varphi(x)\, dx.$$

Integrating by parts four times, and using the rapid decrease of $\varphi \in \mathcal{S}$, we find that

$$|\langle T_n, \varphi \rangle| = \left| \int_{\mathbb{R}} \frac{e^{inx}}{n} \varphi^{(4)}(x)\, dx \right| \to 0 \qquad \text{as } n \to \infty.$$

Thus, we have $T_n \rightharpoonup 0$ in $\mathcal{S}^*(\mathbb{R})$. The cancellation of oscillations for large n in the integration of $n^3 e^{inx}$ against a smooth test function outweighs the polynomial growth in n.

For each $\varphi \in \mathcal{S}$, the map

$$T \mapsto \langle T, \varphi \rangle \tag{11.12}$$

is a linear functional on $\mathcal{S}^*$. The convergence of distributions defined in (11.11) corresponds to convergence with respect to the weakest topology such that every functional of the form (11.12) is continuous. This topology, called the *weak-$*$ topology* of $\mathcal{S}^*$, is the locally convex topology generated by the uncountable family of seminorms $\{p_\varphi \mid \varphi \in \mathcal{S}\}$, where

$$p_\varphi(T) = |\langle T, \varphi \rangle| \qquad \text{for } T \in \mathcal{S}^*. \tag{11.13}$$

Sequences of distributions that converge to the delta function are particularly important. Such sequences are called *delta sequences*. We have already encountered several examples of delta sequences, without thinking of them in terms of distributions.

Example 11.21 A simple delta sequence in $\mathcal{S}(\mathbb{R})$ is given by

$$T_n(\varphi) = \frac{n}{2} \int_{-1/n}^{1/n} \varphi(x)\, dx.$$

For any continuous function φ, we have

$$T_n(\varphi) \to \varphi(0) = \delta(\varphi) \qquad \text{as } n \to \infty,$$

so $T_n \rightharpoonup \delta$. Any approximate identity gives a delta sequence; for example, the Gaussian approximate identity

$$\varphi_n(x) = \sqrt{\frac{n}{2\pi}} e^{-nx^2/2} \tag{11.14}$$

is a delta sequence that consists of elements of $\mathcal{S}(\mathbb{R})$.

The following proposition gives a useful delta sequence of oscillatory functions. We define the *sinc function* by

$$\operatorname{sinc} x = \begin{cases} \sin x/x & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases}$$

The integral of the absolute value of the sinc function does not converge, since it decays like $1/x$ as $|x| \to \infty$, but a contour integral argument gives the following improper Riemann integral

$$\lim_{R\to\infty} \int_{-R}^{R} \operatorname{sinc} x \, dx = \pi. \tag{11.15}$$

Proposition 11.22 For $n \in \mathbb{N}$, let

$$\sigma_n(x) = \frac{\sin nx}{\pi x}. \tag{11.16}$$

Then $\sigma_n \rightharpoonup \delta$ in $\mathcal{S}^*(\mathbb{R})$ as $n \to \infty$.

Proof. From (11.15), we see that

$$\sigma_n(x) = \frac{n}{\pi} \operatorname{sinc} nx$$

has unit integral for every $n \in \mathbb{N}$. To avoid difficulties caused by the lack of absolute convergence of the integral of σ_n at infinity, we split the integral of σ_n against a test function $\varphi \in \mathcal{S}$ into two terms:

$$\int_{-\infty}^{\infty} \frac{\sin nx}{\pi x} \varphi(x)\, dx = \int_{|x|\geq 1} \frac{\sin nx}{\pi x} \varphi(x)\, dx + \int_{|x|\leq 1} \frac{\sin nx}{\pi x} \varphi(x)\, dx. \tag{11.17}$$

An integration by parts implies that the first integral on the right-hand side tends to zero as $n \to \infty$, since

$$\int_{|x|\geq 1} \frac{\sin nx}{\pi x} \varphi(x)\, dx = \frac{1}{n} \left[\cos nx \frac{\varphi(x)}{x} \right]_{-1}^{1} + \frac{1}{n} \int_{|x|\geq 1} \cos nx \left(\frac{\varphi(x)}{x} \right)' dx.$$

We write the second term on the right-hand side of (11.17) as

$$\int_{|x|\leq 1} \frac{\sin nx}{\pi x} \varphi(x)\, dx = \int_{|x|\leq 1} \frac{\sin nx}{\pi x} \left[\varphi(x) - \varphi(0) \right] dx + \varphi(0) \int_{|x|\leq 1} \frac{\sin nx}{\pi x}\, dx.$$

We may write $\varphi(x) = \varphi(0) + x\psi(x)$ where $\psi \in C^\infty$. The first integral on the right-hand side is therefore given by

$$\frac{1}{\pi} \int_{|x|\leq 1} \sin nx \, \psi(x)\, dx,$$

and an integration by parts shows this approaches zero as $n \to \infty$. Making the change of variable $nx \mapsto x$ and using (11.15), we see that the second term approaches

$\varphi(0)$ as $n \to \infty$, which proves the result. Note that the proof shows that $\sigma_n * \varphi \to \varphi$ uniformly for every $\varphi \in \mathcal{S}$. □

The identification $\varphi \mapsto T_\varphi$ continuously embeds the Schwartz space $\mathcal{S}$ into the space $\mathcal{S}^*$ of tempered distributions. This embedding is clearly not onto, but the next result, whose proof we only outline, states that $\mathcal{S}$ is dense in $\mathcal{S}^*$.

Theorem 11.23 The Schwartz space is dense in the space of tempered distributions.

Proof. Let (φ_n) be an approximate identity in $\mathcal{S}$. Then $(\varphi_n * T)$ is a sequence of C^∞-functions of polynomial growth that converges to T in $\mathcal{S}^*$. The Schwartz functions $(\varphi_n * T)e^{-\epsilon|x|^2}$ therefore converge to T in $\mathcal{S}^*$ as $n \to \infty$ and $\epsilon \to 0^+$. □

A similar proof shows that $\mathcal{S}$ is dense in C_0, with respect to uniform convergence, and in L^p for $1 \le p < \infty$.

11.5 The Fourier transform of test functions

In this section, we define the Fourier transform of a Schwartz function, and show that the Fourier transform is a continuous, one-to-one map from $\mathcal{S}$ onto $\mathcal{S}$. In the next section, we will extend the transform by duality to a continuous, one-to-one map from $\mathcal{S}^*$ onto $\mathcal{S}^*$.

Definition 11.24 If $\varphi \in \mathcal{S}(\mathbb{R}^n)$, then the *Fourier transform* $\hat{\varphi} : \mathbb{R}^n \to \mathbb{C}$ is the function defined by

$$\hat{\varphi}(k) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \varphi(x) e^{-ik\cdot x}\, dx \qquad \text{for } k \in \mathbb{R}^n. \tag{11.18}$$

There are many different conventions for where to place the factors of 2π and the signs in the Fourier transform. In the next proposition, we show that the transform of a rapidly decaying function is smooth, and the transform of a smooth function is rapidly decaying. As a result, the Fourier transform maps the Schwartz space into itself. We define the Fourier transform operator $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ by $\mathcal{F}\varphi = \hat{\varphi}$.

Proposition 11.25 If $\varphi \in \mathcal{S}(\mathbb{R}^n)$, then:

(a) $\hat{\varphi} \in C^\infty(\mathbb{R}^n)$, and

$$\partial^\alpha \hat{\varphi} = \mathcal{F}\left[(-ix)^\alpha \varphi\right]; \tag{11.19}$$

(b) $k^\alpha \hat{\varphi}$ is bounded for every multi-index $\alpha \in \mathbb{Z}_+^n$, and

$$(ik)^\alpha \hat{\varphi} = \mathcal{F}\left[\partial^\alpha \varphi\right]. \tag{11.20}$$

The Fourier transform $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ is a continuous linear map on $\mathcal{S}(\mathbb{R}^n)$.

Proof. Equation (11.19) follows by differentiation under the integral sign in (11.18). This differentiation is justified by the dominated convergence theorem and the integrability of $x^\alpha \varphi$ for every $\alpha \in \mathbb{Z}_+^n$. Equation (11.20) follows from an integration by parts in the formula

$$\begin{aligned}
\widehat{\partial^\alpha \varphi}(k) &= \frac{1}{(2\pi)^{n/2}} \int e^{-ik\cdot x} \partial^\alpha \varphi(x)\, dx \\
&= \frac{1}{(2\pi)^{n/2}} \int (ik)^\alpha e^{-ik\cdot x} \varphi(x)\, dx \\
&= (ik)^\alpha \hat{\varphi}(k).
\end{aligned}$$

Thus, for every $\alpha, \beta \in \mathbb{Z}_+^n$, we have

$$(ik)^\alpha \partial^\beta \hat{\varphi} = \mathcal{F}\left[\partial^\alpha (-ix)^\beta \varphi\right]. \tag{11.21}$$

If $\varphi \in \mathcal{S}$, then

$$\begin{aligned}
|\hat{\varphi}(k)| &= \frac{1}{(2\pi)^{n/2}} \left| \int e^{-ik\cdot x} \varphi(x)\, dx \right| \\
&\leq \frac{1}{(2\pi)^{n/2}} \int \frac{\left(1 + |x|^2\right)^{n/2+1} |\varphi(x)|}{(1 + |x|^2)^{n/2+1}}\, dx \\
&\leq C \sup_{x\in\mathbb{R}^n} \left[\left(1 + |x|^2\right)^{n/2+1} |\varphi(x)|\right],
\end{aligned}$$

where the constant C is given by

$$C = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \frac{1}{(1 + |x|^2)^{n/2+1}}\, dx < \infty.$$

Taking the supremum of (11.21) with respect to k, using the Leibnitz rule to expand the function on the right-hand side, and estimating the result, we find for the seminorms in (11.3) that

$$\|\hat{\varphi}\|_{\alpha,\beta} \leq \sum_{\alpha',\beta'} C_{\alpha',\beta'} \|\varphi\|_{\beta',\alpha'}$$

for suitable constants $C_{\alpha',\beta'}$, where $|\alpha'| \leq |\alpha|$ and $|\beta'| \leq |\beta| + n + 2$. Hence, the Fourier transform is a continuous linear map on $\mathcal{S}$. □

An important example of the Fourier transform of a Schwartz function is the transform of a Gaussian, which is another Gaussian.

Proposition 11.26 Suppose that A is an $n \times n$ symmetric, positive definite matrix. The Fourier transform of the n-dimensional Gaussian

$$\varphi(x) = \exp\left(-\frac{1}{2} x \cdot Ax\right) \tag{11.22}$$

is given by

$$\hat{\varphi}(k) = \frac{1}{\sqrt{\det A}} \exp\left(-\frac{1}{2} k \cdot A^{-1} k\right). \tag{11.23}$$

Proof. First, we consider the one-dimensional Gaussian

$$\varphi(x) = \exp\left(-\frac{ax^2}{2}\right),$$

where $a > 0$. We claim that

$$\hat{\varphi}(k) = \frac{1}{\sqrt{a}} \exp\left(-\frac{k^2}{2a}\right). \tag{11.24}$$

To prove this result, it suffices to consider the case $a = 1$. The formula for $a > 0$ then follows from the change of variables $x \mapsto \sqrt{a}x$. Thus, we just need to show that

$$\frac{1}{\sqrt{2\pi}} \int e^{-x^2/2} e^{-ikx}\, dx = e^{-k^2/2}.$$

The left-hand side of this equation may be written as

$$\frac{1}{\sqrt{2\pi}} e^{-k^2/2} \int e^{-(x+ik)^2/2}\, dx.$$

So we want to show that

$$\frac{1}{\sqrt{2\pi}} \int e^{-(x+ik)^2/2}\, dx = 1. \tag{11.25}$$

This integral is independent of k, since

$$\begin{aligned} \frac{d}{dk}\left(\frac{1}{\sqrt{2\pi}} \int e^{-(x+ik)^2/2}\, dx\right) &= -i\frac{1}{\sqrt{2\pi}} \int (x+ik) e^{-(x+ik)^2/2}\, dx \\ &= i\frac{1}{\sqrt{2\pi}} \int \frac{d}{dx} e^{-(x+ik)^2/2}\, dx \\ &= i\frac{1}{\sqrt{2\pi}} e^{-(x+ik)^2/2}\Big|_{-\infty}^{\infty} \\ &= 0, \end{aligned}$$

so (11.25) follows from the standard Gaussian integral,

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}.$$

Now we consider the n-dimensional case. The Fourier tranform of the Gaussian in (11.22) is given by

$$\hat{\varphi}(k) = \frac{1}{(2\pi)^{n/2}} \int e^{-x \cdot Ax/2} e^{-ik \cdot x}\, dx. \tag{11.26}$$

Since A is positive definite, there is an orthogonal matrix Q such that $Q^T AQ = \Lambda$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix with the eigenvalues $\lambda_i > 0$ of A on the main diagonal. We make the change of variables $x = Q\overline{x}$ and $k = Q\overline{k}$ in (11.26). The Jacobian of the transformation $x \mapsto \overline{x}$ is $\det Q = 1$. The resulting expression factors into a product of one-dimensional Fourier integrals, which we may evaluate using the one-dimensional computation:

$$\begin{aligned}
\hat{\varphi}(k) &= \frac{1}{(2\pi)^{n/2}} \int e^{-\overline{x}\cdot\Lambda\overline{x}/2} e^{-i\overline{k}\cdot\overline{x}}\, d\overline{x} \\
&= \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}} \int e^{-\lambda_j \overline{x}_j^2/2} e^{-i\overline{k}_j\overline{x}_j}\, d\overline{x}_j \\
&= \prod_{j=1}^{n} \frac{1}{\sqrt{\lambda_j}} e^{-\overline{k}_j^2/(2\lambda_j)}.
\end{aligned}$$

Rewriting this result in terms of k, and using the facts that $\det A = (\lambda_1\lambda_2 \ldots \lambda_n)$ and $A^{-1} = Q^T \Lambda^{-1} Q$, we obtain (11.23). □

The covariance matrix A of the transform of a Gaussian is the inverse of the covariance matrix of the Gaussian. Thus, the transform of a Gaussian that is localized near the origin is delocalized, and conversely. The intuitive explanation of this behavior is that more high-frequency Fourier components are required to represent a rapidly varying, localized function than a slowly varying, delocalized function. For example, the Fourier transform of the Gaussian approximate identity

$$\varphi_\epsilon(x) = \frac{1}{(2\pi\epsilon)^{n/2}} \exp\left(-\frac{|x|^2}{2\epsilon}\right)$$

is given by

$$\hat{\varphi}_\epsilon(k) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\epsilon|x|^2}{2}\right).$$

As $\epsilon \to 0^+$, we have $\varphi_\epsilon \rightharpoonup \delta$ and $\hat{\varphi}_\epsilon \rightharpoonup (2\pi)^{-n/2}$ in $\mathcal{S}^*$. The spectrum of the Gaussian becomes flatter as it concentrates at the origin. These limits are consistent with the result below that $\hat{\delta} = (2\pi)^{-n/2}$.

The following proposition, whose proof we leave to Exercise 11.13, gives the formulae for the Fourier transform of translates and convolutions. An important result is the fact that the Fourier transform maps the convolution product of two functions to their pointwise product. We will see in Section 11.9 that this is related to the translational invariance of the convolution.

Proposition 11.27 If $\varphi, \psi \in \mathcal{S}$ and $h \in \mathbb{R}^n$, then:

$$\widehat{\tau_h\varphi} = e^{-ik\cdot h}\hat{\varphi}, \tag{11.27}$$

$$\widehat{e^{ix\cdot h}\varphi} = \tau_h\hat{\varphi}, \tag{11.28}$$

$$\widehat{\varphi * \psi} = (2\pi)^{n/2}\hat{\varphi}\hat{\psi}. \tag{11.29}$$

Finally, we prove that $\mathcal{F}$ is invertible on $\mathcal{S}$ with a continuous inverse. First, we give a formula for the inverse.

Definition 11.28 If $\varphi \in \mathcal{S}$, then the *inverse Fourier transform* $\check{\varphi}$ is given by

$$\check{\varphi}(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{ik\cdot x}\varphi(k)\, dk.$$

We define $\mathcal{F}^* : \mathcal{S} \to \mathcal{S}$ by $\mathcal{F}^*\varphi = \check{\varphi}$.

We will prove that $\mathcal{F}^* = \mathcal{F}^{-1}$, meaning that

$$\check{\hat{\varphi}} = \varphi = \hat{\check{\varphi}} \qquad \text{for every } \varphi \in \mathcal{S}. \tag{11.30}$$

To motivate the proof of the inversion formula, we first give a formal calculation, based on the completeness formula in (11.33) below. Writing out $\mathcal{F}^*\hat{\varphi}$, and exchanging the order of integration, we find that

$$\begin{aligned}
\mathcal{F}^*\hat{\varphi}(x) &= \frac{1}{(2\pi)^n}\int e^{ik\cdot x}\left[\int e^{-ik\cdot y}\varphi(y)\, dy\right] dk \\
&= \int_{-\infty}^{\infty} \frac{1}{(2\pi)^n}\left[\int_{-\infty}^{\infty} e^{ik(x-y)}\, dk\right]\varphi(y)\, dy \\
&= \int \delta(x-y)\varphi(y)\, dy \\
&= \varphi(x).
\end{aligned}$$

The exchange of integration in this calculation is not justified by Fubini's theorem because the integral is not absolutely convergent. To make the argument rigorous, we introduce an "ultraviolet cut-off" in the integral before exchanging the order of integration.

Proposition 11.29 The map $\mathcal{F}^*$ is a continuous linear transformation on $\mathcal{S}$, and $\mathcal{F}^* = \mathcal{F}^{-1}$.

Proof. We have $\mathcal{F}^* = R\circ\mathcal{F}$, where R is the reflection defined by $R\varphi(x) = \varphi(-x)$, so the continuity of $\mathcal{F}^*$ on $\mathcal{S}$ follows from the continuity of R and $\mathcal{F}$.

The n-dimensional Fourier transform is the composition of one-dimensional Fourier transforms in each of the components x_i of $x \in \mathbb{R}^n$, $i = 1, \ldots, n$, so it suffices to prove the result for $n = 1$. Introducing a cut-off in the k-integral, and using Fubini's theorem to exchange the order of integration, we find that

$$\begin{aligned}
\mathcal{F}^*\hat{\varphi}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\left[\int_{-\infty}^{\infty} e^{-iky}\varphi(y)\, dy\right] dk \\
&= \frac{1}{2\pi}\lim_{R\to\infty}\int_{-R}^{R}\left[\int_{-\infty}^{\infty} e^{ik(x-y)}\varphi(y)\, dy\right] dk
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2\pi} \lim_{R\to\infty} \int_{-\infty}^{\infty} \left[\int_{-R}^{R} e^{ik(x-y)}\, dk \right] \varphi(y)\, dy \\
&= \lim_{R\to\infty} \int_{-\infty}^{\infty} \left[\frac{\sin R(x-y)}{\pi(x-y)}\, dx \right] \varphi(y)\, dy.
\end{aligned}
$$

From Proposition 11.22, the sequence $(\sin Rx)/\pi x$ is a delta sequence as $R \to \infty$, so $\mathcal{F}^*\hat{\varphi} = \varphi$. An identical argument shows that $\mathcal{F}\check{\varphi} = \varphi$. Therefore $\mathcal{F}, \mathcal{F}^* : \mathcal{S} \to \mathcal{S}$ are one-to-one, onto continuous maps, and $\mathcal{F}^* = \mathcal{F}^{-1}$. □

We could have instead introduced a Gaussian regularization,

$$
\mathcal{F}^*\hat{\varphi}(x) = \frac{1}{2\pi} \lim_{\epsilon\to 0^+} \int_{-\infty}^{\infty} e^{ikx-\epsilon k^2/2} \left[\int_{\infty}^{\infty} e^{iky}\varphi(y)\, dy \right] dk,
$$

exchanged the order of integration, and passed to the limit in the resulting Gaussian approximate identity.

11.6 The Fourier transform of tempered distributions

In this section, we define the Fourier transform of a tempered distribution. First, suppose that $f, \varphi \in \mathcal{S}$. Using the definition of the transform and exchanging the order of integration, which is justified by Fubini's theorem, we find that the action of the Fourier transform $\hat{f}$ on a test function φ is given by

$$
\begin{aligned}
\int \hat{f}(k)\varphi(k)\, dk &= \int \frac{1}{(2\pi)^{n/2}} \left(\int f(x) e^{-ik\cdot x}\, dx \right) \varphi(k)\, dk \\
&= \int f(x) \frac{1}{(2\pi)^{n/2}} \left(\int \varphi(k) e^{-ik\cdot x}\, dk \right) dx \\
&= \int f(x)\hat{\varphi}(x)\, dx. \qquad (11.31)
\end{aligned}
$$

In the notation of (11.9), this result means that $\mathcal{F}' = \mathcal{F}$. We therefore define the Fourier transform of a tempered distribution as follows.

Definition 11.30 The Fourier transform of a tempered distribution T is the tempered distribution $\hat{T} = \mathcal{F}T$ defined by

$$
\langle \hat{T}, \varphi \rangle = \langle T, \hat{\varphi} \rangle \qquad \text{for all } \varphi \in \mathcal{S}. \qquad (11.32)
$$

The inverse Fourier transform $\check{T} = \mathcal{F}^{-1}T$ on $\mathcal{S}^*$ is defined by

$$
\langle \check{T}, \varphi \rangle = \langle T, \check{\varphi} \rangle \qquad \text{for all } \varphi \in \mathcal{S}.
$$

The linearity and continuity of the Fourier transform on $\mathcal{S}$ implies that $\hat{T}$ is a tempered distribution. The map $\mathcal{F} : \mathcal{S}^* \to \mathcal{S}^*$ is a continuous, one-to-one transformation of $\mathcal{S}^*$ onto itself. The fact that $\mathcal{F}^{-1}$ is the inverse of $\mathcal{F}$ on $\mathcal{S}^*$ follows immediately from the corresponding result on $\mathcal{S}$, since

$$\langle \check{\hat{T}}, \varphi \rangle = \langle T, \check{\hat{\varphi}} \rangle = \langle T, \varphi \rangle \qquad \text{for all } \varphi \in \mathcal{S}.$$

The formulae for the Fourier transform of derivatives, translates, and convolutions carry over directly to distributions (see Exercise 11.13). For example,

$$\widehat{\partial^\alpha T} = (ik)^\alpha \hat{T}.$$

We also write the Fourier transform using the integral notation,

$$\hat{T}(k) = \frac{1}{(2\pi)^{n/2}} \int T(x) e^{-ik\cdot x}\, dx,$$

as if T were a function, with an analogous expression for the inverse. This notation should be interpreted simply as a short-hand for the definition in (11.32).

Example 11.31 Let us compute the Fourier transform of the delta function. From (11.32), we have

$$\langle \hat{\delta}, \varphi \rangle = \langle \delta, \hat{\varphi} \rangle = \hat{\varphi}(0).$$

From the formula for the Fourier transform on $\mathcal{S}$, we have

$$\hat{\varphi}(0) = \frac{1}{(2\pi)^{n/2}} \int \varphi(x)\, dx = \frac{1}{(2\pi)^{n/2}} \langle 1, \varphi \rangle.$$

Hence, the Fourier transform of the delta function is a constant,

$$\hat{\delta} = \frac{1}{(2\pi)^{n/2}}.$$

Using the integral notation, we get from the inversion formula the following Fourier representation of the delta function:

$$\delta(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{ik\cdot x}\, dk. \tag{11.33}$$

The formula for the transform of the derivative implies that the transform of the αth derivative of the delta function is a monomial,

$$\widehat{\partial^\alpha \delta} = \frac{1}{(2\pi)^{n/2}} (ik)^\alpha.$$

11.7 The Fourier transform on L^1

The Fourier integral

$$\hat{f}(k) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-ik\cdot x}\, dx \tag{11.34}$$

converges if and only if $f \in L^1(\mathbb{R}^n)$, meaning that

$$\int_{\mathbb{R}^n} |f(x)|\, dx < \infty.$$

We define the Fourier transform $\hat{f}$ of an L^1-function f by (11.34). This definition is consistent with the distributional definition, since Fubini's theorem justifies the exchange in the order of integration in (11.31) when $f \in L^1(\mathbb{R}^n)$.

Example 11.32 Let $f = \chi_{[-R,R]}$ be the characteristic function of the interval $[-R, R]$, sometimes called a "box" function. Then

$$\hat{f}(k) = \sqrt{2\pi}\frac{\sin Rk}{\pi k}.$$

Thus, the Fourier transform of a box function is a sinc function. The slow rate of decay of the Fourier transform as $k \to \infty$, of the order k^{-1}, is caused by the discontinuities in f. Although f belongs to L^1, the transform $\hat{f}$ does not. Thus, we cannot recover f from $\hat{f}$ by use of the inverse Fourier integral, but we can use the distributional definition of the inverse Fourier transform.

Example 11.33 For $a > 0$, let $f(x) = \exp(-a|x|)$. Then

$$\hat{f}(k) = \sqrt{\frac{2}{\pi}}\frac{a}{a^2 + k^2}.$$

The following result, called the Riemann-Lebesgue lemma, gives a basic property of the Fourier transform of L^1-functions. We denote by $C_0(\mathbb{R}^n)$ the space of continuous functions f that approach zero at infinity, meaning that for every $\epsilon > 0$ there is an R such that $|f(x)| < \epsilon$ when $|x| > R$. This space is the completion of $C_c(\mathbb{R}^n)$ with respect to the supremum norm, and is a Banach space.

Theorem 11.34 (Riemann-Lebesgue) If $f \in L^1(\mathbb{R}^n)$, then $\hat{f} \in C_0(\mathbb{R}^n)$, and

$$(2\pi)^{n/2}\|\hat{f}\|_\infty \leq \|f\|_1.$$

Proof. To prove the claim, we first observe that if $\varphi \in \mathcal{S}$, then

$$\begin{aligned} (2\pi)^{n/2}|\hat{\varphi}(k)| &= \left|\int e^{-ik\cdot x}\varphi(x)\, dx\right| \\ &\leq \int |\varphi(x)|\, dx. \end{aligned}$$

Taking the supremum of this inequality over k, we find that

$$(2\pi)^{n/2}\,\|\hat{\varphi}\|_\infty \leq \|\varphi\|_1\,.$$

The Schwartz space $\mathcal{S}$ is dense in L^1. Hence, if $f \in L^1$, there is sequence (φ_m) in $\mathcal{S}$ that converges to f with respect to the L^1-norm. Then $(\hat{\varphi}_m)$ is Cauchy in the supremum norm, since

$$(2\pi)^{n/2} \|\hat{\varphi}_m - \hat{\varphi}_\ell\|_\infty \le \|\varphi_m - \varphi_\ell\|_1 .$$

Since $\mathcal{S}$ is contained in C_0, and C_0 is complete, there is a function $\hat{g} \in C_0$ such that $\hat{\varphi}_m \to \hat{g}$ uniformly. Moreover, $\hat{g} = \hat{f}$ because

$$\begin{aligned} (2\pi)^{n/2} \left|\hat{g}(k) - \hat{f}(k)\right| &= (2\pi)^{n/2} \lim_{m\to\infty} \left|\hat{\varphi}_m(k) - \hat{f}(k)\right| \\ &= \lim_{m\to\infty} \left| \int [\varphi_m(x) - f(x)]\, e^{-ik\cdot x}\, dx \right| \\ &\le \liminf_{m\to\infty} \|\varphi_m - f\|_1 = 0. \end{aligned}$$

□

The Fourier transform is therefore a bounded linear map from L^1 into C_0. We may make L^1 into an algebra with the convolution product, and C_0 into an algebra with the pointwise product. The following proposition shows that the Fourier transform maps the convolution product into the pointwise product, up to a factor of $(2\pi)^{n/2}$, which depends on the normalization of the Fourier transform. Thus the Fourier transform is an algebra isomorphism of L^1 and its image $\mathcal{F}(L^1) \subset C_0$. The image $\mathcal{F}(L^1)$ is strictly smaller than C_0, but a precise description of it is difficult.

Theorem 11.35 (Convolution) If $f, g \in L^1(\mathbb{R}^n)$, then $f * g \in L^1(\mathbb{R}^n)$ and

$$\widehat{f * g} = (2\pi)^{n/2} \hat{f}\hat{g}.$$

Proof. Fubini's theorem implies that

$$\begin{aligned} \int |f * g(x)|\, dx &= \int \left| \int f(x-y) g(y)\, dy \right| dx \\ &\le \int \left[\int |f(x-y)|\, dx \right] |g(y)|\, dy \\ &= \left(\int |f(z)|\, dz \right) \left(\int |g(y)|\, dy \right), \end{aligned}$$

which shows that $f*g \in L^1(\mathbb{R}^n)$. Moreover, the absolute convergence of this integral implies that we can exchange the order of integration in the integral for the Fourier transform of $f * g$:

$$\begin{aligned} \widehat{f * g}(k) &= \frac{1}{(2\pi)^{n/2}} \int e^{-ik\cdot x} \left[\int f(x-y) g(y)\, dy \right] dx \\ &= \frac{1}{(2\pi)^{n/2}} \int e^{-ik\cdot y} \left[\int e^{-ik\cdot(x-y)} f(x-y)\, dx \right] g(y)\, dy \\ &= \frac{1}{(2\pi)^{n/2}} \left(\int e^{-ik\cdot z} f(z)\, dz \right) \left(\int e^{-ik\cdot y} g(y)\, dy \right) \end{aligned}$$

$$= (2\pi)^{n/2}\hat{f}\hat{g}.$$

□

One use of the convolution theorem is the computation of the inverse Fourier transform of the product of two functions whose inverse Fourier transform we know.

Example 11.36 For $a > 0$, we have that

$$\mathcal{F}\left[\frac{1}{\pi}\frac{a}{a^2+x^2}\right] = \frac{1}{\sqrt{2\pi}}e^{-a|k|}.$$

Taking the inverse Fourier transform of the equation

$$\frac{1}{\sqrt{2\pi}}e^{-(a+b)|k|} = \sqrt{2\pi}\,\frac{1}{\sqrt{2\pi}}e^{-a|k|}\,\frac{1}{\sqrt{2\pi}}e^{-b|k|},$$

where $a, b > 0$, and using the convolution theorem, we obtain the semigroup relation

$$\frac{1}{\pi}\frac{a+b}{(a+b)^2+x^2} = \left(\frac{1}{\pi}\frac{a}{a^2+x^2}\right) * \left(\frac{1}{\pi}\frac{b}{b^2+x^2}\right).$$

Finally, we make a few comments about the extension of the Fourier transform to a function of a complex variable, called the *Fourier-Laplace transform.* If $f : \mathbb{R}^n \to \mathbb{C}$ is an integrable function with compact support, then (11.34) defines an entire function $\hat{f} : \mathbb{C}^n \to \mathbb{C}$ (meaning that $\hat{f}(k)$ is a differentiable, or analytic, function of the complex variable k for all $k \in \mathbb{C}^n$), since the integral obtained by differentiation under the integral sign converges for every $k \in \mathbb{C}^n$. The *Paley-Wiener theorem*, which we do not state here, gives a precise characterization of the Fourier transforms of compactly supported functions. Similarly, considering the case of one variable for simplicity, if f is integrable and the support of $f(x)$ is contained in the half-line $x \geq 0$, then the Fourier transform $\hat{f}(k)$ is an analytic function of k in the lower-half plane $\operatorname{Im} k < 0$, because in that case the exponential e^{-ikx} decays as $x \to +\infty$. Setting $k = -iz$, and omitting the normalization factor of $\sqrt{2\pi}$, we obtain the *Laplace transform* of f,

$$\tilde{f}(z) = \int_0^\infty f(x)e^{-xz}\,dx,$$

which is analytic in the right-half plane $\operatorname{Re} z > 0$. More generally, if $\operatorname{supp} f \subset [0, \infty)$ and $f(x)e^{-ax}$ is integrable for some $a \in \mathbb{R}$, then $\tilde{f}(z)$ is analytic in the right-half plane $\operatorname{Re} z > a$. Methods from complex analysis, such as contour integration, may be used to study and invert the Fourier-Laplace transform.

The space of Fourier transforms of test functions in $\mathcal{D} = C_c^\infty$ is a space $\mathcal{L}$ of entire functions. Continuous linear functionals on $\mathcal{L}$, equipped with an appropriate topology, are called *ultradistributions.* The space $\mathcal{L}^*$ of ultradistributions contains the space $\mathcal{S}^*$ of tempered distributions, and the Fourier transform of an arbitrary distribution in $\mathcal{D}^*$ may be defined as an ultradistribution, even if it has exponential growth at infinity. For example, the Fourier transform

$$\widehat{e^{x^2}} = \mathcal{F}\left(\sum_{n=0}^{\infty} \frac{x^{2n}}{n!}\right) = \sqrt{2\pi} \sum_{n=0}^{\infty} \frac{\delta^{(2n)}}{n!}$$

is well-defined as an ultradistribution. The series on the right-hand side does not converge in $\mathcal{S}^*$, however, since Schwartz functions need not have convergent Taylor series expansions.

11.8 The Fourier transform on L^2

We have seen that the Fourier transform is an isomorphism on both the Schwartz space and on the space of tempered distributions equipped with their appropriate topologies. In this section, we will see that the Fourier transform is also an isomorphism on the Hilbert space $L^2(\mathbb{R}^n)$ of square-integrable functions. To avoid confusion with our notation for the duality pairing on $\mathcal{S}^* \times \mathcal{S}$, we denote the L^2-inner product by

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

The duality pairing and inner-product of $f \in L^2$ and $\varphi \in \mathcal{S}$ are related by

$$\langle T_{\overline{f}}, \varphi \rangle = (f, \varphi).$$

Not every square-integrable function on $\mathbb{R}^n$ is integrable; for example, the function $(1 + x^2)^{-1/2}$ belongs to $L^2(\mathbb{R})$ but not $L^1(\mathbb{R})$. Thus, we cannot define the Fourier transform of a general L^2-function directly by means of its Fourier integral. Instead, we will use the L^2-boundedness of the Fourier transform to extend it from $\mathcal{S}(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$.

If $\varphi \in \mathcal{S}$, then $\overline{\hat{\varphi}} = \check{\overline{\varphi}}$, since

$$\frac{1}{(2\pi)^{n/2}} \overline{\int \varphi(x) e^{-ik\cdot x}\, dx} = \frac{1}{(2\pi)^{n/2}} \int \overline{\varphi(x)} e^{ik\cdot x}\, dx.$$

Using (11.30) and (11.31), we see that for every $\varphi, \psi \in \mathcal{S}$

$$(\hat{\varphi}, \hat{\psi}) = \int_{\mathbb{R}^n} \overline{\hat{\varphi}}(x) \hat{\psi}(x)\, dx = \int_{\mathbb{R}^n} \check{\overline{\varphi}}(x) \hat{\psi}(x)\, dx = \int_{\mathbb{R}^n} \overline{\varphi}(x) \check{\hat{\psi}}(x)\, dx = (\varphi, \psi).$$

Thus, the Fourier transform is an isometric map

$$\mathcal{F} : \mathcal{S} \subset L^2 \to \mathcal{S} \subset L^2.$$

The Schwartz space $\mathcal{S}$ is dense in L^2, so the bounded linear transformation theorem implies that there is a unique isometric extension $\mathcal{F} : L^2 \to L^2$. Moreover, $\mathcal{F}^{-1} =$

$\mathcal{F}^*$, where $\mathcal{F}^*$ is the Hilbert space adjoint of $\mathcal{F}$. Consequently, we have the following theorem.

Theorem 11.37 (Plancherel) The Fourier transform $\mathcal{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a unitary map. For every $f, g \in L^2(\mathbb{R}^n)$, we have

$$(\hat{f}, \hat{g}) = (f, g), \tag{11.35}$$

where $\hat{f} = \mathcal{F}f$. In particular,

$$\int_{\mathbb{R}^n} |f(x)|^2 \, dx = \int_{\mathbb{R}^n} \left|\hat{f}(k)\right|^2 \, dk. \tag{11.36}$$

To compute the Fourier transform of a general function $f \in L^2$, we choose any sequence (φ_n) in $\mathcal{S}$ (or, more generally, in $L^1 \cap L^2$) that converges to f in L^2. Then $\hat{f}$ is the L^2-limit of $(\hat{\varphi}_n)$. For example,

$$\begin{aligned} \hat{f}(k) &= \lim_{R\to\infty} \int_{|x|\leq R} f(x) e^{-ik\cdot x} \, dx \\ &= \lim_{\epsilon\to 0^+} \int_{\mathbb{R}^n} f(x) e^{-ik\cdot x - \epsilon|x|^2} \, dx, \end{aligned} \tag{11.37}$$

where the limits are understood in the L^2-sense. The inverse transform may be computed in a similar way.

The Fourier transform is a unitary operator on $L^2(\mathbb{R}^n)$, so its spectrum lies on the unit circle in $\mathbb{C}$. The spectrum turns out to consist entirely of eigenvalues. We will describe it, without proof, in the one-dimensional case. Multi-dimensional eigenfunctions may be constructed from products of one-dimensional eigenfunctions in each of the coordinates.

Since $R\mathcal{F}^{-1} = \mathcal{F}$, where R is the reflection operator on L^2, we have $\mathcal{F}^2 = R$, and $\mathcal{F}^4 = I$. It follows that if $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathcal{F}$, then $\lambda^4 = 1$, so $\lambda \in \{1, i, -1, -i\}$. Each of these values is an eigenvalue of infinite multiplicity. A complete orthonormal set of eigenfunctions is given by the Hermite functions,

$$\varphi_n(x) = \frac{1}{\sqrt{2^n n! \sqrt{\pi}}} e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2}, \tag{11.38}$$

where $n = 0, 1, 2, \ldots$. One can prove that

$$\mathcal{F}\varphi_n = (-i)^n \varphi_n.$$

From Exercise 6.14, the Hermite functions are eigenfunctions of the differential operator

$$Au = -u'' + x^2 u.$$

Taking the Fourier transform of this expression, we find that the terms involving derivatives and multiplication by powers exchange places, so

$$\mathcal{F}Au = k^2\hat{u} - \hat{u}'' = A\mathcal{F}u.$$

Thus, A and $\mathcal{F}$ commute, which explains why they share a common set of eigenfunctions.

Once we know that the Hermite functions form an orthonormal basis of $L^2(\mathbb{R})$, we can give an alternative definition of the L^2-Fourier transform as

$$\mathcal{F}\left(\sum_{n=0}^{\infty} c_n\varphi_n\right) = \sum_{n=0}^{\infty}(-i)^n c_n\varphi_n.$$

The unitarity of the Fourier transform on L^2 can be seen immediately from this formula.

Just as we used Fourier series to define Sobolev spaces of periodic functions, we can use the Fourier transform to define Sobolev spaces of functions with square-integrable derivatives on $\mathbb{R}^n$. Since

$$\widehat{\partial^\alpha f} = (ik)^\alpha \hat{f},$$

the partial derivatives of f of order less than or equal to s are square-integrable if and only if $(ik)^\alpha \hat{f}$ is a square-integrable function for $|k| \leq s$. This is the case if the function

$$\left(1 + |k|^2\right)^{s/2} \hat{f}$$

is square-integrable. More generally, we can define Sobolev spaces of distributions with fractional, or even negative, order L^2-derivatives.

Definition 11.38 Let $s \in \mathbb{R}$. The Sobolev space $H^s(\mathbb{R}^n)$ consists of all distributions $f \in \mathcal{S}^*$ whose Fourier transform $\hat{f} : \mathbb{R}^n \to \mathbb{C}$ is a regular distribution and

$$\int_{\mathbb{R}^n} \left(1 + |k|^2\right)^s \left|\hat{f}(k)\right|^2 \, dk < \infty.$$

A similar proof to the proof of the Sobolev embedding theorem for periodic functions, in Theorem 7.9, shows that if $f \in H^s(\mathbb{R}^n)$ for $s > n/2$, then $f \in C_0(\mathbb{R}^n)$ (see Exercise 11.12).

11.9 Translation invariant operators

There is a close connection between the Fourier transform and the group of translation operators τ_h defined in Example 11.16. Since

$$\tau_h e^{-ik\cdot x} = e^{ik\cdot h} e^{-ik\cdot x},$$

the exponential functions $e^{-ik\cdot x}$, with $k \in \mathbb{R}^n$, are eigenvectors of τ_h in $\mathcal{S}^*$ with eigenvalues $e^{ik\cdot h}$. The Fourier transform is therefore an expansion of a function or distribution with respect to the eigenvectors of τ_h. If $A : \mathcal{S}^* \to \mathcal{S}^*$ is a linear translation invariant operator, meaning that $A\tau_h = \tau_h A$, then we expect that there is a basis of common eigenvectors of τ_h and A, so that A can be diagonalized by use of the Fourier transform. In that case, the action of A on a distribution is to multiply the Fourier transform of the distribution by a C^∞-function $\hat{a}$ of polynomial growth,

$$\widehat{AT} = \hat{a}\hat{T}.$$

The function $\hat{a}$ is called the *symbol* of the operator A. Inverting the transform, we find from the convolution theorem that

$$AT = \frac{1}{(2\pi)^{n/2}} a * T,$$

with a suitable definition of the convolution $a * T$. Since $\tau_h(a * T) = a * (\tau_h T)$, the convolution is translation invariant.

Example 11.39 A constant coefficient linear differential operator $P : \mathcal{S}^* \to \mathcal{S}^*$ is translation invariant, and is given by

$$PT = \sum_{|\alpha| \le d} c_\alpha \partial^\alpha T$$

for constants c_α. The Fourier representation is $\widehat{PT} = \hat{p}\hat{T}$, where

$$\hat{p}(k) = \sum_{|\alpha| \le d} c_\alpha (ik)^\alpha.$$

Thus, the symbol of a differential operator is a polynomial. The convolution form of the operator is

$$PT = \left(\sum_{|\alpha| \le d} c_\alpha \partial^\alpha \delta \right) * T.$$

It can be much simpler to define an operator in terms of its symbol than by an explicit formula for its action on a function.

Example 11.40 The symbol of the differential operator $(-\Delta + I)$ is the quadratic polynomial $(|k|^2 + 1)$. The square-root $(-\Delta + I)^{1/2}$ is the nonlocal operator with symbol $(|k|^2 + 1)^{1/2}$. Its action on a distribution T is given by

$$(-\Delta + I)^{1/2} T = \mathcal{F}^{-1} \left[(|k|^2 + 1)^{1/2} \hat{T} \right].$$

The inverse operator $(-\Delta + I)^{-1}$ has symbol $(|k|^2 + 1)^{-1}$, so

$$(-\Delta + I)^{-1} T = g * T,$$

where g is the Green's function of $(-\Delta + I)$, given by

$$g = \frac{1}{(2\pi)^{n/2}} \mathcal{F}^{-1} \left[\frac{1}{|k|^2 + 1} \right].$$

For $n = 3$, where n is the number of space dimensions, computation of the inverse Fourier transform gives

$$g(x) = \frac{1}{4\pi} \frac{e^{-|x|}}{|x|}.$$

For $n = 2$, the Green's function may be expressed in terms of Bessel functions. We will study some other examples of Green's functions in the next section.

We may also consider translation invariant operators defined on a subspace of $\mathcal{S}^*$. For example, any bounded function $\hat{a} \in L^\infty(\mathbb{R}^n)$ is the symbol of a translation invariant operator $A : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ defined by $\widehat{Af} = \hat{a}\hat{f}$.

Example 11.41 For $g \in L^1(\mathbb{R}^n)$, we define the convolution integral operator $G : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ by

$$Gf(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} g(x - y) f(y)\, dy. \tag{11.39}$$

The symbol of G is $\hat{g}$. Since $g \in L^1$, the Riemann-Lebesgue lemma (Theorem 11.34) implies that $\hat{g} \in C_0$. Thus, the Fourier transform $\mathcal{F}$ diagonalizes G, and $G = \mathcal{F}^* \hat{g} \mathcal{F}$ is unitarily equivalent to multiplication by $\hat{g}$. Unless $\hat{g}$ = constant on a set of nonzero measure, the multiplication operator has a continuous spectrum, given by the closure of the range of $\hat{g}$, so this is also the spectrum of G.

More generally, the map G is well defined on L^2 whenever $\hat{g} \in L^\infty$ is bounded. For example, suppose that $\hat{f}_R$ is the function obtained by truncating the Fourier transform of $f \in L^2(\mathbb{R})$ at wavenumbers k with $|k| \le R$:

$$\hat{f}_R(x) = \begin{cases} \hat{f}(k) & \text{if } |k| \le R, \\ 0 & \text{if } |k| > R. \end{cases}$$

Then $\hat{f}_R = \chi_{[-R,R]} \hat{f}$. Since

$$\mathcal{F}^{-1}\left(\chi_{[-R,R]}\right) = \sqrt{\frac{2}{\pi}} R \operatorname{sinc}(Rx),$$

the function $f_R = \mathcal{F}^{-1}\left[\hat{f}_R\right]$ is given by

$$f_R = \frac{R}{\pi} \operatorname{sinc}(Rx) * f.$$

Example 11.42 The symbol of the translation operator τ_h itself is $e^{-ik \cdot h}$. The translation operators $\{\tau_h \mid h \in \mathbb{R}^n\}$ form a unitary group acting on $L^2(\mathbb{R}^n)$. If $h \neq 0$, then the spectrum of τ_h is the unit circle in $\mathbb{C}$ and is purely continuous.

Example 11.43 The operator $\mathbb{H} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ with symbol

$$\hat{h}(k) = i \operatorname{sgn} k$$

is called the *Hilbert transform.* Since the modulus of the symbol is equal to one, Plancherel's theorem implies that $\mathbb{H}$ is a unitary map of $L^2(\mathbb{R})$ onto itself. Since $\hat{h}^2 = -1$, we have $\mathbb{H}^2 = -I$. From Exercise 11.22 and the convolution theorem,

$$\mathbb{H}f = -\frac{1}{\pi}\left(\text{p.v.}\frac{1}{x}\right) * f.$$

The Hilbert transform is one of the simplest examples of a *singular integral operator.* Its properties are much more transparent from the Fourier representation than the convolution form.

Example 11.44 The operator $R_{pq} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ with symbol

$$\hat{r}_{pq} = \frac{k_p k_q}{|k|^2}$$

is called the *Riesz transform.* Since $|\hat{r}_{pq}| \le 1$, R_{pq} is a bounded linear map on $L^2(\mathbb{R}^n)$. The Riesz transform recovers the second derivatives of a differentiable function from its Laplacian:

$$\frac{\partial^2 f}{\partial x_p \partial x_q} = R_{pq}\Delta f.$$

One can also define *pseudodifferential operators*, whose symbol $\hat{a}(x,k)$ is a function belonging to a suitable class that is allowed to depend on both x and k, so that

$$\begin{aligned} Af(x) &= \frac{1}{(2\pi)^{n/2}} \int \hat{a}(x,k)\hat{f}(k)e^{ik\cdot x}\, dk \\ &= \frac{1}{(2\pi)^n} \int \hat{a}(x,k)e^{ik\cdot(x-y)} f(y)\, dydk. \end{aligned}$$

These operators are not translation invariant, and they allow the use of Fourier methods in the analysis of variable coefficient, linear partial differential equations.

11.10 Green's functions

Constant coefficient, linear partial differential equations on $\mathbb{R}^n$ may be solved by use of the Fourier transform. In particular, we can use the distributional Fourier transform to compute their Green's functions.

The Green's function g of the Laplacian on $\mathbb{R}^n$ is a distributional solution of the equation

$$-\Delta g = \delta. \tag{11.40}$$

The delta function has the physical interpretation of the density of a point source located at the origin, and the Green's function g is the potential of the point source. Taking the Fourier transform of (11.40), we find that

$$|k|^2\hat{g} = \frac{1}{(2\pi)^{n/2}}.$$

A complication in solving this equation for g is that the symbol $|k|^2$ of the Laplacian vanishes at $k = 0$. We therefore need to interpret division by $|k|^2$ in an appropriate sense. From Example 11.8, if $n \geq 3$, then a solution for $\hat{g}$ is the regular distribution

$$\hat{g}(k) = \frac{1}{(2\pi)^{n/2}}\frac{1}{|k|^2},$$

and the Green's function is

$$g(x) = \frac{1}{(2\pi)^{n/2}}\mathcal{F}^{-1}\left(\frac{1}{|k|^2}\right).$$

The solution is not unique. We may add an arbitrary linear combination of δ and first-order partial derivatives of δ to $\hat{g}$. The inverse transform of this distribution is a linear polynomial in x, which is a solution of the homogeneous Laplace equation. We omit this function of integration for simplicity.

We will compute the inverse transform of $\hat{g}$ explicitly when $n = 3$; the computation for $n \geq 4$ is similar. Since $\hat{g}(k)$ decays too slowly as $|k| \to \infty$ to be integrable, we introduce a cut-off, as in (11.37). Using spherical polar coordinates (r, θ, φ) in k-space, with the x-direction corresponding to $\theta = 0$, we find from the inversion formula and the sinc integral in (11.15), that

$$\begin{aligned} g(x) &= \frac{1}{(2\pi)^3}\lim_{R\to\infty}\int_{|x|<R}\frac{e^{ik\cdot x}}{|x|^2}\,dk \\ &= \frac{1}{(2\pi)^3}\lim_{R\to\infty}\int_0^R\int_0^\pi\int_0^{2\pi}\frac{e^{ir|x|\cos\theta}}{r^2}\,r^2\sin\theta\, d\varphi\, d\theta\, dr \\ &= \frac{1}{(2\pi)^2}\lim_{R\to\infty}\int_0^R\frac{2\sin r|x|}{r|x|}\,dr \\ &= \frac{1}{(2\pi)^2}\frac{\pi}{|x|}. \end{aligned}$$

It follows that the three-dimensional, free-space Green's function for Laplace's equation is

$$g(x) = \frac{1}{4\pi|x|},$$

as we found in Section 10.6 by a different method. For $n = 2$, a solution for $\hat{g}$ is

$$\hat{g}(k) = \frac{1}{2\pi}\text{f.p.}\,\frac{1}{|k|^2},$$

where the finite part distribution is defined in Example 11.8. One can show that the inverse Fourier transform of this distribution is of the form

$$g(x) = \frac{1}{2\pi} \log\left(\frac{1}{|x|}\right) + C$$

for a suitable constant C, also in agreement with our previous result.

Next, we consider the initial value problem for the heat or diffusion equation. The Green's function $g(x,t)$ is the solution of the following initial value problem:

$$\begin{aligned} g_t &= \frac{1}{2}\Delta g && \text{for } x \in \mathbb{R}^n,\ t > 0, \\ g(\cdot, t) &\in \mathcal{S}^*(\mathbb{R}^n) && \text{for } t > 0, \\ g(x,0) &= \delta(x) && \text{for } x \in \mathbb{R}^n. \end{aligned}$$

Taking the Fourier transform $\mathcal{F}_x$ with respect to x of this equation, we find that $\hat{g}(k,t) = \mathcal{F}_x g(x,t)$ satisfies the ODE

$$\hat{g}_t = -\frac{1}{2}|k|^2 \hat{g}, \qquad \hat{g}(k,0) = \frac{1}{(2\pi)^{n/2}}.$$

The solution is given by

$$\hat{g}(k,t) = \frac{1}{(2\pi)^{n/2}} e^{-t|k|^2/2}.$$

Using Proposition 11.26 to invert the transform, we obtain that

$$g(x,t) = \frac{1}{(2\pi t)^{n/2}} e^{-|x|^2/(2t)}.$$

The solution $u(x,t)$ of the heat equation with initial condition

$$u(x,0) = f(x),$$

is given by a convolution with the Green's function:

$$u(x,t) = \frac{1}{(2\pi t)^{n/2}} \int_{\mathbb{R}^n} e^{-|x-y|^2/(2t)} f(y)\, dy.$$

Since the Green's function is a Schwartz function, this expression makes sense as a convolution for any initial data $f \in \mathcal{S}^*$. The solution is C^∞ in both x and t when $t > 0$. This is the *smoothing property* of the heat equation. It can be shown that the solution of the initial value problem for the heat equation is not unique (see Exercise 11.24). There is, however, a unique solution of polynomial growth, and this is the one obtained by use of the Fourier transform.

11.11 The Poisson summation formula

The *Poisson summation formula* states that a large class of functions $f : \mathbb{R} \to \mathbb{C}$ satisfy the following identity:

$$\sum_{n=-\infty}^{\infty} f(2\pi n) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{f}(n). \tag{11.41}$$

The presence or absence of factors 2π in this equation depends on the normalization of the Fourier transform. This formula may be used to derive identities between infinite series, or even to sum a series explicitly. It can also be used to connect the Fourier series of a periodic function with the Fourier transform.

Theorem 11.45 Suppose that $f \in C^1(\mathbb{R})$, and there exist constants $C > 0$, $\epsilon > 0$ such that

$$\left| \left(1 + x^2\right)^{1/2+\epsilon} f(x) \right| \leq C, \quad \left| \left(1 + x^2\right)^{1/2+\epsilon} f'(x) \right| \leq C \tag{11.42}$$

for all $x \in \mathbb{R}$. Then we have the identity

$$\sum_{n=-\infty}^{\infty} f(x + 2\pi n) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} e^{inx} \hat{f}(n). \tag{11.43}$$

Proof. The condition in (11.42) implies that the sum

$$g(x) = \sum_{n=-\infty}^{\infty} f(x + 2\pi n) \tag{11.44}$$

converges uniformly, and g is a continuously differentiable 2π-periodic function. Therefore, from Lemma 7.8, the Fourier series of g converges uniformly, and

$$g(x) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \left(\int_0^{2\pi} e^{-iny} g(y)\, dy \right) e^{inx}.$$

Since g is related to f by (11.44), we can rewrite this as (11.43). □

Evaluation of (11.43) at $x = 0$ gives the Poisson summation formula (11.41).

Example 11.46 The *Jacobi theta function* θ is defined for $t > 0$ by

$$\theta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t}.$$

The Poisson summation formula implies that the theta function has the following symmetry property:

$$\theta(t) = \frac{1}{\sqrt{t}} \theta(1/t). \tag{11.45}$$

Theta functions have important connections with Riemann surfaces and the theory of integrable systems. They also arise in the solution of the heat equation on the circle, as in (7.21).

The Poisson summation formula holds, in particular, for Schwartz functions. The convergence of the series on the left-hand side of (11.41) for every $f \in \mathcal{S}$ implies that the series $\sum_{n=-\infty}^{\infty} \delta_{2\pi n}$ converges in $\mathcal{S}^*$. The series on the right-hand side of (11.41) may be written as:

$$\frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{f}(n) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \int e^{-inx} f(x)\, dx.$$

Hence, the series $\sum_{n=-\infty}^{\infty} e^{-inx}$ also converges in $\mathcal{S}^*$. Changing n to $-n$ in this sum, we obtain the following identity of tempered distributions:

$$\sum_{n=-\infty}^{\infty} \delta_{2\pi n}(x) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{inx}. \tag{11.46}$$

This equation may be interpreted as the Fourier series expansion of a periodic array of delta functions (sometimes called the "delta comb"). Its Fourier coefficients are constants, independent of n.

More generally, we say that a distribution $T \in \mathcal{S}(\mathbb{R})$ is periodic with period 2π if $\tau_{2\pi} T = T$. In that case, one can show that

$$\hat{T} = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{T}_n \delta_{2\pi n}$$

for suitable Fourier coefficients $\hat{T}_n \in \mathbb{C}$. The Fourier coefficients have polynomial growth in n, meaning that there are constants $C > 0$ and $d \in \mathbb{N}$ such that

$$\left|\hat{T}_n\right| \leq C \left(1 + n^2\right)^{d/2}.$$

Thus, the Fourier transform of a periodic function or distribution is an $\mathcal{S}^*$-convergent linear combination of delta functions supported at $2\pi n$. The strengths of the delta functions give the Fourier coefficients of the periodic function. The distribution T is given by the $\mathcal{S}^*$-convergent Fourier series

$$T(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{T}_n e^{inx}.$$

11.12 The central limit theorem

A *random variable* describes the observation, or measurement, of a number associated with a random event. We say that a real-valued random variable X is *absolutely continuous* if its distribution may be described by a *probability density*

$p \in L^1(\mathbb{R})$, meaning that for any $a \le b$, the probability that X has a value between a and b is given by

$$\Pr(a \le X \le b) = \int_a^b p(x)\,dx. \tag{11.47}$$

Since a probability is a number between zero and one, the density function p is nonnegative and

$$\int_{-\infty}^{\infty} p(x)\,dx = 1. \tag{11.48}$$

We call any function p with these properties, a probability density. If X is not absolutely continuous (for example, because it takes integer values with probability one), then its distribution is described by a probability measure on $\mathbb{R}$ that does not have a probability density function. We consider absolutely continuous random variables for simplicity, but the central limit theorem does not depend on this restriction.

The expected value of a function $f(X)$ of X is given in terms of the density p by

$$\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)\,dx,$$

provided that this integral converges. The *mean* μ and the *variance* σ^2 of X are given by

$$\mu = \mathbb{E}[X], \qquad \sigma^2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

The expected deviation of X from its mean is therefore of the order of the *standard deviation* σ. If the mean and variance of X are finite, then the random variable Y defined by $X = \mu + \sigma Y$ has mean zero and variance one, so we can normalize the mean of X to zero and the variance of X to one by an affine transformation. In that case,

$$\int_{-\infty}^{\infty} xp(x)\,dx = 0, \quad \int_{-\infty}^{\infty} x^2 p(x)\,dx = 1. \tag{11.49}$$

Example 11.47 We say that a real random variable X is a *Gaussian*, or *normal*, random variable with mean μ and variance σ^2 if its probability density p is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}.$$

If $\mu = 0$ and $\sigma^2 = 1$, then we say that X is a standard Gaussian.

We say that N random variables $\{X_1, X_2, \dots, X_N\}$ are *independent* if

$$\begin{aligned}&\Pr(a_1 \le X_1 \le b_1, a_2 \le X_2 \le b_2, \dots, a_N \le X_N \le b_N) \\ &\qquad = \Pr(a_1 \le X_1 \le b_1)\Pr(a_2 \le X_2 \le b_2)\dots\Pr(a_N \le X_N \le b_N).\end{aligned}$$

In that case,

$$\mathbb{E}[f_1(X_1)f_2(X_2)\dots f_N(X_N)] = \mathbb{E}[f_1(X_1)]\,\mathbb{E}[f_2(X_2)]\dots\mathbb{E}[f_N(X_N)].$$

Suppose that $\{X_1, X_2, \dots, X_N\}$ have a *joint probability density* $p(x_1, x_2, \dots, x_N)$, meaning that

$$\begin{aligned}&\Pr(a_1 \le X_1 \le b_1, a_2 \le X_2 \le b_2, \dots, a_N \le X_N \le b_N) \\ &\qquad = \int_{a_N}^{b_N}\dots\int_{a_2}^{b_2}\int_{a_1}^{b_1} p(x_1, x_2, \dots, x_N)\,dx_1 dx_2\dots dx_N.\end{aligned}$$

Then the random variables are independent if and only if p has the form

$$p(x_1, x_2, \dots, x_N) = p_1(x_1)p_2(x_2)\dots p_N(x_N).$$

Intuitively, independence means that the value taken by one of the random variables has no influence on the values taken by the others.

In many applications it is important to consider the sum of a large number of independent, identically distributed random variables. For example, a standard way to reduce nonsystematic errors in the experimental measurement of a given quantity is to measure the quantity many times and take the average. The central limit theorem explains how this error reduction works, and also gives an estimate of the expected difference between the measured value of the quantity and its true value. As we will see, if the experimental measurements are independent and randomly distributed with mean equal to the true value and with finite variance σ^2, then for sufficiently large N the distribution of the average value of N measurements is approximately Gaussian with mean equal to the true value and variance σ^2/N. Thus, one needs to take four times as many measurements in order to double the accuracy. This example is the original application that led Gauss to introduce the Gaussian distribution.

A second example is the discrete-time *random walk*. Considering the case of one space dimension for simplicity, we suppose that a particle starts at the origin at time zero and moves a random distance $X_n \in \mathbb{R}$ at time $n \in \mathbb{N}$, where X_m and X_n are independent, identically distributed random variables for $m \neq n$. The particle then takes random steps up and down the real line. The total distance moved by the particle after N steps is

$$S_N = \sum_{n=1}^{N} X_n. \tag{11.50}$$

A natural question is: What is the probability distribution of the position S_N of the particle after N steps, given the probability distribution of each individual step? The central limit theorem describes the limiting behavior of S_N as $N \to \infty$. For instance, if each individual step has mean zero and variance one, then the distribution of S_N approaches a Gaussian distribution with mean zero and variance N. The corresponding $\sqrt{N}$-growth of S_N is characteristic of sums of N independent random variables: the sums do not remain bounded as $N \to \infty$, but there is a large amount of cancellation, so the sums grow at a slower rate than the number of their terms.

The Gaussian distribution is universal, in the sense that the distribution of any sum of a large number of independent, identically distributed random variables with finite mean and variance is approximately Gaussian, whatever the details of the probability distribution of the individual random variables. The central limit theorem remains true for sums of non-identical, independent random variables, under a suitable, mild condition (such as the *Lindeberg condition*) that ensures the distribution of the sum is not dominated by the distribution of a small number of the individual random variables. Moreover, some weak dependence between the variables may also be permitted.

Suppose that X and Y are independent random variables with probability density functions p_X and p_Y, respectively. Then

$$\begin{aligned} \Pr(a \le X + Y \le b) &= \int\int_{a \le x+y \le b} p_X(x) p_Y(y)\, dx dy \\ &= \int_a^b \left(\int_{-\infty}^{\infty} p_X(z-y) p_Y(y)\, dy \right) dz. \end{aligned}$$

Thus, the probability density of $X+Y$ is the convolution of the probability densities of X and Y. Hence, the convolution theorem implies that the Fourier transforms of the densities multiply:

$$\hat{p}_{X+Y} = \sqrt{2\pi}\,\hat{p}_X \hat{p}_Y.$$

We can obtain the same result by an equivalent probabilistic argument. The *characteristic function* φ_X of a random variable X is defined by

$$\varphi_X(k) = \mathbb{E}\left[e^{ikX}\right].$$

What we have called the characteristic function χ_A of a set A is then referred to as the *indicator function* of the set. If X is absolutely continuous with probability density p, then

$$\varphi_X(k) = \int_{-\infty}^{\infty} e^{ikx} p(x)\, dx = \sqrt{2\pi}\,\hat{p}(-k).$$

Thus, up to normalization conventions, the characteristic function is the Fourier transform of the probability density. If X and Y are independent, then we have

$$\varphi_{X+Y}(k) = \mathbb{E}\left[e^{ik(X+Y)}\right] = \mathbb{E}\left[e^{ikX}\right]\mathbb{E}\left[e^{ikY}\right] = \varphi_X(k)\varphi_Y(k),$$

which agrees with the previous result. Because the Fourier transform maps convolutions to products, in studying sums of independent random variables it is much simpler to consider the characteristic functions rather than the densities themselves, and we shall use this observation to prove the central limit theorem.

Example 11.48 If X is a Gaussian random variable with mean μ and variance σ^2, then the formula for the Fourier transform of a Gaussian implies that

$$\mathbb{E}\left[e^{ikX}\right] = e^{i\mu k}e^{-\sigma^2 k^2/2}.$$

The product of such characteristic functions is another function of the same form, in which the means and variances add together. Consequently, the sum of independent Gaussian random variables is a Gaussian whose mean and variance is the sum of the individual means and variances, and problems involving Gaussian random variables are stochastically linear. The product of two independent Gaussian random variables is not Gaussian, however.

Suppose that $\{X_1, X_2, \dots, X_N\}$ is a sequence of independent, identically distributed, random variables with finite mean and variance, and probability density p. By making an affine transformation, we may assume that the mean is zero and the variance is one without loss of generality. The probability density p_N of the sum $S_N = X_1 + X_2 + \ldots + X_N$ is given by

$$p_N = \underbrace{p * p * \cdots * p}_{N \text{ times}}.$$

We denote the probability density of $S_N/\sqrt{N}$ by q_N. Then

$$q_N(x) = \sqrt{N} p_N\left(\sqrt{N}x\right). \tag{11.51}$$

We will prove that $S_N/\sqrt{N}$ converges to a standard Gaussian as $N \to \infty$ in the following sense.

Definition 11.49 A sequence (X_n) of random variables *converges in distribution* to a random variable X if

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \qquad \text{for every } f \in C_b(\mathbb{R}),$$

where $C_b(\mathbb{R})$ is the space of bounded, continuous functions $f : \mathbb{R} \to \mathbb{C}$. A sequence (p_n) of probability densities *converges weakly* to a probability density p if

$$\lim_{n\to\infty} \int f(x)p_n(x)\,dx = \int f(x)p(x)\,dx \qquad \text{for every } f \in C_b(\mathbb{R}). \tag{11.52}$$

If X_n and X are absolutely continuous random variables with probability densities p_n and p, respectively, then X_n converges in distribution to X if and only if p_n converges weakly to p. Approximating the characteristic function of the interval $[a, b]$ by continuous functions, one can then also show that

$$\lim_{n\to\infty} \Pr\left(a \le X_n \le b\right) = \Pr\left(a \le X \le b\right).$$

The following theorem, called Lévy's continuity theorem, provides a useful necessary and sufficient condition for weak convergence.

Theorem 11.50 (Continuity) A sequence (p_n) of probability densities converges weakly to a probability density p if and only if $(\hat{p}_n)$ converges pointwise to $\hat{p}$.

Proof. If p_n converges weakly to p, then (11.52) with $f(x) = e^{-ikx}$ implies that $\hat{p}_n(k)$ converges to $\hat{p}(k)$.

We prove the converse statement in several steps. First, we show that if $\hat{p}_n$ converges pointwise to $\hat{p}$, then (11.52) holds for every Schwartz function $f \in \mathcal{S}$. Since the Fourier transform maps $\mathcal{S}$ onto $\mathcal{S}$, an equivalent statement is that

$$\lim_{n\to\infty} \int \hat{f}(x) p_n(x)\, dx = \int \hat{f}(x) p(x)\, dx \qquad \text{for every } f \in \mathcal{S}.$$

From Fubini's theorem, as in (11.31), this statement is equivalent to

$$\lim_{n\to\infty} \int f(k) \hat{p}_n(k)\, dk = \int f(k) \hat{p}(k)\, dk \qquad \text{for every } f \in \mathcal{S}. \tag{11.53}$$

Since p_n is a probability density, the Riemann-Lebesgue lemma, Theorem 11.34, implies that $\hat{p}_n$ is a continuous function with $|\hat{p}_n(k)| \le 1/\sqrt{2\pi}$ for every $k \in \mathbb{R}$. Hence (11.53) follows from the pointwise convergence of $\hat{p}_n$, the integrability of f, and the Lebesgue dominated convergence theorem.

If $f \in C_0(\mathbb{R})$ is a continuous function that vanishes at infinity, then there is a sequence (φ_m) of Schwartz functions that converges uniformly to f. The estimate

$$\begin{aligned}
\left| \int f(x) \left[p_n(x) - p(x)\right]\, dx \right| \quad &\le \quad \left| \int \left[f(x) - \varphi_m(x)\right] p_n(x)\, dx \right| \\
&\qquad + \left| \int \varphi_m(x) \left[p_n(x) - p(x)\right]\, dx \right| \\
&\qquad + \left| \int \left[\varphi_m(x) - f(x)\right] p(x)\, dx \right| \\
&\le \quad 2\|f - \varphi_m\|_\infty + \left| \int \varphi_m(x) \left[p_n(x) - p(x)\right]\, dx \right|,
\end{aligned}$$

and (11.52) for $f = \varphi_m \in \mathcal{S}$, implies that (11.52) holds for $f \in C_0(\mathbb{R})$.

In order to show (11.52) for $f \in C_b(\mathbb{R})$, we first prove that

$$\lim_{R\to\infty} \limsup_{n\to\infty} \int_{|x|\ge R} p_n(x)\, dx = 0. \tag{11.54}$$

This condition means that probability cannot escape to infinity as $n \to \infty$. The family of probability measures associated with the densities p_n is then said to be *tight.* For each $R > 0$, we choose $\varphi_R \in C_0(\mathbb{R})$ such that $0 \leq \varphi_R(x) \leq 1$ for all $x \in \mathbb{R}$ and $\varphi_R(x) = 1$ when $|x| \leq R$. Then, by the dominated convergence theorem, and the fact that $\int p(x)\, dx = 1$, we have

$$\lim_{R\to\infty} \int \varphi_R(x) p(x)\, dx = 1. \tag{11.55}$$

Also, since $\varphi_R \in C_0(\mathbb{R})$,

$$\lim_{n\to\infty} \int \varphi_R(x) p_n(x)\, dx = \int \varphi_R(x) p(x)\, dx.$$

Using the fact that $\int p_n(x)\, dx = 1$, we therefore have

$$\begin{aligned} \limsup_{n\to\infty} \int_{|x|\geq R} p_n(x)\, dx &\geq \limsup_{n\to\infty} \int \left[1 - \varphi_R(x)\right] p_n(x)\, dx \\ &\geq 1 - \int \varphi_R(x) p(x)\, dx, \end{aligned}$$

and (11.54) follows from (11.55).

For $f \in C_b(\mathbb{R})$ and $R > 0$, we define $f_R = \varphi_R f$. Since $f(x) = f_R(x)$ for $|x| \leq R$, we have the following estimate:

$$\begin{aligned} \left| \int f(x) \left[p_n(x) - p(x)\right]\, dx \right| &\leq \left| \int \left[f(x) - f_R(x)\right] p_n(x)\, dx \right| \\ &\quad + \left| \int f_R(x) \left[p_n(x) - p(x)\right]\, dx \right| \\ &\quad + \left| \int \left[f_R(x) - f(x)\right] p(x)\, dx \right| \\ &\leq 2\|f\|_\infty \left[\int_{|x|\geq R} \left[p_n(x) + p(x)\right]\, dx \right] \\ &\quad + \left| \int f_R(x) \left[p_n(x) - p(x)\right]\, dx \right| \end{aligned}$$

It then follows from (11.52) for $f = f_R \in C_0(\mathbb{R})$ and (11.54) that we can make the right hand side of this equation arbitrarily small for all sufficiently large n. Hence, equation (11.52) holds for all $f \in C_b(\mathbb{R})$. □

We can now prove the following central limit theorem.

Theorem 11.51 (Central limit) Let S_N be the sum of N independent, identically distributed, absolutely continuous, real random variables with mean zero and variance one. Then $S_N/\sqrt{N}$ converges in distribution as $N \to \infty$ to a Gaussian random variable with mean zero and variance one.

Proof. From Theorem 11.50, we just have to show that the Fourier transform of the density q_N of $S_N/\sqrt{N}$ converges pointwise to the Fourier transform of the standard Gaussian density. Taking the Fourier transform of (11.51), we find that

$$\hat{q}_N(k) = \hat{p}_N\left(\frac{k}{\sqrt{N}}\right). \tag{11.56}$$

Since p_N is the N-fold convolution of p, the convolution theorem, Theorem 11.35, implies that

$$\hat{p}_N\left(\frac{k}{\sqrt{N}}\right) = (2\pi)^{(N-1)/2}\left[\hat{p}\left(\frac{k}{\sqrt{N}}\right)\right]^N. \tag{11.57}$$

We Taylor expand e^{-iz} as

$$e^{-iz} = 1 - iz - \frac{1}{2}z^2\left[1 + r(z)\right],$$

where $r(z)$ is a continuous function that vanishes at $z = 0$ and is uniformly bounded on the real line. Using the conditions in (11.48) and (11.49), we find that

$$\begin{aligned}
\hat{p}\left(\frac{k}{\sqrt{N}}\right) &= \frac{1}{\sqrt{2\pi}}\int e^{-ikx/\sqrt{N}}p(x)\,dx \\
&= \frac{1}{\sqrt{2\pi}}\int\left\{1 - \frac{ikx}{\sqrt{N}} - \frac{k^2x^2}{2N}\left[1 + r\left(\frac{kx}{\sqrt{N}}\right)\right]\right\}p(x)\,dx \\
&= \frac{1}{\sqrt{2\pi}}\left[1 - \frac{k^2}{2N}(1 + R_N)\right],
\end{aligned}$$

where

$$R_N = \int x^2 r\left(\frac{kx}{\sqrt{N}}\right)p(x)\,dx.$$

The integrand converges pointwise to zero as $N \to \infty$, and the dominated convergence theorem implies that $\lim_{N\to\infty} R_N = 0$. Computing the Nth power of $\hat{p}(k/\sqrt{N})$, and using (11.56)–(11.57), we obtain that

$$\lim_{N\to\infty}\hat{q}_N(k) = \lim_{N\to\infty}\frac{1}{\sqrt{2\pi}}\left[1 - \frac{k^2}{2N}(1 + R_N)\right]^N = \frac{1}{\sqrt{2\pi}}e^{-k^2/2}.$$

□

We can rescale the discrete random walk (11.50) to obtain a continuous-time stochastic process $W(t)$, called *Brownian motion*, or the *Wiener process*, that satisfies

$$W(t) = \lim_{N\to\infty}\frac{S_{Nt}}{\sqrt{N}}.$$

Here, we extend S_N to a function S_t of a continuous time variable t by supposing, for example, that the particle moves at a constant velocity from its position at time

N to its position at time $N+1$. This limit has to be interpreted in an appropriate probabilistic sense, which we will not make precise here.

As the central limit theorem suggests, Brownian motion $W(t)$ is a Gaussian process of mean 0 and variance t. Its sample paths are continuous, nowhere differentiable functions of time with probability one. The probability density $p(x,t)$ of finding the particle at position $W(t) = x$ at time t, assuming that $W(0) = 0$, is given by

$$p(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}.$$

The density p is the Green's function of the heat equation

$$p_t = \frac{1}{2} p_{xx}, \qquad p(x,0) = \delta(x).$$

Brownian motion is the simplest, and most fundamental, example of a *diffusion process*. These processes may also be described by *stochastic differential equations*, and they have widespread applications, from statistical physics to the modeling of financial markets.

11.13 References

See Hochstadt [22] for proofs and further discussion of the eigenfunctions of the Fourier transform. Distributions are discussed in Reed and Simon [45]. For an introduction to the theory of stochastic differential equations, see Øksendal [41].

11.14 Exercises

Exercise 11.1 Let X be a locally convex space. Prove the following.

(a) The addition of vectors in X and the multiplication by a scalar are continuous.
(b) A topology defined by a family of seminorms has a base of convex open neighborhoods. Such a topological space is called *locally convex*.
(c) If for all $x \in X$ there exists $\alpha \in \mathcal{A}$ such that $p_\alpha(x) > 0$, then the topology defined by $\{p_\alpha \mid \alpha \in \mathcal{A}\}$ is Hausdorff.

Exercise 11.2 Suppose that $\{p_1, p_2, p_3, \ldots\}$ is a countable family of seminorms on a linear space X. Prove that (11.4) defines a metric on X, and prove that metric topology defined by d coincides with the one defined by the family of seminorms $\{p_1, p_2, p_3, \ldots\}$.

Exercise 11.3 Let (x_n) be a sequence in a locally convex space whose topology is defined by a countably infinite set of seminorms. Prove that (x_n) is a Cauchy sequence for the metric d defined in (11.4) if and only if for every $\alpha \in \mathcal{A}$ and $\epsilon > 0$, there is an N such that $p_\alpha(x_n - x_m) < \epsilon$ for all $n, m \geq N$.

Exercise 11.4 If $\varphi \in \mathcal{S}(\mathbb{R})$, prove that

$$\varphi\delta' = \varphi(0)\delta' - \varphi'(0)\delta.$$

Exercise 11.5 Prove that

$$\lim_{\epsilon \to 0^+} \frac{1}{x + i\epsilon} = \text{p.v.}\frac{1}{x} - i\pi\delta(x) \qquad \text{in } \mathcal{S}^*(\mathbb{R}).$$

Exercise 11.6 Show that the distributional derivative of $\log|x| : \mathbb{R} \to \mathbb{R}$ is p.v.$1/x$.

Exercise 11.7 Show that there is no product $\cdot : \mathcal{S}^* \times \mathcal{S}^* \to \mathcal{S}^*$ on the space of tempered distributions that is commutative, associative, and agrees with the usual product of a tempered distribution and a smooth function of polynomial growth. HINT. Compute the product $x \cdot \delta(x) \cdot \text{p.v.}(1/x)$ in two different ways.

Exercise 11.8 Suppose that $\omega \in \mathcal{S}(\mathbb{R})$ is a test function such that

$$\int_{\mathbb{R}} \omega(x)\, dx = 1.$$

Show that every test function $\varphi \in \mathcal{S}(\mathbb{R})$ may be written as

$$\varphi(x) = \omega(x)\left(\int_{\mathbb{R}} \varphi(y)\, dy\right) + \psi'(x)$$

for some test function $\psi \in \mathcal{S}(\mathbb{R})$. Deduce that if T is a tempered distribution such that $T' = 0$, then T is constant.

Exercise 11.9 Let $k \in \mathcal{S}$ and define the convolution operator

$$Kf(x) = \int k(x - y)f(y)\, dy \qquad \text{for all } f \in \mathcal{S}.$$

Prove that $K : \mathcal{S} \to \mathcal{S}$ is a continuous linear operator for the topology of $\mathcal{S}$.

Exercise 11.10 For every $h \in \mathbb{R}^n$ define a linear transformation $\tau_h : \mathcal{S} \to \mathcal{S}$ by $\tau_h(f)(x) = f(x - h)$.

(a) Prove that for all $h \in \mathbb{R}^n$, τ_h is continuous in the topology of $\mathcal{S}$.
(b) Prove that for all $f \in \mathcal{S}$, the map $h \mapsto \tau_h f$ is continuous from $\mathbb{R}^n$ to $\mathcal{S}$.

HINT. For (b), prove that for $f \in C(\mathbb{R}^n)$ one has $\lim_{h\to 0} \|\tau_h f - f\|_\infty = 0$ if and only if f is uniformly continuous. Also, note that it is sufficient to prove continuity at $h = 0$, due to the group property of τ_h.

Exercise 11.11 The density ρ of an array of N point masses of mass $m_j > 0$ located at $x_j \in \mathbb{R}^n$ is a sum of δ functions

$$\rho(x) = \sum_{j=1}^{N} m_j \delta(x - x_j).$$

Compute the Fourier transform $\hat{\rho}$ of ρ. Show that for any $\varphi \in \mathcal{S}$, and for any $k_1, \ldots, k_N \in \mathbb{R}^n$, $z_1, \ldots, z_N \in \mathbb{C}$ we have

$$\int \overline{\varphi(k)} \hat{\rho}(k - \ell) \varphi(\ell) \, dk d\ell \geq 0, \qquad \sum_{p,q=1}^{N} \overline{z}_p \hat{\rho}\left(k_p - k_q\right) z_q \geq 0.$$

The Fourier transform $\hat{\rho}$ is said to be of *positive type*.

Exercise 11.12 Prove that if $s > n/2$, then $H^s(\mathbb{R}^n) \subset C_0(\mathbb{R}^n)$, and there is a constant C such that

$$\|f\|_\infty \leq C\|f\|_{H^s} \qquad \text{for all } f \in H^s(\mathbb{R}^n).$$

Exercise 11.13 Prove equations (11.27)–(11.29) for the Fourier transform of translates and convolutions. Prove the corresponding results for derivatives and translates of tempered distributions and for the convolution of a test function with a tempered distribution.

Exercise 11.14 The *Airy equation* is the ODE

$$u'' - xu = 0.$$

The solutions, called *Airy functions*, are the simplest functions that make a transition from oscillatory behavior (for $x < 0$) to exponential behavior (for $x > 0$). Take the Fourier transform, and deduce that

$$u(x) = c \int e^{ikx + ik^3/3} \, dk,$$

where c is an arbitrary constant. This nonconvergent integral is a simple example of an *oscillatory integral*. Here, it may be interpreted distributionally as an inverse Fourier transform. Why do you find only one linearly independent solution?

Exercise 11.15 Let $f_n : \mathbb{R} \to \mathbb{R}$ be the function

$$f_n(x) = \begin{cases} n^2 & \text{if } -1/n < x < 0, \\ -n^2 & \text{if } 0 < x < 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

Show that the sequence (f_n) converges in $\mathcal{S}^*(\mathbb{R})$ as $n \to \infty$, and determine its distributional limit.

Exercise 11.16 Let $f \in L^1(\mathbb{R}^3)$ be a rotationally invariant function in the sense that there is a function $g : \mathbb{R}^+ \to \mathbb{C}$ such that

$$f(x) = g(|x|).$$

Prove that the Fourier transform of f is a continuous function $\hat{f}$ that is also rotation invariant, and $\hat{f}(k) = h(|k|)$, where

$$h(k) = \frac{1}{k}\sqrt{\frac{2}{\pi}}\int_0^\infty r\sin(kr)g(r)\,dr.$$

Exercise 11.17 Show that

$$A^{-1} = \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} x \otimes x \exp\left(-\frac{1}{2}x \cdot Ax\right)\,dx.$$

We therefore call A^{-1} the *covariance matrix* of the n-dimensional Gaussian probability distribution with density $(2\pi)^{-n/2}\exp(-x \cdot Ax/2)$.

Exercise 11.18 Prove that if $g \in L^2$ satisfies $g(-x) = \overline{g(x)}$, then $\hat{g}$ is real-valued.

Exercise 11.19 Give a counterexample to show that the Riemann-Lebesgue lemma does not hold for all functions in L^2. That is, find a function $f \in L^2(\mathbb{R})$ such that $\hat{f}$ is not continuous.

Exercise 11.20 Show that $\delta \in H^s(\mathbb{R}^n)$ if and only if $s < -n/2$.

Exercise 11.21 Show that the integral equation

$$u(x) + \int_{-\infty}^\infty e^{-(x-y)^2/2}u(y)\,dy = f(x)$$

has a unique solution $u \in L^2(\mathbb{R})$ for every $f \in L^2(\mathbb{R})$, and give an expression for u in terms of f.

Exercise 11.22 Show that

$$\mathcal{F}^{-1}[i\operatorname{sgn} k] = -\frac{2}{\sqrt{\pi}}\left(\text{p.v.}\frac{1}{x}\right).$$

Exercise 11.23 Show that the solution of the heat equation on a one-dimensional semi-infinite rod,

$$\begin{aligned} u_t &= \frac{1}{2}u_{xx} && 0 < x < \infty,\ t > 0, \\ u(0,t) &= 0 && t > 0, \\ u(x,0) &= f(x) && 0 < x < \infty, \end{aligned}$$

is given by

$$u(x,t) = \int_0^\infty \frac{e^{-(x-y)^2/(2t)} - e^{-(x+y)^2/(2t)}}{\sqrt{2\pi t}} u_0(y)\, dy.$$

This solution illustrates the *method of images.*

Exercise 11.24 Let

$$f(t) = \begin{cases} \exp(-1/t^2) & \text{if } t > 0, \\ 0 & \text{if } t \le 0. \end{cases}$$

Show that

$$u(x,t) = \sum_{n=0}^{\infty} \frac{f^{(n)}(t)}{(2n)!} x^{2n}$$

is a nonzero solution of the one-dimensional heat equation $u_t = u_{xx}$ with zero initial data $u(x,0) = 0$.

Exercise 11.25 Find the Green's function $g(x,t)$ of the one-dimensional wave equation,

$$g_{tt} - g_{xx} = 0,$$
$$g(x,0) = 0, \quad g_t(x,0) = \delta(x).$$

Exercise 11.26 Consider the wave equation

$$u_{tt} = \Delta u$$

for $u(x,t)$, where $t > 0$ and $x \in \mathbb{R}^n$, with initial data

$$u(x,0) = 0, \quad u_t(x,0) = v_0(x).$$

For simplicity, assume that $v_0 \in \mathcal{S}(\mathbb{R}^n)$. Find the equation satsified by the Fourier transform of u,

$$\hat{u}(k,t) = \frac{1}{(2\pi)^{n/2}} \int e^{-ik\cdot x} u(x,t)\, dx,$$

and show that

$$u(x,t) = \frac{1}{(2\pi)^{n/2}} \int \frac{\sin(|k|t)}{|k|} e^{ik\cdot x} \hat{v}_0(k)\, dk,$$

where $\hat{v}_0$ is the Fourier transform of v_0 and $|\cdot|$ denotes the Euclidean norm.

For $n = 3$, let $d\Omega_t$ denote the surface integration measure on the sphere of radius t, so $\int_{|x|=t} d\Omega_t = 4\pi t^2$. Prove that the solution can be written as

$$u(x,t) = \frac{1}{4\pi t} \int_{|y|=t} v_0(x+y)\, d\Omega_t.$$

For $n = 2$, prove that the solution is

$$u(x,t) = \frac{1}{2\pi} \int_{|y|\le t} \frac{v_0(x+y)}{\sqrt{t^2 - |y|^2}}\, dy.$$

Interpret these formulae physically.

Exercise 11.27 Prove (11.45).

Exercise 11.28 Prove the following identity for all $a > 0$:

$$\sum_{n=-\infty}^{\infty} \frac{1}{n^2 + a^2} = \frac{\pi}{a} \frac{1 + e^{-2\pi a}}{1 - e^{-2\pi a}}.$$

By consideration of the limit $a \to 0^+$, show that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Exercise 11.29 We define the *Wigner distribution* $W(x,k)$ of a Schwartz function $\varphi(x)$, where $x, k \in \mathbb{R}^n$, by

$$W(x,k) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \varphi\left(x - \frac{y}{2}\right) \overline{\varphi\left(x + \frac{y}{2}\right)} e^{ik\cdot y}\, dy.$$

Compute the Wigner distribution of a Gaussian $\exp(-x \cdot Ax)$, where A is a positive definite matrix. Show that W is real-valued, and

$$W(x,k) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{\varphi}\left(k - \frac{\ell}{2}\right) \overline{\hat{\varphi}\left(k + \frac{\ell}{2}\right)} e^{-i\ell\cdot x}\, d\ell,$$

$$\int_{\mathbb{R}^n} W(x,k)\, dk = |\varphi(x)|^2, \qquad \int_{\mathbb{R}^n} W(x,k)\, dx = |\hat{\varphi}(k)|^2.$$

Thus, the Wigner distribution W has some properties of a phase space (that is, an (x,k)-space) density of φ. Show, however, that W is not necessarily nonnegative.

Exercise 11.30 Let $\varphi : \mathbb{R} \to \mathbb{C}$ be any Schwartz function such that

$$\int_{-\infty}^{\infty} |\varphi(x)|^2\, dx = 1,$$

and define

$$E_x = \int_{-\infty}^{\infty} x^2 |\varphi(x)|^2\, dx, \qquad E_k = \int_{-\infty}^{\infty} k^2 |\hat{\varphi}(k)|^2\, dk.$$

Prove the *Heisenberg uncertainty principle*:

$$E_x E_k \ge \frac{1}{4}.$$

Show that equality is attained when φ is a suitable Gaussian.