Fourier analysis

Lecture notes, Fall 2024

Mikko Salo

Department of Mathematics and Statistics
University of Jyväskylä

Contents

CHAPTER 1

Introduction

Joseph Fourier laid the foundations of the mathematical field now known as Fourier analysis in his 1822 treatise on heat flow, although related ideas were used before by Bernoulli, Euler, Gauss and Lagrange. The basic question is to represent periodic functions as sums of elementary pieces. If $f : \mathbb{R} \to \mathbb{C}$ has period 2π and the elementary pieces are sine and cosine functions, then the desired representation would be a *Fourier series*

$$f(x) = \sum_{k=0}^{\infty} (a_k \cos(kx) + b_k \sin(kx)). \tag{1.1}$$

Since $e^{ikx} = \cos(kx) + i\sin(kx)$, we may alternatively consider the series

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}. \tag{1.2}$$

The bold claim of Fourier was that *every* function has such a representation. It turns out that this is true in a sense not just for functions, but even for a large class of *generalized functions* or *distributions* (this includes all reasonable measures and more).

Integrating (1.2) against e^{-ilx} (assuming this is justified), we see that the coefficients c_l should be given by $c_l = \hat{f}(l)$ where

$$\hat{f}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx}\, dx. \tag{1.3}$$

Then (1.2) can be rewritten as

$$f(x) = \sum_{k=-\infty}^{\infty} \hat{f}(k) e^{ikx}. \tag{1.4}$$

These formulas can be thought of as an *analysis – synthesis* pair: (1.4) synthesizes f as a sum of exponentials e^{ikx}, whereas (1.3) analyzes f to obtain the coefficients $\hat{f}(k)$ that describe how much of the exponential e^{ikx} is contained in f.

Fourier analysis can also be performed in nonperiodic settings, replacing the 2π-periodic functions $\{e^{ikx}\}_{k\in\mathbb{Z}}$ by exponentials $\{e^{i\omega t}\}_{\omega\in\mathbb{R}}$. Suppose that $f : \mathbb{R} \to \mathbb{C}$ is a reasonably nice function. The *Fourier transform* of f is the function

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\,dt, \tag{1.5}$$

and the function f then has the Fourier representation

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega t}\,d\omega. \tag{1.6}$$

Thus, f may be recovered from its Fourier transform $\hat{f}$ by taking the *inverse Fourier transform* as in (1.6). This is a similar analysis – synthesis pair as for Fourier series, and if $f(t)$ is an audio signal (for instance a music clip), then (1.6) gives the frequency representation of the signal: f is written as the integral (=continuous sum) of the exponentials $e^{i\omega t}$ vibrating at frequency ω, and $\hat{f}(\omega)$ describes how much of the frequency ω is contained in the signal.

The extension of the above ideas to higher dimensional cases is straightforward. The Fourier transform and inverse Fourier transform formulas for functions $f : \mathbb{R}^n \to \mathbb{C}$ are given by

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

$$f(x) = (2\pi)^{-n}\int_{\mathbb{R}^n} \hat{f}(\xi)e^{ix\cdot\xi}\,d\xi, \quad x \in \mathbb{R}^n.$$

Like in the case of Fourier series, also the Fourier transform can be defined on a large class of generalized functions (the space of *tempered distributions*), which gives rise to the very useful weak theory of the Fourier transform.

REMARK. There are many conventions on where to put the factors of 2π in the definition of Fourier transform, and they all have their benefits and disadvantages. In this course we will follow the conventions given above. This will be useful in applications to partial differential equations, since no factors of 2π appear when taking Fourier transforms of derivatives.

Let us next give some elementary examples of the above concepts. More substantial applications of Fourier analysis to different parts of mathematics will be covered later in the course.

EXAMPLE 1. (Heat equation) Consider a homogeneous circular metal ring $\{(\cos(x), \sin(x))\,;\, x \in [-\pi, \pi]\}$, which we identify with the interval $[-\pi, \pi]$ on the real line. Denote by $u(x,t)$ the temperature of the ring at point x at time t. If the initial temperature is $f(x)$, then the temperature at time t is obtained by solving the *heat equation*

$$\begin{aligned} \partial_t u(x,t) - \partial_x^2 u(x,t) &= 0 \qquad && \text{in } [-\pi,\pi] \times \{t > 0\}, \\ u(x,0) &= f(x) && \text{for } x \in [-\pi,\pi]. \end{aligned}$$

Since the medium is a ring, the equation actually includes the boundary conditions $u(-\pi,t) = u(\pi,t)$ and $\partial_x u(-\pi,t) = \partial_x u(\pi,t)$ for $t > 0$. We write f as the Fourier series (1.2), and try to find a solution in the form

$$u(x,t) = \sum_{k=-\infty}^{\infty} u_k(t) e^{ikx}.$$

Inserting these expressions in the equation (and assuming that everything converges nicely), we get

$$\begin{aligned} \sum_{k=-\infty}^{\infty} (u_k'(t) + k^2 u_k(t)) e^{ikx} &= 0, \\ \sum_{k=-\infty}^{\infty} u_k(0) e^{ikx} &= \sum_{k=-\infty}^{\infty} c_k e^{ikx}. \end{aligned}$$

Equating the e^{ikx} parts leads to the ODEs

$$\begin{aligned} u_k'(t) + k^2 u_k(t) &= 0, \\ u_k(0) &= c_k. \end{aligned}$$

Solving these gives $u_k(t) = c_k e^{-k^2 t}$, so the temperature distribution $u(x,t)$ of the metal ring is given by

$$u(x,t) = \sum_{k=-\infty}^{\infty} c_k e^{-k^2 t} e^{ikx}.$$

EXAMPLE 2. (Audio filtering) The 2010 FIFA World Cup in football took place in South Africa and introduced TV viewers around the world to the *vuvuzela*, a traditional musical horn that was played by thousands of spectators at the games. This sometimes drowned out the voices of TV commentators, which prompted the development of vuvuzela filters. The main frequency components of vuvuzela noise are at ~ 235 Hz and ~ 470 Hz, and in principle the noise could be removed by replacing the original audio signal $f(t)$, with Fourier representation (1.6), by its filtered version

$$f_{\text{filtered}}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(\omega) \hat{f}(\omega) e^{i\omega t} \, dt$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is a cutoff function that vanishes around 235 and 470 Hz and is equal to one elsewhere.

Example 3. (Measuring temperature) Suppose that $u(x)$ is the temperature at point x in a room, and one wants to measure the temperature by a thermometer. The bulb of the thermometer is not a single point but rather has cylindrical shape, and one can think that the thermometer measures a weighted average of the temperature near the bulb. Thus, one measures

$$\int u(x)\varphi(x) \, dx$$

where φ is a function determined by the shape and properties of the thermometer and φ is concentrated near the bulb. If one has two different thermometers, the measured temperatures could be given by

$$\int u(x)\varphi_1(x) \, dx, \qquad \int u(x)\varphi_2(x) \, dx.$$

Thus, temperature measurements can be thought to arise from "testing" the temperature distribution $u(x)$ by different functions $\varphi(x)$. This is the main idea behind *distribution theory*: instead of thinking of functions in terms of pointwise values, one thinks of functions as objects that are tested against test functions. The same idea makes it possible to consider objects that are much more general than functions.

In this course we mostly concern ourselves with the weak and L^2 theory of Fourier series and transforms, together with the relevant distribution spaces, with an emphasis on aspects related to partial differential equations. We also give a number of applications. There are many other possible topics for a course on Fourier analysis, including the following:

L^p **harmonic analysis.** The terms *Fourier analysis* and *harmonic analysis* may be considered roughly synonymous. Harmonic analysis is concerned with expansions of functions in terms of "harmonics", which can be complex exponentials or other similar objects (like spherical harmonics on the sphere, or eigenfunctions of the Laplace operator on Riemannian manifolds). One is often interested in estimates for related operators in L^p norms. A representative question is the *Fourier restriction conjecture* (posed by Stein in the 1960's): one version asks whether for any $q > \frac{2n}{n-1}$ there is $C > 0$ such that

$$\|\widehat{f\,dS}\|_{L^q(\mathbb{R}^n)} \leq C\|f\|_{L^\infty(S^{n-1})}, \qquad f \in L^\infty(S^{n-1}),$$

where

$$\widehat{f\,dS}(\xi) = \int_{S^{n-1}} f(x)e^{-ix\cdot\xi}\,dS(x), \qquad \xi \in \mathbb{R}^n.$$

The theory of singular integrals and Calderón-Zygmund operators are closely related topics.

Time-frequency analysis. The usual Fourier transform on the real line is not optimal for many signal processing purposes: while it provides perfect frequency localization (the number $\hat{f}(\omega)$ describes how much of the exponential $e^{i\omega t}$ vibrating *exactly* at frequency ω is contained in the signal), there is no time localization (the evaluation of $\hat{f}(\omega)$ requires integrating f over all times). Often one is interested in the content of the signal over short time periods, and then it is more appropriate to use *windowed Fourier transforms* that involve a cutoff function in time and represent a tradeoff between time and frequency localization.

A closely related concept is the *continuous wavelet transform*, which decomposes a signal $f(t)$ as

$$f(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Tf(a,b)\psi^{a,b}(t)\,db\,\frac{da}{a^2},$$

where the wavelet coefficients are given by

$$Tf(a,b) = \int_{-\infty}^{\infty} f(t)\psi^{a,b}(t)\,dt.$$

Here $\psi^{a,b}$ is a function living near time b at scale a,

$$\psi^{a,b}(t) = |a|^{-1/2}\psi(\frac{t-b}{a}),$$

and ψ is a suitable compactly supported function (so called *mother wavelet*) whose graph might look like a Mexican hat. Transforms of this type are central in signal and image processing (for instance JPEG compression) and they are of great theoretical value as well, providing characterizations of many function spaces.

Another related topic is *microlocal analysis*, where one tries to study functions in space and frequency variables simultaneously. This viewpoint, together with the machinery of pseudodifferential and Fourier integral operators, is central in the modern theory of partial differential equations and constitutes a kind of "variable coefficient" Fourier analysis.

Abstract harmonic analysis. Fourier analysis can be performed on locally compact topological groups. The theory is the most complete on locally compact *abelian* groups. If G is such a group, there is a unique (up to scalar multiple) translation invariant measure called *Haar measure*, and a corresponding space $L^1(G)$. The Fourier transform of $f \in L^1(G)$ is a function acting on $\hat{G}$, the *Pontryagin dual group* of G. This is the set of *characters* of G, that is, continuous homomorphisms

$$\chi : G \to S^1, \quad \chi(x+y) = \chi(x)\chi(y).$$

If $G = \mathbb{R}^n$ the continuous homomorphisms are given by $\chi(x) = e^{ix\cdot\xi}$ for $\xi \in \mathbb{R}^n$, whereas if $G = \mathbb{R}/2\pi\mathbb{Z}$ they are given by $\chi(x) = e^{ik\cdot x}$ for $k \in \mathbb{Z}$. There is also an L^2 theory for the Fourier transform, and some aspects extend to *compact non-abelian* groups.

References. As references for Fourier analysis and distribution theory, the following textbooks are useful (some parts of the course will follow parts of these books). They are roughly in ascending order of difficulty:

- E. Stein and R. Sharkarchi: Fourier analysis.
- R. Strichartz: A guide to distribution theory.
- W. Rudin: Functional analysis.
- L. Schwartz: Théorie des distributions.
- J. Duoandikoetxea: Fourier analysis.
- L. Hörmander: The analysis of linear partial differential operators, vol. I.

Notation

We will write $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}$ for the real numbers, complex numbers, and the integers, respectively. $\mathbb{R}_+$ will be the set of positive real numbers and $\mathbb{Z}_+$ the set of positive integers, with $\mathbb{N} = \mathbb{Z}_+ \cup \{0\}$ the set of natural numbers. For vectors x in $\mathbb{R}^n$ the expression $|x|$ denotes the Euclidean length, while for vectors k in $\mathbb{Z}^n$ we write $|k| = \sum_{i=1}^n |k_i|$. We will also use the notation $\langle x \rangle = (1 + |x|^2)^{1/2}$.

To facilitate discussion of functions in several variables the multi-index notation is used. The set of multi-indices is denoted by $\mathbb{N}^n$ and it consists of all n-tuples $\alpha = (\alpha_1, \ldots, \alpha_n)$ where the α_i are nonnegative integers. We write $|\alpha| = \alpha_1 + \ldots + \alpha_n$ and $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$.

For partial derivatives, the notation

$$\partial^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}$$

will be used. We will also write $D_j = \frac{1}{i}\frac{\partial}{\partial x_j}$, and correspondingly

$$D^\alpha = D_1^{\alpha_1} \cdots D_n^{\alpha_n}.$$

The Laplacian in $\mathbb{R}^n$ is defined as

$$\Delta = \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2}.$$

If Ω is an open set in $\mathbb{R}^n$ then $C^k(\Omega)$ will be the space of those complex functions f in Ω for which $\partial^\alpha f$ is continuous for $|\alpha| \leq k$. Of course $C^\infty(\Omega)$ is the space of infinitely differentiable functions on Ω.

CHAPTER 2

Fourier series

We wish to represent functions of n variables as Fourier series. If f is a function in $\mathbb{R}^n$ which is 2π-periodic in each variable, then a natural multidimensional analogue of (1.2) would be

$$f(x) = \sum_{k \in \mathbb{Z}^n} c_k e^{ik \cdot x}.$$

This is the form of Fourier series which we will study. Note that the terms on the right-hand side are 2π-periodic in each variable.

There are many subtle issues related to various modes of convergence for the series above. We will discuss L^2 convergence and pointwise convergence. In the end of the chapter we will consider a number of applications of Fourier series.

2.1. Fourier series in L^2

The convergence in L^2 norm for Fourier series of L^2 functions is a straightforward consequence of Hilbert space theory. Consider the cube $Q = [-\pi, \pi]^n$, and define an inner product on $L^2(Q)$ by

$$(f, g) = (2\pi)^{-n} \int_Q f \bar{g}\, dx, \quad f, g \in L^2(Q).$$

With this inner product, $L^2(Q)$ is a separable infinite-dimensional Hilbert space. Recall that this means that

- $(\,\cdot\,,\,\cdot\,)$ is an inner product on $L^2(Q)$ with norm $\|u\| = (u, u)^{1/2}$,
- all Cauchy sequences converge (Riesz-Fischer theorem),
- there is a countable dense subset (this follows by looking at simple functions with rational coefficients, or from Lemma 2.1.2 below).

The space of functions which are locally square integrable and 2π-periodic in each variable may be identified with $L^2(Q)$. Therefore, we will consider Fourier series of functions in $L^2(Q)$.

LEMMA 2.1.1. *The set $\{e^{ik \cdot x}\}_{k \in \mathbb{Z}^n}$ is an orthonormal subset of $L^2(Q)$.*

PROOF. A direct computation: if $k, l \in \mathbb{Z}^n$ then

$$\begin{aligned}
(e^{ik\cdot x}, e^{il\cdot x}) &= (2\pi)^{-n} \int_Q e^{i(k-l)\cdot x}\,dx \\
&= (2\pi)^{-n} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{i(k_1-l_1)x_1} \cdots e^{i(k_n-l_n)x_n}\,dx_n \cdots dx_1 \\
&= \begin{cases} 1, & k = l, \\ 0, & k \neq l. \end{cases}
\end{aligned}$$

□

We recall a Hilbert space fact. If $\{e_j\}_{j=1}^\infty$ is an orthonormal subset of a separable Hilbert space H, then the following are equivalent:

(1) $\{e_j\}_{j=1}^\infty$ is an orthonormal basis, in the sense that any $f \in H$ may be written as the series

$$f = \sum_{j=1}^\infty (f, e_j) e_j$$

with convergence in H,

(2) for any $f \in H$ one has

$$\|f\|^2 = \sum_{j=1}^\infty |(f, e_j)|^2,$$

(3) if $f \in H$ and $(f, e_j) = 0$ for all j, then $f \equiv 0$.

If any of these conditions is satisfied, the orthonormal system $\{e_j\}$ is called *complete*. The main point is that $\{e^{ik\cdot x}\}_{k\in\mathbb{Z}^n}$ is complete in $L^2(Q)$.

LEMMA 2.1.2. *If $f \in L^2(Q)$ satisfies $(f, e^{ik\cdot x}) = 0$ for all $k \in \mathbb{Z}^n$, then $f \equiv 0$.*

The proof is given below. The main result on Fourier series of L^2 functions is now immediate. Below we denote by $\ell^2(\mathbb{Z}^n)$ the space of complex sequences $c = (c_k)_{k\in\mathbb{Z}^n}$ with norm

$$\|c\|_{\ell^2(\mathbb{Z}^n)} = \Big(\sum_{k\in\mathbb{Z}^n} |c_k|^2 \Big)^{1/2}.$$

THEOREM 2.1.3. *(Fourier series of L^2 functions) If $f \in L^2(Q)$, then one has the Fourier series*

$$f(x) = \sum_{k\in\mathbb{Z}^n} \hat{f}(k) e^{ik\cdot x}$$

with convergence in $L^2(Q)$, where the Fourier coefficients are given by

$$\hat{f}(k) = (f, e^{ik\cdot x}) = (2\pi)^{-n} \int_Q f(x) e^{-ik\cdot x}\, dx.$$

One has the Parseval identity

$$\|f\|_{L^2(Q)}^2 = \sum_{k\in\mathbb{Z}^n} |\hat{f}(k)|^2.$$

Conversely, if $c = (c_k) \in \ell^2(\mathbb{Z}^n)$, then the series

$$f(x) = \sum_{k\in\mathbb{Z}^n} c_k e^{ik\cdot x}$$

converges in $L^2(Q)$ to a function f satisfying $\hat{f}(k) = c_k$.

Proof. The facts on the Fourier series of $f \in L^2(Q)$ follow directly from the discussion above, since $\{e^{ik\cdot x}\}_{k\in\mathbb{Z}^n}$ is a complete orthonormal system in $L^2(Q)$. For the converse, if $(c_k) \in \ell^2(\mathbb{Z}^n)$, then

$$\| \sum_{\substack{k\in\mathbb{Z}^n \\ M\leq|k|\leq N}} c_k e^{ik\cdot x} \|_{L^2(Q)}^2 = \sum_{\substack{k\in\mathbb{Z}^n \\ M\leq|k|\leq N}} |c_k|^2$$

by orthogonality. Since the right hand side can be made arbitrarily small by choosing M and N large, we see that $f_N = \sum_{k\in\mathbb{Z}^n,|k|\leq N} c_k e^{ik\cdot x}$ is a Cauchy sequence in $L^2(Q)$, and converges to $f \in L^2(Q)$. One obtains $\hat{f}(k) = (f, e^{ik\cdot x}) = c_k$ again by orthogonality. □

It remains to prove Lemma 2.1.2. We begin with the most familiar case, $n = 1$. It is useful to introduce the following notion.

Definition. A sequence $(K_N(x))_{N=1}^\infty$ of 2π-periodic continuous functions on the real line is called an *approximate identity* if

(1) $K_N \geq 0$ for all N,
(2) $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_N(x)\, dx = 1$ for all N, and
(3) for all $\delta > 0$ one has

$$\lim_{N\to\infty} \sup_{\delta\leq|x|\leq\pi} K_N(x) = 0.$$

Thus, an approximate identity (K_N) for large N resembles a Dirac mass at 0, extended in a 2π-periodic way. We now show that there is an approximate identity consisting of trigonometric polynomials.

LEMMA 2.1.4. *The sequence*

$$Q_N(x) = c_N \left(\frac{1 + \cos x}{2} \right)^N,$$

where $c_N = 2\pi \left(\int_{-\pi}^{\pi} \left(\frac{1+\cos x}{2}\right)^N \, dx \right)^{-1}$, *is an approximate identity.*

PROOF. (1) and (2) are clear. To show (3), we first estimate c_N by

$$\begin{aligned} 1 &= \frac{c_N}{\pi} \int_0^{\pi} \left(\frac{1 + \cos x}{2} \right)^N \, dx \geq \frac{c_N}{\pi} \int_0^{\pi} \left(\frac{1 + \cos x}{2} \right)^N \sin x \, dx \\ &= \frac{c_N}{\pi} \int_{-1}^{1} \left(\frac{1 + t}{2} \right)^N \, dt = \frac{2c_N}{\pi} \int_0^1 s^N \, ds = \frac{2c_N}{\pi(N+1)}. \end{aligned}$$

Then for $\delta < |x| < \pi$, we have

$$Q_N(x) \leq Q_N(\delta) = c_N \left(\frac{1 + \cos \delta}{2} \right)^N \leq \frac{\pi(N+1)}{2} \left(\frac{1 + \cos \delta}{2} \right)^N$$

which shows (3) since $(1 + \cos \delta)/2 < 1$ for all $\delta > 0$. □

It is possible to approximate L^p functions by convolving them against an approximate identity. Here, the *convolution* of two 2π-periodic functions is defined as the 2π-periodic function

$$(f * g)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) g(x - y) \, dy.$$

This integral is well defined for a.e. x if one of the functions is in L^1 and the other in L^∞, or more generally if $f, g \in L^1$ by Fubini's theorem. We define the L^p norm by

$$\|f\|_{L^p([-\pi,\pi])} = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^p \, dx \right)^{1/p}$$

LEMMA 2.1.5. *Let* (K_N) *be an approximate identity. If* $f \in L^p([-\pi, \pi])$ *where* $1 \leq p < \infty$, *or if* f *is a continuous* 2π*-periodic function and* $p = \infty$, *then*

$$\|K_N * f - f\|_{L^p([-\pi,\pi])} \to 0 \quad \text{as } N \to \infty.$$

PROOF. Since $\frac{1}{2\pi} K_N$ has integral 1, we have

$$(K_N * f)(x) - f(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(y) [f(x - y) - f(x)] \, dy.$$

Let first f be continuous and $p = \infty$. To estimate the L^∞ norm of $K_N * f - f$, we fix $\varepsilon > 0$ and compute

$$|(K_N * f)(x) - f(x)| \le \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(y) |f(x-y) - f(x)|\, dy$$
$$\le \frac{1}{2\pi} \int_{|y|\le\delta} K_N(y)|f(x-y) - f(x)|\, dy + \frac{1}{2\pi} \int_{\delta\le|y|\le\pi} K_N(y)|f(x-y)-f(x)|\, dy.$$

Here $\delta > 0$ is chosen so that

$$|f(x-y) - f(x)| < \frac{\varepsilon}{2} \quad \text{whenever } x \in \mathbb{R} \text{ and } |y| \le \delta.$$

This is possible because f is uniformly continuous. Further, we use the definition of an approximate identity and choose N_0 so that

$$\sup_{\delta\le|x|\le\pi} K_N(x) < \frac{\pi\varepsilon}{2\|f\|_{L^\infty}}, \quad \text{for } N \ge N_0.$$

With these choices, we obtain

$$|(K_N * f)(x) - f(x)| \le \frac{\varepsilon}{4\pi} \int_{|y|\le\delta} K_N(y)\, dy + \frac{\|f\|_{L^\infty}}{\pi} \sup_{\delta\le|x|\le\pi} K_N(x) < \varepsilon$$

whenever $N \ge N_0$. The result is proved in the case $p = \infty$.

Let now $f \in L^p([-\pi,\pi])$ and $1 \le p < \infty$. We will use the integral form of Minkowski's inequality,

$$\Big(\int_X \Big| \int_Y F(x,y)\, d\nu(y) \Big|^p d\mu(x) \Big)^{1/p} \le \int_Y \Big(\int_X |F(x,y)|^p\, d\mu(x) \Big)^{1/p} d\nu(y),$$

which is valid on σ-finite measure spaces (X,μ) and (Y,ν), cf. the usual Minkowski inequality $\|\sum_y F(\,\cdot\,,y)\|_{L^p} \le \sum_y \|F(\,\cdot\,,y)\|_{L^p}$. Using this, we obtain

$$2\pi\|K_N * f - f\|_{L^p([-\pi,\pi])} \le \int_{-\pi}^{\pi} K_N(y) \|f(\,\cdot\, - y) - f\|_{L^p([-\pi,\pi])}\, dy$$
$$= \int_{\delta\le|y|\le\pi} K_N(y)\|f(\,\cdot\,-y) - f\|_{L^p}\, dy + \int_{|y|\le\delta} K_N(y)\|f(\,\cdot\,-y)-f\|_{L^p}\, dy$$
$$\le 2\|f\|_{L^p} \sup_{\delta\le|x|\le\pi} K_N(x) + 2\pi \sup_{|y|\le\delta} \|f(\,\cdot\, - y) - f\|_{L^p}.$$

Since translation is a continuous operation on L^p spaces, for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$\|f(\,\cdot\, - y) - f\|_{L^p([-\pi,\pi])} < \varepsilon \quad \text{whenever } |y| \le \delta.$$[1]

Thus the second term can be made arbitrarily small by choosing δ sufficiently small, and then the first term is also small if N is large. This shows the result. □

As a side product of the above results, we get the following version of the Weierstrass approximation theorem for periodic functions.

THEOREM 2.1.6. *If f is a continuous 2π-periodic function, then for any $\varepsilon > 0$ there is a trigonometric polynomial P such that*

$$\|f - P\|_{L^\infty(\mathbb{R})} < \varepsilon.$$

PROOF. It is enough to choose $P = Q_N * f$ for N large and use Lemma 2.1.5 for $p = \infty$. □

We can now finish the proof of the basic facts on Fourier series of L^2 functions.

PROOF OF LEMMA 2.1.2. Let first $n = 1$. If $f \in L^2([-\pi,\pi])$ and $(f, e^{ikx}) = 0$ for all $k \in \mathbb{Z}$, then the inner product of f against any trigonometric polynomial vanishes. Thus, for any x,

$$0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) Q_N(x-y)\,dy = (Q_N * f)(x)$$

Lemmas 2.1.4 and 2.1.5 imply that $\lim_{N\to\infty} Q_N * f = f$ in the L^2 sense, so $f \equiv 0$ as required.

Now let $n \ge 2$, and assume that $f \in L^2(Q)$ and $(f, e^{ik\cdot x}) = 0$ for all $k \in \mathbb{Z}^n$. Since $e^{ik\cdot x} = e^{ik_1x_1}\cdots e^{ik_nx_n}$, we have

$$\int_{-\pi}^{\pi} h(x_1; k_2, \ldots, k_n) e^{-ik_1x_1}\,dx_1 = 0$$

for all $k_1 \in \mathbb{Z}$, where

$$h(x_1; k_2, \ldots, k_n) = \int_{[-\pi,\pi]^{n-1}} f(x_1, x_2, \ldots, x_n) e^{-i(k_2x_2+\ldots+k_nx_n)}\,dx_2\cdots dx_n.$$

[1] To see this, use Lusin's theorem to find $g \in C_c((-\pi,\pi))$ with $\|f-g\|_{L^p} < \varepsilon/3$. Extend f and g in a 2π-periodic way, and note that

$$\|f(\,\cdot\, - y) - f\|_{L^p} \le \|f(\,\cdot\, - y) - g(\,\cdot\, - y)\|_{L^p} + \|g(\,\cdot\, - y) - g\|_{L^p} + \|g - f\|_{L^p}.$$

The first and third terms are $< \varepsilon/3$, and so is the second term by uniform continuity if $|y| < \delta$ for δ small enough.

Now $h(\,\cdot\,; k_2, \ldots, k_n)$ is in $L^2([-\pi, \pi])$ by Cauchy-Schwarz inequality. By the completeness of the system $\{e^{ik_1x_1}\}$ in one dimension, we obtain that $h(\,\cdot\,; k_2, \ldots, k_n) = 0$ for all $k_2, \ldots, k_n \in \mathbb{Z}$. Applying the same argument in the other variables shows that $f \equiv 0$. □

2.2. Pointwise convergence

Although pointwise convergence of Fourier series is not the main topic of this course, it may be of interest to mention a few classical results. We will focus on the case $n = 1$. Note that the Fourier coefficients

$$\hat{f}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx}\, dx, \qquad k \in \mathbb{Z},$$

are well defined for any $f \in L^1([-\pi, \pi])$, and

$$|\hat{f}(k)| \leq \|f\|_{L^1}, \qquad k \in \mathbb{Z}.$$

The partial sums of the Fourier series of a function $f \in L^1([-\pi, \pi])$, extended as a 2π-periodic function into $\mathbb{R}$, are given by

$$\begin{aligned} S_m f(x) = \sum_{k=-m}^{m} \hat{f}(k) e^{ikx} &= \sum_{k=-m}^{m} \Big(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) e^{-iky}\, dy \Big) e^{ikx} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} D_m(x-y) f(y)\, dy \end{aligned}$$

where $D_m(x)$ is the *Dirichlet kernel*

$$\begin{aligned} D_m(x) &= \sum_{k=-m}^{m} e^{ikx} = e^{-imx}(1 + e^{ix} + \ldots + e^{i2mx}) \\ &= e^{-imx} \frac{e^{i(2m+1)x} - 1}{e^{ix} - 1} = \frac{e^{i(m+\frac{1}{2})x} - e^{-i(m+\frac{1}{2})x}}{e^{i\frac{1}{2}x} - e^{-i\frac{1}{2}x}} = \frac{\sin((m+\frac{1}{2})x)}{\sin(\frac{1}{2}x)}. \end{aligned}$$

Thus partial sums of the Fourier series of f are given by convolution against the Dirichlet kernel,

$$S_m f(x) = (D_m * f)(x).$$

One might expect that that the Dirichlet kernel would behave like an approximate identity, which would imply that the partial sums $S_m f = D_m * f$ would converge to f uniformly if f is continuous. However, D_m is *not* an approximate identity in the sense of the earlier definition because it takes both positive and negative values. In fact, one has $\frac{1}{2\pi} \int_{-\pi}^{\pi} D_m(x)\, dx = 1$, but $\int_{-\pi}^{\pi} |D_m(x)|\, dx \to \infty$ as $m \to \infty$.

Thus the convergence of the partial sums $S_m f$ to f may depend on the oscillation (cancellation between positive and negative values) of the Dirichlet kernel. This makes the pointwise convergence of Fourier series somewhat subtle, and in fact there exist continuous functions whose Fourier series *diverge at uncountably many points*.

By assuming something slightly stronger than continuity, pointwise convergence holds:

THEOREM 2.2.1. *(Dini's criterion) If $f \in L^1([-\pi,\pi])$ and if x is a point such that for some $\delta > 0$*

$$\int_{|y|<\delta} \left| \frac{f(x+y) - f(x)}{y} \right| dy < \infty,$$

then $S_m f(x) \to f(x)$ as $m \to \infty$.

Note that if f is Hölder continuous near x, so that for some $\alpha > 0$

$$|f(x) - f(y)| \leq C|x-y|^\alpha \quad \text{for } y \text{ near } x,$$

then f satisfies the above condition at x. Note also that the condition is local: the behavior away from x will not affect the convergence of the Fourier series at x. This general phenomenon is illustrated by the following result.

THEOREM 2.2.2. *(Riemann localization principle) If $f \in L^1([-\pi,\pi])$ satisfies $f \equiv 0$ near x, then*

$$\lim_{m\to\infty} S_m f(x) = 0.$$

The proof of these results will rely on a fundamental result due to Riemann and Lebesgue.

THEOREM 2.2.3. *(Riemann-Lebesgue lemma) If $f \in L^1([-\pi,\pi])$, then $\hat{f}(k) \to 0$ as $k \to \pm\infty$.*

PROOF. Since $f(x)$ and e^{-ikx} are periodic, we have

$$\begin{aligned} 2\pi \hat{f}(k) &= \int_{-\pi}^{\pi} f(x) e^{-ikx}\, dx = \int_{-\pi}^{\pi} f(x - \pi/k) e^{-ik(x-\pi/k)}\, dx \\ &= -\int_{-\pi}^{\pi} f(x-\pi/k) e^{-ikx}\, dx. \end{aligned}$$

Rearranging gives

$$2\pi\hat{f}(k) = \frac{1}{2}\left[\int_{-\pi}^{\pi} f(x)e^{-ikx}\,dx - \int_{-\pi}^{\pi} f(x-\pi/k)e^{-ikx}\,dx\right]$$
$$= \frac{1}{2}\int_{-\pi}^{\pi} \left[f(x) - f(x-\pi/k)\right] e^{-ikx}\,dx.$$

If f were continuous, taking absolute values of the above and using that $\sup_x |f(x) - f(x-\pi/k)| \to 0$ as $k \to \pm\infty$ would give $\hat{f}(k) \to 0$. (This uses the fact that a continuous periodic function is uniformly continuous.) In general, if $f \in L^1([-\pi,\pi])$, given any $\varepsilon > 0$ we choose a continuous periodic function g with $\|f-g\|_{L^1} < \varepsilon/2$. Then

$$|\hat{f}(k)| \le |(f-g)\hat{\ }(k)| + |\hat{g}(k)| < \varepsilon/2 + |\hat{g}(k)|.$$

The above argument for continuous functions shows that $|\hat{g}(k)| < \varepsilon/2$ for $|k|$ large enough, which concludes the proof. □

PROOF OF THEOREM 2.2.2. If $f|_{(x-\delta,x+\delta)} = 0$, then

$$S_m f(x) = \frac{1}{2\pi}\int_{\delta<|y|<\pi} D_m(y) f(x-y)\,dy = \frac{1}{2\pi}\int_{-\pi}^{\pi} \sin((m+\frac{1}{2})y) g(y)\,dy$$

where

$$g(y) = \frac{f(x-y)}{\sin(\frac{1}{2}y)}\chi_{\{\delta<|y|<\pi\}}.$$

Here $g \in L^1([-\pi,\pi])$, and by writing $\sin t = \frac{e^{it}-e^{-it}}{2i}$ we have

$$S_m f(x) = -(e^{-iy/2}g/2i)\hat{\ }(m) + (e^{iy/2}g/2i)\hat{\ }(-m).$$

The Riemann-Lebesgue lemma shows that $S_m f(x) \to 0$ as $m \to \infty$. □

PROOF OF THEOREM 2.2.1. Since $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_m(x)\,dx = 1$, we write

$$2\pi\left[S_m f(x) - f(x)\right] = \int_{-\pi}^{\pi} D_m(y)\left[f(x-y) - f(x)\right]\,dy$$
$$= \int_{|y|<\delta} D_m(y)\left[f(x-y) - f(x)\right]\,dy + \int_{\delta<|y|<\pi} D_m(y)\left[f(x-y) - f(x)\right]\,dy.$$

Since $D_m(y) = \frac{\sin((m+\frac{1}{2})y)}{\sin(\frac{1}{2}y)}$ and since $|\sin(\frac{1}{2}y)| \sim |\frac{1}{2}y|$ for y small, the first integral satisfies

$$\left|\int_{|y|<\delta} D_m(y)\left[f(x-y) - f(x)\right]\,dy\right| \le C\int_{|y|<\delta}\left|\frac{f(x-y) - f(x)}{y}\right|\,dy.$$

By assumption, the last expression can be made arbitrarily small by taking δ small. Also the second integral converges to zero as $m \to \infty$ by the same argument as in the proof of Theorem 2.2.2. □

As discussed above, the problem with pointwise convergence is that the Dirichlet kernel D_m takes negative values. One does get an approximate identity if a different summation method is used: instead of the partial sums $S_m f$ consider the *Cesàro sums*

$$\sigma_N f(x) = \frac{1}{N+1} \sum_{m=0}^{N} S_m f(x).$$

This can be written in convolution form as

$$\sigma_N f(x) = \frac{1}{N+1} \sum_{m=0}^{N} (D_m * f)(x) = (F_N * f)(x)$$

where F_N is the *Fejér kernel*,

$$\begin{aligned} F_N(x) &= \frac{1}{N+1} \sum_{m=0}^{N} \frac{e^{i(m+\frac{1}{2})x} - e^{-i(m+\frac{1}{2})x}}{e^{i\frac{1}{2}x} - e^{-i\frac{1}{2}x}} \\ &= \frac{1}{N+1} \frac{e^{i\frac{1}{2}x} \frac{e^{i(N+1)x}-1}{e^{ix}-1} - e^{-i\frac{1}{2}x} \frac{e^{-i(N+1)x}-1}{e^{-ix}-1}}{e^{i\frac{1}{2}x} - e^{-i\frac{1}{2}x}} \\ &= \frac{1}{N+1} \frac{e^{i(N+1)x} - 1 + e^{-i(N+1)x} - 1}{(e^{i\frac{1}{2}x} - e^{-i\frac{1}{2}x})^2} \\ &= \frac{1}{N+1} \frac{\sin^2(\frac{N+1}{2}x)}{\sin^2(\frac{1}{2}x)}. \end{aligned}$$

Clearly this is nonnegative, and in fact F_N is an approximate identity (exercise). It follows from Lemma 2.1.5 that Cesàro sums of the Fourier series an L^p function always converge in the L^p norm if $1 \leq p < \infty$.

Theorem 2.2.4. *(Cesàro summability of Fourier series) Assume that $f \in L^p([-\pi, \pi])$ where $1 \leq p < \infty$, or that f is a continuous 2π-periodic function and $p = \infty$. Then*

$$\|\sigma_N f - f\|_{L^p([-\pi,\pi])} \to 0 \quad \text{as } N \to \infty.$$

2.3. Properties of Fourier series

In this section we describe the basic properties of Fourier series and their behaviour under various operations, such as differentiation.

In order for such operations to be well defined, we will work with a particularly nice class of periodic functions.

DEFINITION. Let $\mathscr{P}$ be the set of all C^∞ functions $\mathbb{R}^n \to \mathbb{C}$ that are 2π-periodic in each variable (periodic for short). Elements of $\mathscr{P}$ are called *periodic test functions*.

EXAMPLE. Any trigonometric polynomial $\sum_{|k|\le N} c_k e^{ik\cdot x}$ is in $\mathscr{P}$.

REMARK. The name *test functions* comes from distribution theory. There the first step is to consider a class of very nice functions, called test functions, and operations on them. Later, distributions are defined as elements of the dual space of test functions. In this course we will not have time to consider periodic distributions. However, we will develop the theory of tempered distributions in connection with the Fourier transform on $\mathbb{R}^n$.

The set $\mathscr{P}$ is an infinite-dimensional vector space under the usual addition and scalar multiplication of functions. We now consider some basic operations on functions in $\mathscr{P}$. To introduce some notation, define the reflection

$$\tilde{u}(x) = u(-x),$$

translation (for $x_0 \in \mathbb{R}^n$)

$$(\tau_{x_0} u)(x) = u(x - x_0),$$

and convolution (for $u, v \in \mathscr{P}$)

$$(u * v)(x) = (2\pi)^{-n} \int_{[-\pi,\pi]^n} u(x-y)v(y)\,dy.$$

Here are some basic properties of convolution:

LEMMA 2.3.1. *If $u, v \in \mathscr{P}$ then $u * v \in \mathscr{P}$. Moreover, for functions in $\mathscr{P}$ we have*

(1) $u * v = v * u$ *(commutativity)*

(2) $u * (v * w) = (u * v) * w$ *(associativity)*

(3) $\partial^\alpha (u * v) = (\partial^\alpha u) * v = u * (\partial^\alpha v)$ *(derivative)*

PROOF. Let $u, v \in \mathscr{P}$. We first observe that $u * v$ is uniformly continuous: if $x, y \in \mathbb{R}^n$ write

$$|u * v(x) - u * v(y)| \leq (2\pi)^{-n} \int_{[-\pi,\pi]^n} |u(x-z) - u(y-z)||v(z)|\,dz$$
$$\leq (2\pi)^{-n} \|v\|_{L^1([-\pi,\pi]^n} \sup_{z \in [-\pi,\pi]^n} |u(x-z) - u(y-z)|.$$

Since u is uniformly continuous, for any $\varepsilon > 0$ we may find $\delta > 0$ so that the last expression is $< \varepsilon$ when $|x-y| < \delta$. Thus $u * v$ is a continuous periodic function.

For derivatives, note that

$$\frac{u * v(x + he_j) - u * v(x)}{h} = (2\pi)^{-n} \int_{[-\pi,\pi]^n} \frac{u(x + he_j - y) - u(x-y)}{h} v(y)\,dy.$$

By the mean value theorem, if $|h| \leq 1$ there is θ with $|\theta| \leq 1$ so that

$$\left| \frac{u(x + he_j - y) - u(x-y)}{h} \right| = |\partial_j u(x + \theta e_j - y)| \leq \|\nabla u\|_{L^\infty}.$$

The last bound is independent of x, y, h, and dominated convergence allows to take the limit as $h \to 0$ in the earlier integral to obtain that

$$\partial_j (u * v)(x) = (\partial_j u) * v(x).$$

Since $\partial_j u \in \mathscr{P}$, the first part of the argument shows that $\partial_j (u * v)$ is a continuous periodic function for each j. Iterating this argument implies that $u * v \in \mathscr{P}$ and $\partial^\alpha (u * v) = (\partial^\alpha u) * v$. Parts (1) – (3) are left as an exercise. □

THEOREM 2.3.2. *(Operations on test functions) If $f, v \in \mathscr{P}$, then the following operations map $\mathscr{P}$ to $\mathscr{P}$:*

(1) $u \mapsto \tilde{u}$ *(reflection)*
(2) $u \mapsto \overline{u}$ *(conjugation)*
(3) $u \mapsto \tau_{x_0} u$ *(translation)*
(4) $u \mapsto \partial^\alpha u$ *(derivative)*
(5) $u \mapsto fu$ *(multiplication)*
(6) $u \mapsto v * u$ *(convolution).*

PROOF. Exercise. □

We move to the study of Fourier series of test functions. Clearly test functions are in L^2, and hence their Fourier coefficients are l^2 sequences and their Fourier series converge in L^2. Of course, much more is true. We begin with the definition of rapidly decreasing sequences.

DEFINITION. A sequence $c = (c_k)_{k\in\mathbb{Z}^n}$ is said to be *rapidly decreasing* if for any $N > 0$ there is $C_N > 0$ such that

$$|c_k| \leq C_N \langle k\rangle^{-N}.$$

Here $\langle k\rangle := (1 + |k|^2)^{1/2}$. The set of rapidly decreasing sequences is denoted by $\mathscr{S} = \mathscr{S}(\mathbb{Z}^n)$.

We denote by $\mathscr{F}$ the map ("Fourier transform" in the periodic setting) which takes a test function to its sequence of Fourier coefficients. Thus,

$$\mathscr{F} : \mathscr{P}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{Z}^n), \quad u \mapsto (\hat{u}(k))_{k\in\mathbb{Z}^n}.$$

THEOREM 2.3.3. *(Fourier series of test functions) $\mathscr{F}$ is a linear bijection from $\mathscr{P}(\mathbb{R}^n)$ to $\mathscr{S}(\mathbb{Z}^n)$. Any $u \in \mathscr{P}$ can be written as the uniformly convergent Fourier series*

$$u = \sum_{k\in\mathbb{Z}^n} \hat{u}(k)e^{ik\cdot x}.$$

For the proof, we need a simple lemma:

LEMMA 2.3.4. *The series*

$$\sum_{k\in\mathbb{Z}^n} \langle k\rangle^{-s}$$

converges iff $s > n$.

PROOF. Exercise. □

PROOF OF THEOREM 2.3.3. If $u \in \mathscr{P}$ then also $D^\alpha u \in \mathscr{P}$. We use repeatedly the integration by parts formula

$$\int_{[-\pi,\pi]^n} v\partial_j w\,dx = -\int_{[-\pi,\pi]^n} (\partial_j v)w\,dx, \quad v, w \in C^1([-\pi,\pi]^n) \text{ periodic,}$$

to obtain that

$$(D^\alpha u)\hat{}(k) = (D^\alpha u, e^{ik\cdot x}) = (u, D^\alpha(e^{ik\cdot x})) = k^\alpha(u, e^{ik\cdot x}) = k^\alpha \hat{u}(k).$$

In particular,

$$((1-\Delta)^N u)\hat{}(k) = \langle k\rangle^{2N}\hat{u}(k).$$

Thus

$$|\hat{u}(k)| \le \langle k\rangle^{-2N} |((1-\Delta)^N u)\hat{}(k)| \le \langle k\rangle^{-2N} \|(1-\Delta)^N u\|_{L^\infty([-\pi,\pi]^n)}.$$

This shows that for $u \in \mathscr{P}$, the sequence $(\hat{u}(k))$ is in $\mathscr{S}$.

Conversely, if $(c_k) \in \mathscr{S}$, define

$$u(x) = \sum_{k\in\mathbb{Z}^n} c_k e^{ik\cdot x}.$$

Since $|c_k e^{ik\cdot x}| \le C_N \langle k\rangle^{-N}$ for $N > n$, the series converges uniformly by Lemma 2.3.4 and the Weierstrass M-test. By a similar argument we see that $\sum_{k\in\mathbb{Z}^n} D^\alpha(c_k e^{ik\cdot x})$ converges uniformly for any α, and the limit function is then equal to $D^\alpha u$. This shows that $u \in \mathscr{P}$, the series converges uniformly, and also $c_k = \hat{u}(k)$.

We have shown that $\mathscr{F} : \mathscr{P} \to \mathscr{S}$ is bijective, and clearly it is linear. □

We conclude this section by collecting some properties of the Fourier transform on test functions. This illustrates the general philosophy that the Fourier transform exchanges certain operations:

- *translation* is exchanged with *modulation* (=multiplication by suitable complex exponentials);
- *derivatives* are exchanged with *multiplication by polynomials*;
- *convolutions* are exchanged with *products*.

Here, the convolution of two rapidly decreasing sequences $c = (c_k)$ and $d = (d_k)$ is defined as the rapidly decreasing sequence

$$(c * d)_k = \sum_{l\in\mathbb{Z}^n} c_{k-l} d_l.$$

THEOREM 2.3.5. *If $f \in \mathscr{P}$ then the Fourier transform on $\mathscr{P}$ has the following properties.*

(1)	$(\tau_{x_0} u)\hat{}(k) = e^{-ik\cdot x_0}\hat{u}(k)$	*(translation)*
(2)	$(e^{ik_0\cdot x} u)\hat{}(k) = \tau_{k_0}\hat{u}(k)$	*(modulation)*
(3)	$(D^\alpha u)\hat{}(k) = k^\alpha \hat{u}(k)$	*(derivative)*
(4)	$(u * v)\hat{}(k) = \hat{u}(k)\hat{v}(k)$	*(convolution)*
(5)	$(fu)\hat{}(k) = (\hat{f} * \hat{u})(k)$	*(product)*

PROOF. Exercise. □

2.4. Applications

In this section we show how Fourier analysis leads to remarkably simple proofs of fundamental results in diverse areas. We will prove the isoperimetric inequality in geometry, the Weyl equidistribution theorem in number theory, the (compact) Sobolev embedding theorem in function space theory, and elliptic regularity for partial differential equations.

2.4.1. Isoperimetric inequality. Let $\gamma : [a,b] \to \mathbb{R}^2$ be a C^1 simple closed curve (this means that $\gamma(a) = \gamma(b)$, γ has no other self-intersections, and $\gamma'(t) \neq 0$ everywhere), and let $U \subset \mathbb{R}^2$ be the bounded region enclosed by γ. The fact that U exists is a special case of the Jordan curve theorem, which has an easier proof in the C^1 case by a winding number argument. The length of γ is defined by

$$L = \int_a^b |\dot{\gamma}(t)|\,dt,$$

and the area of U is

$$A = |U| = \int_U dx.$$

THEOREM 2.4.1. *(Isoperimetric inequality) One has*

$$4\pi A \leq L^2$$

with equality iff γ is a circle.

We will prove the isoperimetric inequality by using Fourier series. The next result will be useful:

THEOREM 2.4.2. *(Poincaré inequality) If u is a C^1 function on $\mathbb{R}$ with period 2π, and if $\int_0^{2\pi} u(x)\,dx = 0$, then*

$$\int_0^{2\pi} |u(x)|^2\,dx \leq \int_0^{2\pi} |u'(x)|^2\,dx$$

with equality iff $u(x) = a\cos x + b\sin x$ for some $a, b \in \mathbb{C}$.

PROOF. Since u and u' are periodic and L^2, we have the Fourier series

$$u(x) = \sum_{k\neq 0} c_k e^{ikt},$$

$$u'(x) = \sum_{k\neq 0} ikc_k e^{ikt},$$

where $c_k = \hat{u}(k)$ and we have $c_0 = \frac{1}{2\pi}\int_0^{2\pi} u(x)\,dx = 0$. Then by the Parseval identity and periodicity

$$\int_0^{2\pi} |u(x)|^2\,dx = 2\pi \sum_{k\neq 0} |c_k|^2 \leq 2\pi \sum_{k\neq 0} |ikc_k|^2 = \int_0^{2\pi} |u'(x)|^2\,dx$$

with equality iff $c_k = 0$ for $k = 0, \pm 2, \pm 3, \ldots$. □

Proof of Theorem 2.4.1. We reparametrize γ so that

$$\gamma : [0, 2\pi] \to \mathbb{R}^2$$

and

$$|\dot{\gamma}(t)| = \frac{L}{2\pi}, \qquad t \in [0, 2\pi].$$

This can be achieved by taking $t = \frac{2\pi}{L}s$ where s is the arc length parameter. Then

$$\left(\frac{L}{2\pi}\right)^2 = \frac{1}{2\pi}\int_0^{2\pi} |\dot{\gamma}(t)|^2\,dt$$

and by the divergence theorem $\int_U (\partial_1 F_1 + \partial_2 F_2)\,dx = \int_{\partial U} F \cdot \nu\,dS$,

$$\begin{aligned} A &= \int_U dx = \int_U \partial_2 x_2\,dx = \int_{\partial U} x_2 \nu_2\,dS = -\int_0^{2\pi} \gamma_2(t) \frac{\dot{\gamma}_1(t)}{|\dot{\gamma}(t)|} |\dot{\gamma}(t)|\,dt \\ &= -\int_0^{2\pi} \gamma_2(t)\dot{\gamma}_1(t)\,dt \end{aligned}$$

since $\nu(\gamma(t)) = \frac{1}{|\dot{\gamma}(t)|}(\dot{\gamma}_2(t), -\dot{\gamma}_1(t))$. Then

$$\begin{aligned} L^2 - 4\pi A &= 2\pi \int_0^{2\pi} (|\dot{\gamma}(t)|^2 + 2\gamma_2(t)\dot{\gamma}_1(t))\,dt \\ &= 2\pi \left(\int_0^{2\pi} (\dot{\gamma}_1(t) + \gamma_2(t))^2\,dt + \int_0^{2\pi} (\dot{\gamma}_2(t)^2 - \gamma_2(t)^2)\,dt \right). \end{aligned}$$

We may assume that $\int_0^{2\pi} \gamma_2(t)\,dt = 0$ by subtracting a constant (i.e. translating U in the x_2 direction). Thus, Theorem 2.4.2 implies that

$$L^2 - 4\pi A \geq 0$$

and equality holds iff γ is a circle (exercise). □

2.4.2. Weyl's equidistribution theorem. Let $\alpha \in \mathbb{R}_+$, and consider the sequence $(n\alpha)_{n=1}^{\infty}$. We wish to consider the distribution of this sequence modulo 1, or equivalently the sequence $(\{n\alpha\})_{n=1}^{\infty} \subset [0,1)$ where $\{x\}$ is the fractional part of x. If α is rational, it is easy to see that the sequence $(\{n\alpha\})$ consists of finitely many rational numbers. If α is irrational, it is a theorem of Kronecker that $(\{n\alpha\})$ is a dense subset of $[0,1)$.

In this section we will show a stronger theorem due to Weyl. We say that a sequence $(x_n)_{n=1}^{\infty} \subset [0,1)$ is *equidistributed* if for any interval $[a,b] \subset [0,1)$ one has

$$\lim_{N\to\infty} \frac{\#\{x_j\,;\, x_j \in [a,b] \text{ and } j \leq N\}}{N} = b - a.$$

THEOREM 2.4.3. *(Weyl's equidistribution theorem) If $\alpha \in \mathbb{R}_+$ is irrational, the sequence $(\{n\alpha\})_{n=1}^{\infty}$ is equidistributed.*

PROOF. Let $[a,b] \subset [0,1)$ and let $\chi(x)$ be the characteristic function of $[a,b]$ extended as a 1-periodic function to $\mathbb{R}$. We need to show that for the choice $f = \chi$, one has

$$\frac{1}{N}\sum_{n=1}^{N} f(n\alpha) \to \int_0^1 f(x)\,dx \quad \text{as } N \to \infty. \tag{2.1}$$

We first show (2.1) for $f(x) = e^{2\pi i k x}$ where $k \in \mathbb{Z}$. In fact, one has

$$\frac{1}{N}\sum_{n=1}^{N} f(n\alpha) = \frac{1}{N} e^{2\pi i k\alpha} \sum_{n=0}^{N-1} e^{2\pi i k n\alpha} = \frac{e^{2\pi i k\alpha}}{N} \frac{1 - e^{2\pi i k N\alpha}}{1 - e^{2\pi i k\alpha}}$$

using the fact that α is irrational (so $1 - e^{2\pi i k\alpha} \neq 0$). The last expression converges to zero as $N \to \infty$.

It follows that (2.1) holds for any trigonometric polynomial of the form

$$f(x) = \sum_{k=-M}^{M} c_k e^{2\pi i k x}.$$

By the Weierstrass approximation theorem for trigonometric polynomials (Theorem 2.1.6 for 1-periodic functions), it is easy to see that (2.1) holds for any continuous 1-periodic function f.

To see that (2.1) holds for $f = \chi$, we fix $\varepsilon > 0$, extend χ in a 1-periodic way to $\mathbb{R}$, and choose continuous 1-periodic functions $f_\pm$ such that

$$f_- \leq \chi \leq f_+ \quad \text{in } \mathbb{R}$$

and

$$b - a - \frac{\varepsilon}{2} \le \int_0^1 f_-(x)\,dx, \qquad \int_0^1 f_+(x)\,dx \le b - a + \frac{\varepsilon}{2}.$$

Since

$$\frac{1}{N}\sum_{n=1}^{N} f_-(n\alpha) \le \frac{1}{N}\sum_{n=1}^{N} \chi(n\alpha) \le \frac{1}{N}\sum_{n=1}^{N} f_+(n\alpha)$$

and since (2.1) holds for $f_\pm$, we may choose N_0 so large that for $N \ge N_0$ one has

$$\int_0^1 f_-(x)\,dx - \frac{\varepsilon}{2} \le \frac{1}{N}\sum_{n=1}^{N} \chi(n\alpha) \le \int_0^1 f_-(x)\,dx + \frac{\varepsilon}{2}.$$

This implies that

$$b - a - \varepsilon \le \frac{1}{N}\sum_{n=1}^{N} \chi(n\alpha) \le b - a + \varepsilon, \qquad N \ge N_0,$$

which proves the claim. □

2.4.3. Sobolev spaces. In this section we consider L^2 Sobolev spaces of periodic functions. These spaces correspond to the C^k spaces of continuously differentiable functions, but measure regularity in terms of derivatives being in L^2 instead of being continuous. Sobolev spaces are a central concept in the theory of partial differential equations.

Let $\mathbb{T}^n = \mathbb{R}^n/2\pi\mathbb{Z}^n$ be the n-dimensional torus. Note that $L^2(Q)$ above may be identified with $L^2(\mathbb{T}^n)$. However, $C^k(Q)$ is different from $C^k(\mathbb{T}^n)$; in fact $C^k(\mathbb{T}^n)$ (resp. $C^\infty(\mathbb{T}^n)$) can be identified with the C^k (resp. C^∞) 2π-periodic functions in $\mathbb{R}^n$.

To define Sobolev spaces, we need to define weak derivatives of L^2 functions.

Definition. (Weak derivative) If $u \in L^2(\mathbb{T}^n)$ and $\alpha \in \mathbb{N}^n$, we say that u has *αth weak derivative* in $L^2(\mathbb{T}^n)$ if there is $u_\alpha \in L^2(\mathbb{T}^n)$ such that

$$\int_{\mathbb{T}^n} u D^\alpha \varphi\,dx = (-1)^{|\alpha|} \int_{\mathbb{T}^n} u_\alpha \varphi\,dx, \qquad \varphi \in \mathscr{P}.$$

In this case we write $D^\alpha u = u_\alpha$.

It follows from the completeness of the set $\{e^{ik\cdot x}\}_{k\in\mathbb{Z}^n} \subset \mathscr{P}$ that the αth weak derivative $D^\alpha u$, if it exists, is unique.

DEFINITION. (Sobolev space $W^{m,2}(\mathbb{T}^n)$) If $m \geq 0$ is an integer, we denote by $W^{m,2}(\mathbb{T}^n)$ the space of all $u \in L^2(\mathbb{T}^n)$ such that $D^\alpha u \in L^2(\mathbb{T}^n)$ for all $\alpha \in \mathbb{N}^n$ satisfying $|\alpha| \leq m$.

The index m in $W^{m,2}$ measures the smoothness (number of derivatives), and the index 2 reflects the fact that we consider Sobolev spaces based on L^2 spaces.

EXAMPLE 2.4.1. Clearly $\mathscr{P}$ is a subset of $W^{m,2}(\mathbb{T}^n)$ for any m.

LEMMA 2.4.4. *$W^{m,2}(\mathbb{T}^n)$ is a Hilbert space when equipped with the inner product*

$$(u,v)_{W^{m,2}(\mathbb{T}^n)} = \sum_{|\alpha| \leq m} (D^\alpha u, D^\alpha v).$$

PROOF. Exercise. □

It is important that Sobolev spaces on the torus can be characterized in terms of Fourier coefficients.

LEMMA 2.4.5. *Let $u \in L^2(\mathbb{T}^n)$. Then $u \in W^{m,2}(\mathbb{T}^n)$ if and only if $(\langle k \rangle^m \hat{u}(k))_{k \in \mathbb{Z}^n} \in \ell^2(\mathbb{Z}^n)$.*

PROOF. By the Parseval identity, one has

$$\begin{aligned} u \in W^{m,2}(\mathbb{T}^n) &\Leftrightarrow D^\alpha u \in L^2(\mathbb{T}^n) \quad \text{for } |\alpha| \leq m \\ &\Leftrightarrow k^\alpha \hat{u}(k) \in \ell^2(\mathbb{Z}^n) \quad \text{for } |\alpha| \leq m \\ &\Leftrightarrow (k_1^2, \ldots, k_n^2)^\alpha |\hat{u}(k)|^2 \in \ell^1(\mathbb{Z}^n) \quad \text{for } |\alpha| \leq m. \end{aligned}$$

If the last condition is satisfied, then

$$\langle k \rangle^{2m} |\hat{u}(k)|^2 = \sum_{|\beta| \leq m} c_\beta (k_1^2, \ldots, k_n^2)^\beta |\hat{u}(k)|^2 \in \ell^1(\mathbb{Z}^n),$$

consequently $\langle k \rangle^m \hat{u}(k) \in \ell^2(\mathbb{Z}^n)$. Conversely, if $\langle k \rangle^m \hat{u}(k) \in \ell^2(\mathbb{Z}^n)$, then $k^\alpha \hat{u}(k) \in \ell^2(\mathbb{Z}^n)$ for $|\alpha| \leq m$ because $|k_j| \leq \langle k \rangle$. □

The previous result motivates the following definition, which defines Sobolev spaces also for non-integer smoothness indices.

DEFINITION. (Sobolev space $H^s(\mathbb{T}^n)$) If $s \geq 0$, we denote by $H^s(\mathbb{T}^n)$ the space of all $u \in L^2(\mathbb{T}^n)$ for which the sequence $(\langle k \rangle^s \hat{u}(k))_{k \in \mathbb{Z}^n}$ is in $\ell^2(\mathbb{Z}^n)$.

It would also be possible to define $H^s(\mathbb{T}^n)$ for $s < 0$ analogously, but the elements of these spaces would not be L^2 functions (they would be periodic distributions).

EXAMPLE 2.4.2. If $u \in H^s(\mathbb{T}^n)$, it follows that $D^\alpha u \in H^{s-|\alpha|}(\mathbb{T}^n)$.

LEMMA 2.4.6. *$H^s(\mathbb{T}^n)$ is a Hilbert space when equipped with the inner product*

$$(u, v)_{H^s(\mathbb{T}^n)} = \sum_{k \in \mathbb{Z}^n} \langle k \rangle^{2s} \hat{u}(k) \overline{\hat{v}(k)}.$$

PROOF. Exercise. □

2.4.4. Sobolev embedding. Sobolev embedding theorems come in many forms, and one of their main uses is to relate various weak regularity and integrability properties to classical regularity. It is easy to prove a version that corresponds to our present situation. The next result allows to obtain classical C^l differentiability from H^s regularity if s is sufficiently large.

THEOREM 2.4.7. *(Sobolev embedding theorem) If $s > n/2 + l$ where $l \in \mathbb{N}$, then $H^s(\mathbb{T}^n) \subset C^l(\mathbb{T}^n)$.*

PROOF. Let $u \in H^s(\mathbb{T}^n)$, so that $\langle k \rangle^s \hat{u} \in \ell^2(\mathbb{Z}^n)$ and

$$u(x) = \sum_{k \in \mathbb{Z}^n} \hat{u}(k) e^{ik \cdot x}.$$

Let $M_k = |\hat{u}(k) e^{ik \cdot x}| = \langle k \rangle^{-s} (\langle k \rangle^s |\hat{u}(k)|)$. We have

$$\sum_{k \in \mathbb{Z}^n} M_k \leq \| \langle k \rangle^{-s} \|_{\ell^2(\mathbb{Z}^n)} \| \langle k \rangle^s \hat{u}(k) \|_{\ell^2(\mathbb{Z}^n)} < \infty,$$

by Lemma 2.3.4 since $s > n/2$. Since the terms in the Fourier series of u are continuous functions, this Fourier series converges uniformly into a continuous function in $\mathbb{T}^n$ by the Weierstrass M-test. Moreover, if $|\alpha| \leq l$ we may repeat the above argument for

$$D^\alpha u(x) = \sum_{k \in \mathbb{Z}^n} k^\alpha \hat{u}(k) e^{ik \cdot x},$$

and the condition $s > n/2 + l$ guarantees that $D^\alpha u$ is a continuous periodic function for $|\alpha| \leq l$. □

To be precise, the statement $H^s(\mathbb{T}^n) \subset C^l(\mathbb{T}^n)$ means that any $u \in L^2(\mathbb{T}^n)$ that belongs to $H^s(\mathbb{T}^n)$ satisfies $u = v$ a.e. for some v in $C^l(\mathbb{T}^n)$, and we identify the function u with the C^l function v. The proof also implies the norm estimate

$$\|u\|_{C^l(\mathbb{T}^n)} \leq C\|u\|_{H^s(\mathbb{T}^n)}, \qquad u \in H^s(\mathbb{T}^n),$$

which means that the embedding $H^s(\mathbb{T}^n) \subset C^l(\mathbb{T}^n)$ is continuous.

2.4.5. Compact Sobolev embedding. Many Sobolev embeddings are much better than merely continuous: often they are *compact*. Compact operators are bounded linear operators between infinite dimensional spaces that share some of the good properties of operators between finite dimensional spaces (i.e. matrices), such as the possibility to extract convergent subsequences, the fact that existence and uniqueness of solutions to certain equations are equivalent (Fredholm alternative), and discrete spectrum.

Definition. Let X and Y be complex Banach spaces. A bounded linear operator $T : X \to Y$ is said to be *compact* if for any bounded sequence $(x_j) \subset X$, the sequence (Tx_j) has a convergent subsequence. Equivalently, T is compact if $\overline{T(B)}$ is compact in Y where B is the unit ball $B = \{x \in X\,;\, \|x\| \leq 1\}$.

Example 2.4.3. (Finite rank operators) A bounded linear operator $T : X \to Y$ is said to have *finite rank* if its range $T(X)$ is a finite dimensional subspace of Y. Any finite rank operator is compact since bounded sequences in $\mathbb{C}^n$ have convergent subsequences.

Example 2.4.4. (Integral operators) Let $\Omega \subset \mathbb{R}^n$ be an open set, and let $k \in L^2(\Omega \times \Omega)$. The integral operator

$$K : L^2(\Omega) \to L^2(\Omega), \quad Kf(x) = \int_\Omega k(x,y)f(y)\,dy$$

is compact (it is called a Hilbert-Schmidt operator). In particular, if Ω is bounded and k is continuous on $\overline{\Omega} \times \overline{\Omega}$, then K is compact. These examples indicate that many integral operators whose integral kernels are sufficiently nice are compact.

Example 2.4.5. (Limits of compact operators) If $T_j : X \to Y$ are compact operators, and if $T_j \to T$ in the operator norm where $T : X \to Y$ is a bounded linear operator, then T is compact. This is easy to see by using the fact that a closed set K in a complete metric space is compact iff it is totally bounded, meaning that for any $\varepsilon > 0$ the set K is covered by finitely many balls with radius ε.

EXAMPLE 2.4.6. (Limits of finite rank operators) If $T_j : X \to Y$ are finite rank operators and if $T_j \to T$ in the operator norm, then T is compact by the previous examples. If X and Y are Hilbert spaces, the converse is also true: any compact operator is the limit of finite rank operators.

We are now in a position to prove the compact Sobolev embedding theorem, often attributed to Rellich and Kondrachov, in the present periodic setting. The next theorem actually implies other standard versions of compact Sobolev embedding on bounded domains in $\mathbb{R}^n$.

THEOREM 2.4.8. *(Compact Sobolev embedding) The inclusion map* $i : H^s(\mathbb{T}^n) \to L^2(\mathbb{T}^n)$ *is compact if* $s > 0$.

PROOF. For $N \in \mathbb{Z}_+$ define the projection

$$P_N : H^s(\mathbb{T}^n) \to L^2(\mathbb{T}^n), \quad P_N u(x) = \sum_{|k|\leq N} \hat{u}(k) e^{ik\cdot x}.$$

Then P_N is finite rank, and to show that i is compact it is enough to prove that

$$\|i - P_N\|_{H^s(\mathbb{T}^n)\to L^2(\mathbb{T}^n)} \to 0 \qquad \text{as } N \to \infty.$$

Let $u \in H^s(\mathbb{T}^n)$, and note that

$$\|(i - P_N)u\|_{L^2(\mathbb{T}^n)} = \left(\sum_{|k|>N} |\hat{u}(k)|^2\right)^{1/2} = \left(\sum_{|k|>N} \langle k\rangle^{-2s}\langle k\rangle^{2s}|\hat{u}(k)|^2\right)^{1/2}$$
$$\leq \langle N\rangle^{-2s}\|u\|_{H^s(\mathbb{T}^n)}.$$

Thus $\|i - P_N\|_{H^s(\mathbb{T}^n)\to L^2(\mathbb{T}^n)} \leq \langle N\rangle^{-2s}$, which implies the claim. □

2.4.6. Elliptic regularity. The final result in this section will be elliptic regularity in the periodic case. Consider a constant coefficient differential operator $P(D)$ of order m acting on 2π-periodic functions in $\mathbb{R}^n$,

$$P(D) = \sum_{|\alpha|\leq m} a_\alpha D^\alpha,$$

where a_α are complex constants. The *principal part* of $P(D)$ is

$$P_m(D) = \sum_{|\alpha|=m} a_\alpha D^\alpha.$$

The *symbol* of $P(D)$ is the polynomial

$$P(\xi) = \sum_{|\alpha|\leq m} a_\alpha \xi^\alpha, \qquad \xi \in \mathbb{R}^n,$$

and the *principal symbol* of $P(D)$ is the polynomial

$$P_m(\xi) = \sum_{|\alpha|=m} a_\alpha \xi^\alpha, \qquad \xi \in \mathbb{R}^n.$$

We say that $P(D)$ is *elliptic* if

$$P_m(\xi) \neq 0 \quad \text{whenever } \xi \in \mathbb{R}^n \setminus \{0\}.$$

The following proof also indicates how Fourier series are used in the solution of partial differential equations.

THEOREM 2.4.9. *(Elliptic regularity in H^s) Let $P(D)$ be an elliptic differential operator with constant coefficients, and assume that $u \in L^2(\mathbb{T}^n)$ solves the equation*

$$P(D)u = f$$

for some $f \in H^s(\mathbb{T}^n)$. Then $u \in H^{s+m}(\mathbb{T}^n)$.

In fact the same result is true for any periodic distribution u and any $s \in \mathbb{R}$. A combination of the previous theorem and the Sobolev embedding theorem, Theorem 2.4.7, yields immediately a corresponding elliptic regularity result for C^∞ data.

THEOREM 2.4.10. *(Elliptic regularity in C^∞) If f is C^∞ in the previous theorem, then also u is C^∞.*

PROOF OF THEOREM 2.4.9. Taking the Fourier coefficients of both sides of $P(D)u = f$ gives

$$P(k)\hat{u}(k) = \hat{f}(k), \quad k \in \mathbb{Z}^n. \tag{2.2}$$

Now $P_m(\xi)$ is a homogeneous polynomial of degree m, so we have

$$|P_m(k)| = |k|^m |P_m(k/|k|)| \geq c|k|^m$$

for some $c > 0$, since the ellipticity condition implies that $P_m(\xi) \neq 0$ on the compact set $\{\xi \in \mathbb{R}^n\,;\, |\xi| = 1\}$. Then for $|k| \geq 1$,

$$\begin{aligned} |P(k)| &= |P_m(k) + \sum_{|\alpha| \leq m-1} a_\alpha k^\alpha| \\ &\geq |P_m(k)| - \sum_{|\alpha| \leq m-1} |a_\alpha||k|^{|\alpha|} \\ &\geq c|k|^m - C|k|^{m-1}. \end{aligned}$$

If $N > 0$ is sufficiently large, it follows that

$$|P(k)| \geq \frac{1}{2} c|k|^m \qquad \text{for } |k| \geq N.$$

From (2.2) we obtain

$$|\hat{u}(k)| = \Big|\frac{\hat{f}(k)}{P(k)}\Big| \leq \frac{2}{c|k|^m}|\hat{f}(k)|, \qquad |k| \geq N.$$

Since $\langle k \rangle^s \hat{f}(k) \in \ell^2(\mathbb{Z}^n)$ this shows that $\langle k \rangle^{s+m} \hat{u}(k) \in \ell^2(\mathbb{Z}^n)$, which implies $u \in H^{s+m}(\mathbb{T}^n)$ as required. $\square$

CHAPTER 3

Fourier transform

In this chapter we will discuss Fourier analysis for non-periodic functions and distributions in $\mathbb{R}^n$. If f is a complex valued function in $\mathbb{R}^n$ (say in L^1), its Fourier transform is defined by

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

For the purposes of Fourier analysis it will be useful to have a test function space which is invariant under the Fourier transform. We first introduce a space that will satisfy this criterion.

3.1. Schwartz space

The first step in distribution theory is to consider classes of very nice functions, called *test functions*, and operations on them. Later, distributions will be defined as elements of the dual space of test functions. The test function space relevant for the Fourier transform is as follows.

Definition. The *Schwartz space* $\mathscr{S}(\mathbb{R}^n)$, or the space of *rapidly decreasing functions*, is the set of infinitely differentiable complex functions on $\mathbb{R}^n$ for which the seminorms

$$\|\varphi\|_{\alpha,\beta} = \|x^\alpha \partial^\beta \varphi(x)\|_{L^\infty(\mathbb{R}^n)} \tag{3.1}$$

are finite for all $\alpha, \beta \in \mathbb{N}^n$. Equivalently, $\mathscr{S}(\mathbb{R}^n)$ is the space of functions for which the norms

$$\|\varphi\|_N = \sum_{|\beta|\leq N} \|\langle x\rangle^N \partial^\beta \varphi\|_{L^\infty(\mathbb{R}^n)}$$

are finite for all $N \in \mathbb{N}$.

Example 3.1.1. $\mathscr{S}$ is the set of those smooth functions which together with their derivatives decrease more rapidly than the inverse of any polynomial. Any compactly supported C^∞ function is in $\mathscr{S}(\mathbb{R}^n)$, and also functions like $\exp(-|x|^2)$ are in Schwartz space. The function $\exp(-|x|)$ is not in Schwartz space since it is not C^∞ near the origin.

The set $\mathscr{S}$ is an infinite-dimensional vector space under the usual addition and scalar multiplication of functions. To obtain a reasonable dual space, we need a suitable topology. In practice it will be enough to know how sequences converge, and we would like to say that a sequence $(\varphi_j)_{j=1}^\infty$ converges to φ if for any fixed N,

$$\|\varphi_j - \varphi\|_N \to 0 \text{ as } j \to \infty,$$

of equivalently if for any fixed $\alpha, \beta \in \mathbb{N}^n$,

$$\|\varphi_j - \varphi\|_{\alpha,\beta} \to 0 \text{ as } j \to \infty,$$

Sequential convergence is sufficient for describing topological properties in metric spaces, but not in general topological spaces. (If the space is not first countable, one should use *nets* or *filters* instead, and many distribution spaces are not first countable!) However, here there are no complications since there is a natural metric space topology on $\mathscr{S}$ for which sequential convergence coincides with the notion above.

Theorem 3.1.1. *($\mathscr{S}$ as a metric space) If $u, v \in \mathscr{S}$, define*

$$d(u,v) = \sum_{N=0}^\infty 2^{-N} \frac{\|u-v\|_N}{1+\|u-v\|_N}.$$

Then $(\mathscr{S}, d)$ is a metric space. Moreover, $u_j \to u$ in $(\mathscr{S}, d)$ iff for any $N \in \mathbb{N}$ one has

$$\|u_j - u\|_N \to 0.$$

Proof. Since $0 \leq t/(1+t) \leq 1$ for $t \geq 0$ we have that $d(u,v)$ is defined for all $u, v \in \mathscr{S}$ and $0 \leq d(u,v) \leq \sum_{N=0}^\infty 2^{-N} = 2$. If $d(u,v) = 0$ then $\|u-v\|_N = 0$ for all N, and the case $N = 0$ implies $u = v$. Clearly $d(u,v) = d(v,u)$, and the triangle inequality follows since

$$\begin{aligned}\frac{\|u-w\|_N}{1+\|u-w\|_N} &= \frac{1}{\frac{1}{\|u-w\|_N}+1} \\ &\leq \frac{1}{\frac{1}{\|u-v\|_N+\|v-w\|_N}+1} \leq \frac{\|u-v\|_N}{1+\|u-v\|_N} + \frac{\|v-w\|_N}{1+\|v-w\|_N}.\end{aligned}$$

Thus d is a metric on $\mathscr{S}$.

Let (u_j) be a sequence in $\mathscr{S}$. If $u_j \to u$ in $(\mathscr{S}, d)$ then $d(u_j, u) \to 0$, which implies that $\frac{\|u_j - u\|_N}{1+\|u_j-u\|_N} \to 0$ for all N. Thus $\|u_j - u\|_N \leq 1$ for j sufficiently large, and we obtain that $\|u_j - u\|_N \to 0$ for all N. For the converse, suppose that $\|u_j - u\|_N \to 0$ for all N. Given $\varepsilon > 0$, first choose N_0 so that

$$\sum_{N=N_0+1}^{\infty} 2^{-N} \leq \varepsilon/2.$$

Then choose j_0 so large that for $j \geq j_0$ we have

$$\sum_{N=0}^{N_0} 2^{-N} \frac{\|u_j - u\|_N}{1 + \|u_j - u\|_N} \leq \varepsilon/2.$$

Then $d(u_j, u) \leq \varepsilon$ for $j \geq j_0$, showing that $d(u_j, u) \to 0$. □

The previous theorem is an instance of a general phenomenon: a complex vector space X whose topology is given by a countable separating family of seminorms is in fact a metric space. Here, a map $\rho : X \to \mathbb{R}$ is called a *seminorm* if for any $u, v \in X$ and for $c \in \mathbb{C}$,

(1) $\rho(u) \geq 0$ (nonnegativity)

(2) $\rho(u+v) \leq \rho(u) + \rho(v)$ (subadditivity)

(3) $\rho(cu) = |c|\rho(u)$ (homogeneity)

Thus, a seminorm ρ is almost like a norm but it is allowed that $\rho(u) = 0$ for some nonzero elements $u \in X$. A family $\{\rho_\alpha\}_{\alpha \in A}$ is called *separating* if for any nonzero $u \in X$ there is $\alpha \in A$ with $\rho_\alpha(x) \neq 0$.

THEOREM 3.1.2. *Let X be a vector space and let $\{\rho_N\}_{N=0}^{\infty}$ be a countable separating family of seminorms. The function*

$$d(u, v) = \sum_{N=0}^{\infty} 2^{-N} \frac{\rho_N(u-v)}{1 + \rho_N(u-v)}, \qquad u, v \in X,$$

is a metric on X. Moreover, $u_j \to u$ in (X, d) iff for any N one has

$$\rho_N(u_j - u) \to 0.$$

PROOF. Exercise. □

If X and $\{\rho_N\}$ are as in the theorem, we say that the metric space topology of (X, d) is the topology on X *induced* by the family of seminorms $\{\rho_N\}$. This notion will be used several times later. In particular, the topology on $\mathscr{S}$ is the one induced by the seminorms (actually norms) $\{\|\cdot\|_N\}$, and it is also equal to the topology induced by the seminorms $\{\|x^\alpha \partial^\beta \cdot \|_{L^\infty}\}_{\alpha,\beta\in\mathbb{N}^n}$. From now on we will always consider $\mathscr{S}$ with this topology.

It will be important that the test function space is complete.

THEOREM 3.1.3. *(Completeness) Any Cauchy sequence in $\mathscr{S}$ converges.*

PROOF. Let (u_j) be a Cauchy sequence in $\mathscr{S}$, that is, for any $\varepsilon > 0$ there is M such that

$$d(u_j, u_k) \leq \varepsilon \qquad \text{for } j,k \geq M.$$

In particular, this implies for any fixed N that

$$2^{-N}\frac{\|u_j - u_k\|_N}{1+\|u_j - u_k\|_N} \leq \varepsilon \qquad \text{for } j,k \geq M.$$

Then for any multi-indices α,β and for any $\varepsilon > 0$ there exists $M > 0$ such that $j,k \geq M$ implies $\|u_j - u_k\|_{\alpha,\beta} < \varepsilon$. The last condition may be written

$$\|x^\alpha \partial^\beta u_j - x^\alpha \partial^\beta u_k\|_{L^\infty} < \varepsilon. \tag{3.2}$$

Hence the sequence $(x^\alpha \partial^\beta u_j)$ is Cauchy in the complete space $C(\mathbb{R}^n)$ and converges uniformly to a continuous bounded function $g_{\alpha,\beta}$.

Denote by g the limit function $g_{0,0}$. Since all the sequences $(\partial^\beta u_j)$ converge uniformly we have that g is C^∞ and $\partial^\beta g = g_{0,\beta}$. It now follows from (3.2) that $x^\alpha \partial^\beta g = g_{\alpha,\beta}$ and $g \in \mathscr{S}$, and clearly $u_k \to g$ in $\mathscr{S}$. □

REMARK. The previous results show that $\mathscr{S}$ is a *Fréchet space.* By definition, a Fréchet space is a locally convex Hausdorff topological vector space whose topology is given by an invariant metric and which is complete as a metric space (a metric d on a vector space is said to be *invariant* if $d(u+w, v+w) = d(u,v)$ for all u,v,w). This terminology will not be important in what follows.

We wish to consider various operations on $\mathscr{S}$. To introduce some notation, define the reflection

$$\tilde{u}(x) = u(-x),$$

and translation (for $x_0 \in \mathbb{R}^n$)

$$(\tau_{x_0} u)(x) = u(x - x_0).$$

In order to have a sufficiently general multiplication on $\mathscr{S}$ we are led to introduce a new space of functions.

Definition. The space $\mathscr{O}_M(\mathbb{R}^n)$ is the set of all C^∞ functions $\mathbb{R}^n \to \mathbb{C}$ which together with all their derivatives are polynomially bounded; that is, $f \in \mathscr{O}_M$ if $f \in C^\infty(\mathbb{R}^n)$ and for any $\alpha \in \mathbb{N}^n$ there exist $C = C_\alpha > 0, N = N_\alpha \geq 0$ such that

$$|\partial^\alpha f(x)| \leq C\langle x\rangle^N, \qquad x \in \mathbb{R}^n.$$

Members of $\mathscr{O}_M$ are sometimes called C^∞ functions of slow growth.

Theorem 3.1.4. *If $f \in \mathscr{O}_M$ and $v \in \mathscr{S}$, then the following operations are continuous maps from $\mathscr{S}$ into $\mathscr{S}$.*

(1) $u \mapsto \tilde{u}$ *(reflection)*

(2) $u \mapsto \overline{u}$ *(conjugation)*

(3) $u \mapsto \tau_{x_0} u$ *(translation)*

(4) $u \mapsto \partial^\alpha u$ *(derivative)*

(5) $u \mapsto fu$ *(multiplication)*

Proof. Parts (1) and (2) are clear, and for (3) we may use the identity $x^\alpha = (x - x_0 + x_0)^\alpha = \sum_{\gamma\leq\alpha} c_\gamma (x-x_0)^\gamma$ to obtain

$$\begin{aligned}\|\tau_{x_0}u\|_{\alpha,\beta} &= \sup_{x\in\mathbb{R}^n} |x^\alpha \partial^\beta u(x-x_0)| \\ &\leq C\sum_{\gamma\leq\alpha}\sup_{x\in\mathbb{R}^n}|(x-x_0)^\gamma\partial^\beta u(x-x_0)| = C\sum_{\gamma\leq\alpha}\|u\|_{\gamma,\beta}.\end{aligned}$$

This shows that $\tau_{x_0}u_j \to 0$ in $\mathscr{S}$ whenever $u_j \to 0$ in $\mathscr{S}$. Part (4) follows from

$$\|\partial^\beta u\|_{\alpha',\beta'} = \|u\|_{\alpha',\beta'+\beta}.$$

It remains to show (5). Since $f \in \mathscr{O}_M$, given any β we may choose C and N such that $|\langle x\rangle^{-N}\partial^\gamma f(x)| \leq C$ whenever $\gamma \leq \beta$. Now we have

$$\begin{aligned}\|fu\|_{\alpha,\beta} &= \|x^\alpha \partial^\beta(fu)\|_{L^\infty} \\ &= \|x^\alpha \sum_{\gamma\leq\beta} c_\gamma (\partial^{\beta-\gamma}f)(\partial^\gamma u)\|_{L^\infty} \\ &\leq C\sum_{\gamma\leq\beta}\|x^\alpha\langle x\rangle^N(\langle x\rangle^{-N}\partial^{\beta-\gamma}f)(\partial^\gamma u)\|_{L^\infty} \\ &\leq C\sum_{\gamma\leq\beta}\|x^\alpha\langle x\rangle^N\partial^\gamma u\|_{L^\infty}.\end{aligned}$$

This shows that $fu_j \to 0$ in $\mathscr{S}$ whenever $u_j \to 0$ in $\mathscr{S}$. $\square$

The following elementary fact is often useful.

LEMMA 3.1.5. *The integral*

$$\int_{\mathbb{R}^n} \langle x \rangle^{-s}\, dx$$

is finite iff $s > n$.

EXAMPLE 3.1.2. The space $\mathscr{S}(\mathbb{R}^n)$ is contained in $L^p(\mathbb{R}^n)$ for all $p \geq 1$. If $\varphi \in \mathscr{S}(\mathbb{R}^n)$, then for $p = 1$ the claim follows from

$$\begin{aligned} \|\varphi\|_{L^1} &= \int_{\mathbb{R}^n} \langle x \rangle^{-n-1} \left[\langle x \rangle^{n+1} |\varphi(x)| \right]\, dx \\ &\leq C \| \langle x \rangle^{n+1} \varphi(x) \|_{L^\infty}. \end{aligned}$$

For $p > 1$ the result is given by the inequality

$$\begin{aligned} \|\varphi\|_{L^p} &= \left(\int_{\mathbb{R}^n} |\varphi(x)| |\varphi(x)|^{p-1}\, dx \right)^{1/p} \\ &\leq \|\varphi\|_{L^1}^{1/p} \|\varphi\|_{L^\infty}^{1-1/p}. \end{aligned}$$

These expressions also show that the inclusion map $i : \mathscr{S}(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$ is continuous, that is,

$$\varphi_j \to 0 \text{ in } \mathscr{S} \implies \varphi_j \to 0 \text{ in } L^p.$$

3.2. Fourier transform on Schwartz space

DEFINITION. The *Fourier transform* of a function $f \in \mathscr{S}(\mathbb{R}^n)$ is the function $\hat{f} : \mathbb{R}^n \to \mathbb{C}$ defined by

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)\, dx, \qquad \xi \in \mathbb{R}^n. \tag{3.3}$$

The *inverse Fourier transform* of $f \in \mathscr{S}(\mathbb{R}^n)$ is defined by

$$\check{f}(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} f(\xi)\, d\xi, \qquad x \in \mathbb{R}^n. \tag{3.4}$$

The Fourier transform is also denoted by $\mathscr{F}\{f(x)\}$ and the inverse transform by $\mathscr{F}^{-1}\{f(\xi)\}$ (this notation will be justified soon). We use the name "Fourier transform" both for the function $\hat{f}$ which is the image of some $f \in \mathscr{S}$, and for the linear map $\mathscr{F}$ defined on the function space $\mathscr{S}$ by the formula (3.3).

Since $\mathscr{S} \subset L^1$ it is clear that the integral (3.3) exists for all $\xi \in \mathbb{R}^n$. The following theorem justifies our choice of the Schwartz space as the first setting in which the Fourier transform is discussed. It says that not only does the Fourier transform map $\mathscr{S}$ into $\mathscr{S}$, but also that the map is an isomorphism (a linear bijective continuous map with continuous inverse).

THEOREM 3.2.1. *(Fourier transform on Schwartz space) The Fourier transform is an isomorphism from $\mathscr{S}(\mathbb{R}^n)$ onto $\mathscr{S}(\mathbb{R}^n)$. The inverse map is the inverse Fourier transform: one has $\mathscr{F}^{-1}\mathscr{F}f = \mathscr{F}\mathscr{F}^{-1}f = f$ for $f \in \mathscr{S}$.*

To show this, the first point to observe is that the rapid decrease of functions in $\mathscr{S}$ ensures that the Fourier transform is infinitely differentiable.

LEMMA 3.2.2. *For any $f \in \mathscr{S}(\mathbb{R}^n)$, the Fourier transform $\hat{f}$ is a C^∞ function from $\mathbb{R}^n$ to $\mathbb{C}$ and $\partial^\alpha \hat{f} \in L^\infty(\mathbb{R}^n)$ for all $\alpha \in \mathbb{N}^n$.*

PROOF. The function $\hat{f}$ is bounded since

$$|\hat{f}(\xi)| \leq \int_{\mathbb{R}^n} \left|e^{-ix\cdot\xi} f(x)\right| \, dx = \|f\|_{L^1}. \tag{3.5}$$

For differentiability consider the expression

$$\frac{\hat{f}(\xi + he_k) - \hat{f}(\xi)}{h} = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x) \frac{e^{-ihx_k} - 1}{h} \, dx. \tag{3.6}$$

The estimate

$$\left|\frac{e^{-ihx_k} - 1}{h}\right| = \left|\int_0^{x_k} e^{-iht} \, dt\right| \leq |x_k|$$

shows that the integrand on the right side of (3.6) is in L^1, so an application of the dominated convergence theorem gives

$$\frac{\partial}{\partial \xi_k} \hat{f}(\xi) = \mathscr{F}\{(-ix_k) f(x)\}. \tag{3.7}$$

It follows that the first partial derivatives of $\hat{f}$ are bounded functions. Since $x^\alpha f(x)$ is in $\mathscr{S}$ for any multi-index α, we may repeat the process to see that derivatives of any order are bounded continuous functions in $\mathbb{R}^n$. □

THEOREM 3.2.3. *(Properties of Fourier transform) Let $f \in \mathscr{S}(\mathbb{R}^n)$, $x_0, \xi_0 \in \mathbb{R}^n$, $c > 0$ and $\alpha, \beta \in \mathbb{N}^n$. Then the following identities hold:*

(1) $\mathscr{F}\{\tau_{x_0} f(x)\} = e^{-ix_0\cdot\xi}\hat{f}(\xi)$ *(translation)*

(2) $\mathscr{F}\{e^{ix\cdot\xi_0} f(x)\} = \tau_{\xi_0}\hat{f}(\xi)$ *(modulation)*

(3) $\mathscr{F}\{f(cx)\} = c^{-n}\hat{f}(\xi/c)$ *(scaling)*

(4) $\mathscr{F}\{D^\alpha f(x)\} = \xi^\alpha \hat{f}(\xi)$ *(derivative)*

(5) $\mathscr{F}\{(-x)^\beta f(x)\} = D^\beta\hat{f}(\xi)$ *(polynomial)*

PROOF. The identities (1), (2) and (3) follow from linear changes of variables in the defining integral. For (4), integration by parts gives

$$\mathscr{F}\left\{\frac{\partial}{\partial x_k} f(x)\right\} = \int_{\mathbb{R}^n} e^{-ix\cdot\xi}\frac{\partial f}{\partial x_k}(x)\, dx = -\int_{\mathbb{R}^n}\frac{\partial}{\partial x_k}\left(e^{-ix\cdot\xi}\right) f(x)\, dx$$
$$= (i\xi_k)\int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)\, dm_n = (i\xi_k)\hat{f}(\xi).$$

Thus $\mathscr{F}\{D_{x_k} f\} = \xi_k \hat{f}$, and (4) follows by iteration. Part (5) is given by repeated application of the formula (3.7) which was obtained in the proof of Lemma 3.2.2. □

LEMMA 3.2.4. *$\mathscr{F}$ and $\mathscr{F}^{-1}$ map $\mathscr{S}(\mathbb{R}^n)$ to $\mathscr{S}(\mathbb{R}^n)$ continuously.*

PROOF. Let $f \in \mathscr{S}$ and let α, β be multi-indices. Lemma 3.2.2 showed that $\hat{f}$ is a bounded C^∞ function. From Theorem 3.2.3, parts (4) and (5) we have

$$\begin{aligned}\|\hat{f}\|_{\alpha,\beta} &= \sup_{\xi\in\mathbb{R}^n} |\xi^\alpha D^\beta \hat{f}(\xi)| = \sup_{\xi\in\mathbb{R}^n} |(i\xi)^\alpha D^\beta\hat{f}(\xi)| \\ &= \sup_{\xi\in\mathbb{R}^n} |\mathscr{F}\left\{D^\alpha\left[(-ix)^\beta f(x)\right]\right\}|.\end{aligned} \tag{3.8}$$

Now $D^\alpha\left[(-ix)^\beta f(x)\right]$ is in $\mathscr{S}$, so the Fourier transform of this function is bounded, and $\|\hat{f}\|_{\alpha,\beta}$ is finite. Hence $\hat{f}$ is in $\mathscr{S}$.

Clearly the Fourier transform is linear. To establish the continuity of $\mathscr{F}$, we note that (3.8) and (3.5) give

$$\|\hat{f}\|_{\alpha,\beta} \le (2\pi)^{-n/2}\|D^\alpha[(-ix)^\beta f(x)]\|_{L^1}.$$

Now by the Leibniz rule, $D^\alpha[(-ix)^\beta f(x)] = \sum_{k=1}^m c_k x^{\alpha_k} D^{\beta_k} f(x)$ for some constants c_k and multi-indices α_k, β_k, so

$$\|\hat{f}\|_{\alpha,\beta} \le C\sum_{k=1}^m \|x^{\alpha_k} D^{\beta_k} f\|_{L^1} \le C\sum_{k=1}^m \|x^{\alpha_k+n+1} D^{\beta_k} f\|_{L^\infty}.$$

Now if $f_j \to 0$ in $\mathscr{S}$ we have $\|\hat{f}_j\|_{\alpha,\beta} \to 0$ for all α, β, showing that $\mathscr{F}$ is continuous. The proof that $\mathscr{F}^{-1}$ is continuous is similar. $\square$

It remains to establish the Fourier inversion theorem. The proof rests on the following simple lemma on the Fourier transform of a Gaussian function.

LEMMA 3.2.5. *The function $\phi_n \in \mathscr{S}(\mathbb{R}^n)$ given by*

$$\phi_n(x) = e^{-\frac{1}{2}|x|^2}$$

satisfies $\hat{\phi}_n = (2\pi)^{n/2}\phi_n$ and $\phi_n(0) = (2\pi)^{-n}\int_{\mathbb{R}^n} \hat{\phi}_n(x)\,dx$.

PROOF. We have

$$\hat{\phi}_1(\xi) = \int_{-\infty}^{\infty} e^{-ix\xi}e^{-\frac{1}{2}x^2}\,dx = e^{-\frac{1}{2}\xi^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x+i\xi)^2}\,dx. \tag{3.9}$$

Integrating $e^{-\frac{1}{2}z^2}$ along the rectangular contour with corners at $(\pm R, 0)$ and $(\pm R, \xi)$ gives

$$\int_{-R}^{R} e^{-\frac{1}{2}(x+i\xi)^2}dx = \int_{-R}^{R} e^{-\frac{1}{2}x^2}dx + \int_0^{\xi}\left\{e^{-\frac{1}{2}(R+iy)^2}dy - e^{-\frac{1}{2}(-R+iy)^2}dy\right\}$$

Taking the limit as $R \to \infty$, the last integral on the right becomes zero and we are left with the known integral $\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}dx = \sqrt{2\pi}$. Thus

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x+i\xi)^2}dx = \sqrt{2\pi}. \tag{3.10}$$

Now (3.9) and (3.10) give that $\hat{\phi}_1 = (2\pi)^{1/2}\phi_1$.

Moving to n dimensions, we see that $\phi_n(x) = \phi_1(x_1)\cdots\phi_1(x_n)$, from which one obtains $\hat{\phi}_n(\xi) = \hat{\phi}_1(\xi_1)\cdots\hat{\phi}_n(\xi_n)$ by repeated use of Fubini's theorem. Hence $\hat{\phi}_n = (2\pi)^{n/2}\phi_n$. The second assertion is evident. $\square$

THEOREM 3.2.6. *(Fourier inversion theorem) For any $f \in \mathscr{S}(\mathbb{R}^n)$ one has the inversion formula $\mathscr{F}^{-1}\mathscr{F}f = f$, that is,*

$$f(x) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\hat{f}(\xi)\,d\xi.$$

PROOF. For $f, g \in L^1(\mathbb{R}^n)$, an application of Fubini's theorem to the integral

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

gives the identity

$$\int_{\mathbb{R}^n} \hat{f}(x)g(x)\,dx = \int_{\mathbb{R}^n} f(y)\hat{g}(y)\,dy. \tag{3.11}$$

Let now f be any function in $\mathscr{S}(\mathbb{R}^n)$ and choose $g(x) = \varphi(x/c)$, where $\varphi \in \mathscr{S}$ and $c > 0$. The scaling property of the Fourier transform gives

$$\begin{aligned}\int_{\mathbb{R}^n} \hat{f}(x)\varphi(x/c)\,dx &= \int_{\mathbb{R}^n} f(y)c^n\hat{\varphi}(cy)\,dy \\ &= \int_{\mathbb{R}^n} f(y/c)\hat{\varphi}(y)\,dy.\end{aligned}$$

Both the integrands are in L^1, so dominated convergence implies that we may take $c \to \infty$ to obtain

$$\varphi(0)\int_{\mathbb{R}^n} \hat{f}(x)\,dx = f(0)\int_{\mathbb{R}^n} \hat{\varphi}(y)\,dy.$$

If φ is taken to be the Gaussian ϕ_n in Lemma 3.2.5 then we obtain that $f(0) = (2\pi)^{-n}\int_{\mathbb{R}^n} \hat{f}(x)\,dx$. This gives the inversion theorem for $x = 0$, and the general case is a consequence of the translation property of the Fourier transform (Theorem 3.2.3, part (1)). □

Proof of Theorem 3.2.1. We have seen that $\mathscr{F}$ and $\mathscr{F}^{-1}$ are continuous maps from $\mathscr{S}$ to $\mathscr{S}$, and that $\mathscr{F}^{-1}\mathscr{F}f = f$. Since $\mathscr{F}^{-1}f = (2\pi)^{-n}(\mathscr{F}f)\tilde{}$, it follows that $\mathscr{F}$ is bijective and the proof is concluded. □

Theorem 3.2.7. *For $f, g \in \mathscr{S}(\mathbb{R}^n)$ one has*

(1) $\mathscr{F}^2 f = (2\pi)^n \tilde{f}$ *and* $\mathscr{F}^4 f = (2\pi)^{2n} f$ *(symmetry)*

(2) $\int_{\mathbb{R}^n} \hat{f}(x)g(x)\,dx = \int_{\mathbb{R}^n} f(x)\hat{g}(x)\,dx$ *(Parseval identity)*

(3) $\int_{\mathbb{R}^n} f(x)\overline{g(x)}\,dx = (2\pi)^{-n}\int_{\mathbb{R}^n} \hat{f}(\xi)\overline{\hat{g}(\xi)}\,d\xi$ *(Parseval identity)*

(4) $\int_{\mathbb{R}^n} |f(x)|^2\,dx = (2\pi)^{-n}\int_{\mathbb{R}^n} |\hat{f}(\xi)|^2\,d\xi$ *(Parseval identity)*

Proof. Part (1) is evident from the Fourier inversion theorem. The first of Parseval's identities was established in (3.11), and the others are special cases. □

3.3. The space of tempered distributions

Section 3.1 discussed the rapidly decreasing functions which will be the test functions of choice in Fourier analysis. The next step is to define the corresponding class of distributions, namely the tempered distributions, which will possess a distributional Fourier transform. Tempered distributions are just elements of the dual space of $\mathscr{S}$.

Definition. Let $\mathscr{S}'(\mathbb{R}^n)$ be the set of continuous linear functionals on $\mathscr{S}(\mathbb{R}^n)$. Thus

$$\mathscr{S}'(\mathbb{R}^n) = \{T : \mathscr{S}(\mathbb{R}^n) \to \mathbb{C}\,;\ T \text{ linear and } T(\varphi_j) \to 0 \\ \text{whenever } \varphi_j \to 0 \text{ in } \mathscr{S}(\mathbb{R}^n)\}.$$

The elements of $\mathscr{S}'$ are called *tempered distributions*. The pairing of a distribution and test function will also be written as

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

The terminology is due to Schwartz and means that $\mathscr{S}'$ is in a sense the space of distributions with polynomial (or slow) growth. We make a remark at this point: to check that a linear functional $T : \mathscr{S} \to \mathbb{C}$ is continuous, it is indeed enough to check that

$$\varphi_j \to 0 \text{ in } \mathscr{S} \implies T(\varphi_j) \to 0.$$

This follows immediately from the linearity of T, and will be used many times below.

We can give several examples of tempered distributions.

Example 3.3.1. (Polynomially bounded functions) If $f : \mathbb{R}^n \to \mathbb{C}$ is any measurable polynomially bounded function f, in the sense that $|f(x)| \leq C\langle x \rangle^N$ for a.e. $x \in \mathbb{R}^n$, define

$$T_f : \mathscr{S}(\mathbb{R}^n) \to \mathbb{C}, \quad T_f(\varphi) = \int_{\mathbb{R}^n} f\varphi \, dx.$$

Then T_f is a tempered distribution, since it is linear and for any φ in $\mathscr{S}$ we have

$$\begin{aligned} |T_f(\varphi)| = \Big| \int_{\mathbb{R}^n} f\varphi \, dx \Big| &\leq C \int_{\mathbb{R}^n} \langle x \rangle^N |\varphi(x)| \, dx \\ &\leq C \| \langle x \rangle^{N+n+1} \varphi \|_{L^\infty} \end{aligned}$$

by Lemma 3.1.5. Thus $T(\varphi_j) \to 0$ whenever $\varphi_j \to 0$ in $\mathscr{S}$. Moreover, it is possible to identify the distribution T_f with the function f, since the condition $T_{f_1} = T_{f_2}$ for two such functions f_1 and f_2 implies that

$$\int_{\mathbb{R}^n} (f_1 - f_2)\varphi \, dx = 0, \qquad \varphi \in \mathscr{S}(\mathbb{R}^n),$$

which implies that $f_1 = f_2$ a.e. by a convolution approximation result to be discussed later.

EXAMPLE 3.3.2. (L^p functions) Each space $L^p(\mathbb{R}^n)$, $1 \leq p \leq \infty$, is contained in $\mathscr{S}'(\mathbb{R}^n)$. This follows from Hölder's inequality since if $\frac{1}{p} + \frac{1}{p'} = 1$ and $f \in L^p$ then

$$|T_f(\varphi)| = \Big| \int_{\mathbb{R}^n} f(x)\varphi(x)\,dx \Big| \leq \int_{\mathbb{R}^n} |f(x)\varphi(x)|\,dx \leq \|f\|_{L^p} \|\varphi\|_{L^{p'}}.$$

Here $\|\varphi_j\|_{L^{p'}} \to 0$ whenever $\varphi_j \to 0$ in $\mathscr{S}$ by Example 3.1.2, so $T_f \in \mathscr{S}'$.

EXAMPLE 3.3.3. (Measures) Let μ be a positive or complex regular Borel measure on $\mathbb{R}^n$. We say that the measure μ is polynomially bounded if for some N the total variation $|\mu|$ satisfies

$$\int_{\mathbb{R}^n} \langle x \rangle^{-N} \, d|\mu|(x) < \infty.$$

An equivalent condition is that for any $M > 0$ the measure of the ball $B(0, M)$ satisfies $|\mu|(B(0, M)) \leq C\langle M \rangle^N$.

Any polynomially bounded measure μ gives rise to a tempered distribution defined by

$$T_\mu(\varphi) = \int_{\mathbb{R}^n} \varphi(x)\,d\mu(x),$$

since for any $\varphi \in \mathscr{S}$ one has

$$\begin{aligned} \Big| \int_{\mathbb{R}^n} \varphi(x)\,d\mu(x) \Big| &\leq \int_{\mathbb{R}^n} |\varphi(x)|\,d|\mu|(x) \\ &\leq \|\langle x \rangle^N \varphi(x)\|_{L^\infty} \int_{\mathbb{R}^n} \langle x \rangle^{-N}\,d|\mu|(x). \end{aligned}$$

It is also true that a positive measure which is in $\mathscr{S}'$ is necessarily polynomially bounded.

EXAMPLE 3.3.4. (Dirac measure) A particular case of the previous example is the measure δ defined by

$$\delta(E) = \begin{cases} 1, & 0 \in E, \\ 0, & \text{otherwise}. \end{cases}$$

This measure is called the *Dirac measure*, or slightly imprecisely *Dirac delta function*, and it satisfies

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

EXAMPLE 3.3.5. (Derivative of Dirac measure) An example of a tempered distribution that is not a measure is the linear functional

$$\delta' : \mathscr{S}(\mathbb{R}) \to \mathbb{C}, \quad \varphi \mapsto -\varphi'(0).$$

This is an element on $\mathscr{S}'(\mathbb{R})$. If δ' were a measure then one would have $|\varphi'(0)| \leq C\|\varphi\|_{L^\infty}$ for all $\varphi \in \mathscr{S}(\mathbb{R})$, which is impossible.

Thus, $\mathscr{S}'$ is a set that contains all polynomially bounded functions, measures and more. It turns out that most operations that are defined on test functions can also be defined for distributions by duality.

EXAMPLE 3.3.6. Consider the reflection operator on $\mathscr{S}$ that sends φ to $\tilde{\varphi}(x) = \varphi(-x)$. We wish to define the reflection of a distribution $T \in \mathscr{S}'$ as another distribution $\tilde{T}$. A reasonable requirement is that the operation should extend the reflection on $\mathscr{S}$, i.e. if $u \in \mathscr{S}$ then the reflection of T_u should be $T_{\tilde{u}}$. If this holds then we have

$$\tilde{T}_u(\varphi) = T_{\tilde{u}}(\varphi) = \int_{\mathbb{R}^n} u(-x)\varphi(x)\,dx = \int_{\mathbb{R}^n} u(x)\varphi(-x)\,dx = T_u(\tilde{\varphi}).$$

Motivated by this computation we *define* the reflection of $T \in \mathscr{S}'$ as the distribution $\tilde{T}$ given by

$$\tilde{T}(\varphi) = T(\tilde{\varphi}).$$

Here $\tilde{T}$ is continuous since the composition $\varphi \mapsto \tilde{\varphi} \mapsto T(\tilde{\varphi})$ is continuous from $\mathscr{S}$ to the scalars.

One may carry out similar computations as in the preceding example for the conjugation and translation to motivate the definitions $\overline{T}(\varphi) = \overline{T(\bar{\varphi})}$ and $(\tau_{x_0}T)(\varphi) = T(\tau_{-x_0}\varphi)$.

It is a remarkable fact that there is a natural notion of derivative on $\mathscr{S}'$. For $u \in \mathscr{S}$ the usual requirement that $\partial^\alpha T_u$ should be equal to $T_{\partial^\alpha u}$ leads to

$$\begin{aligned}(\partial^\alpha T_u)(\varphi) = T_{\partial^\alpha u}(\varphi) &= \int_{\mathbb{R}^n} (\partial^\alpha u)(x)\varphi(x)\,dx \\ &= (-1)^{|\alpha|}\int_{\mathbb{R}^n} u(x)\partial^\alpha\varphi(x)\,dx = (-1)^{|\alpha|}T_u(\partial^\alpha\varphi)\end{aligned}$$

where we have integrated repeatedly by parts.

DEFINITION. For any $T \in \mathscr{S}'$ we define the distribution $\partial^\alpha T$ by

$$(\partial^\alpha T)(\varphi) = (-1)^{|\alpha|}T(\partial^\alpha\varphi).$$

The distribution $\partial^\alpha T$ is called the αth *distributional derivative* or *weak derivative* of T.

Note that $\partial^\alpha T$ is a continuous linear functional on $\mathscr{S}$ since differentiation is continuous on $\mathscr{S}$. It follows that *any distribution has well defined derivatives of any order*, even if it arises from a function which is not differentiable in the classical sense. (The downside is that these derivatives are only defined in a weak sense, and saying anything more may require precise arguments that depend on the case at hand.) The definition of derivative also accommodates a form of integration by parts which is valid for distributions.

Example 3.3.7. If u is a C^k function whose derivatives up to order k are polynomially bounded, then the derivatives $\partial^\alpha u$ exist as continuous functions if $|\alpha| \leq k$. On the other hand, u gives rise to a distribution T_u, which has distributional derivatives $\partial^\alpha T_u$ for any α. It is an easy exercise to check that

$$\partial^\alpha T_u = T_{\partial^\alpha u}, \qquad |\alpha| \leq k,$$

showing that the distributional derivatives up to order k coincide with the corresponding classical derivatives.

Example 3.3.8. As an example of weak derivatives consider the function

$$u(x) = |x|, \qquad x \in \mathbb{R}.$$

Now u is not differentiable in the classical sense, but it determines a distribution T_u (below we write $u = T_u$) by

$$u : \mathscr{S} \to \mathbb{C}, \quad u(\varphi) = \int_{-\infty}^{\infty} |x|\varphi(x)\,dx.$$

The distribution u has a weak derivative given by

$$\begin{aligned} u'(\varphi) = -u(\varphi') &= -\int_{-\infty}^{0} (-x)\varphi'(x)\,dx - \int_0^\infty x\varphi'(x)\,dx \\ &= -\int_{-\infty}^{0} \varphi(x)\,dx + \int_0^\infty \varphi(x)\,dx \end{aligned}$$

where we have used integration by parts and the fact that $|\varphi(x)| \leq C_N \langle x\rangle^{-N}$ for any N since $\varphi \in \mathscr{S}$. Hence u' can be identified with the function

$$u'(x) = \begin{cases} -1, & x < 0, \\ 1, & x > 0. \end{cases}$$

Differentiation of u' leads to the distribution u'' with

$$u''(\varphi) = -u'(\varphi') = -\int_{-\infty}^{0} (-1)\varphi'(x)\,dx - \int_{0}^{\infty} (1)\varphi'(x)\,dx = 2\varphi(0).$$

Hence u'' is the measure 2δ. The derivative of the Dirac measure is given by

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

This explains the notation in Example 3.3.5.

For later purposes, we define a notion of convergence in $\mathscr{S}'$. There is a topology on $\mathscr{S}'$ (the so called weak* topology) which is compatible with this notion of convergence, but we will not specify the topology or use any of its properties.

Definition. If $(T_j)_{j=1}^{\infty} \subset \mathscr{S}'$ and $T \in \mathscr{S}'$, we say that $T_j \to T$ in $\mathscr{S}'$ if

$$T_j(\varphi) \to T(\varphi) \text{ for any } \varphi \in \mathscr{S}.$$

The next simple result states that limits in $\mathscr{S}'$ are unique, and that convergence in $\mathscr{S}$ or L^p implies convergence in $\mathscr{S}'$.

Lemma 3.3.1. *(Convergence in $\mathscr{S}'$)*

(a) *If $T_j \to T$ in $\mathscr{S}'$ and $T_j \to S$ in $\mathscr{S}'$, then $T = S$.*

(b) *If (φ_j) is a sequence in $\mathscr{S}$ or L^p ($1 \le p \le \infty$) with $\varphi_j \to \varphi$ in $\mathscr{S}$ or L^p, then $\varphi_j \to \varphi$ in $\mathscr{S}'$.*

The operations on Schwartz functions given in Theorem 3.1.4 extend to tempered distributions by duality. Perhaps the most striking point is that *any tempered distribution, no matter how irregular, has distributional derivatives of any order*, and these derivatives are still tempered distributions.

Theorem 3.3.2. *(Operations on tempered distributions) Let f be a function in $\mathscr{O}_M(\mathbb{R}^n)$. The following operations map $\mathscr{S}'$ into $\mathscr{S}'$, and they extend the corresponding operations on $\mathscr{S}$:*

(1)	$\tilde{T}(\varphi) = T(\tilde{\varphi})$	*(reflection)*
(2)	$\overline{T}(\varphi) = \overline{T(\overline{\varphi})}$	*(conjugation)*
(3)	$(\tau_{x_0}T)(\varphi) = T(\tau_{-x_0}\varphi)$	*(translation)*
(4)	$(\partial^\alpha T)(\varphi) = (-1)^{\lvert\alpha\rvert} T(\partial^\alpha \varphi)$	*(derivative)*
(5)	$(fT)(\varphi) = T(f\varphi)$	*(multiplication)*

PROOF. Follows from Theorem 3.1.4. □

We have seen that the polynomially bounded continuous functions and their weak derivatives are among the tempered distributions. The structure theorem for tempered distributions says that there are no others. First we observe a result that is of interest in its own right.

LEMMA 3.3.3. *(Any tempered distribution has finite order) For any* $T \in \mathscr{S}'$ *there exist* $C > 0$ *and* $N \in \mathbb{N}$ *such that*

$$|T(\varphi)| \le C \sum_{|\beta| \le N} \|\langle x\rangle^N \partial^\beta \varphi\|, \qquad \varphi \in \mathscr{S}.$$

REMARK 3.3.4. If T and N are as in the lemma, we say that *the distribution* T *has order* N. The lemma states that any tempered distribution has finite order. It is also true that the distributions that have order 0 are exactly the bounded measures (this is a consequence of the Riesz representation theorem in measure theory). This lemma is the first place where we use in an essential way the fact that distributions are *continuous* linear functionals, instead of just linear functionals.

PROOF OF LEMMA 3.3.3. Let $T \in \mathscr{S}'$. We argue by contradiction and assume that for any $N > 0$ there is $\varphi_N \in \mathscr{S}$ such that

$$|T(\varphi_N)| \ge N \sum_{|\alpha| \le N} \|\langle x\rangle^N \partial^\alpha \varphi_N\|_{L^\infty}.$$

Define

$$\psi_N := \frac{1}{N}\left(\sum_{|\alpha| \le N} \|\langle x\rangle^N \partial^\alpha \varphi_N\|_{L^\infty}\right)^{-1} \varphi_N.$$

For any fixed $K \in \mathbb{N}$ and $\beta \in \mathbb{N}^n$, if $N \ge \max(K, |\beta|)$ we have

$$\|\langle x\rangle^K \partial^\beta \psi_N\|_{L^\infty} \le \|\langle x\rangle^N \partial^\beta \psi_N\|_{L^\infty} \le \frac{1}{N}..$$

Thus for each K and β, $\langle x\rangle^K \partial^\beta \psi_N \to 0$ uniformly as $N \to \infty$, which shows that $\psi_N \to 0$ in $\mathscr{S}$. Since T is a continuous linear functional we also have

$$T(\psi_N) \to 0 \quad \text{as } N \to \infty.$$

But $|T(\psi_N)| \ge 1$ for all N by the inequality above, which gives a contradiction. □

THEOREM 3.3.5. *(Structure theorem for tempered distributions) Any* $T \in \mathscr{S}'(\mathbb{R}^n)$ *can be written as*

$$T = \partial^\alpha f$$

for some $\alpha \in \mathbb{N}^n$ *and some polynomially bounded continuous function* f.

PROOF. We only give the proof when $n = 1$. Let $T \in \mathscr{S}'(\mathbb{R})$ and let N be an even integer such that for all $\varphi \in \mathscr{S}(\mathbb{R})$ one has

$$|T(\varphi)| \le C \sum_{\beta=0}^{N-1} \|\langle x\rangle^N \partial^\beta \varphi\|_{L^\infty} \tag{3.12}$$

where C and N do not depend on φ. Let x_0 be a point where the function $|\langle x\rangle^N \partial^\beta \varphi(x)|$ attains its maximum. If $x_0 < 0$ choose $I = (-\infty, x_0)$, otherwise take $I = (x_0, \infty)$. We obtain the estimate

$$\begin{aligned}
\|\langle x\rangle^N \partial^\beta \varphi(x)\|_{L^\infty} = \langle x_0\rangle^N |\partial^\beta \varphi(x_0)| &= \langle x_0\rangle^N \Big| \int_I \partial^{\beta+1}\varphi(x)\, dx \Big| \\
&\le \int_I \langle x\rangle^N |\partial^{\beta+1}\varphi(x)|\, dx \\
&\le \int_{-\infty}^{\infty} \langle x\rangle^N |\partial^{\beta+1}\varphi(x)|\, dx.
\end{aligned}$$

It is convenient to introduce the weighted L^1 space $L^1_w = L^1(\mathbb{R}, d\mu)$ where $d\mu(x) = \langle x\rangle^N\, dx$. Using the last estimate in (3.12) gives that

$$|T(\varphi)| \le C \sum_{\beta=0}^{N} \|\partial^\beta \varphi\|_{L^1_w} \tag{3.13}$$

which is valid for all φ in $\mathscr{S}$.

Denote by $\mathscr{L}$ the direct sum $L^1_w \oplus \ldots \oplus L^1_w$ ($N+1$ times). The space $\mathscr{L}$ becomes a Banach space with norm

$$\|(\varphi_0, \ldots, \varphi_N)\| = \|\varphi_0\|_{L^1_w} + \ldots + \|\varphi_N\|_{L^1_w},$$

and there is an injective map

$$\pi : \mathscr{S} \to \mathscr{L}, \quad \varphi \mapsto (\varphi, \varphi', \ldots, \varphi^{(N)})$$

from $\mathscr{S}$ onto $\pi(\mathscr{S}) \subset \mathscr{L}$. Define

$$U : \pi(\mathscr{S}) \to \mathbb{C}, \quad U(\varphi, \varphi', \ldots, \varphi^{(N)}) = T(\varphi).$$

According to (3.13) the map U can be interpreted as a bounded linear functional on $\pi(\mathscr{S}) \subset \mathscr{L}$, and hence has a continuous extension into all of $\mathscr{L}$ by the Hahn-Banach theorem. The extended U splits into bounded linear functionals U_j on L^1_w so that

$$U(\varphi_0, \ldots, \varphi_N) = U_0(\varphi_0) + \ldots + U_N(\varphi_N).$$

Since the dual of L^1_w is $L^\infty(\mathbb{R}, d\mu) = L^\infty(\mathbb{R}, dx)$, any bounded linear functional on L^1_w is of the form

$$S(\varphi) = \int_{-\infty}^{\infty} f\varphi \, d\mu$$

for some $f \in L^\infty(\mathbb{R})$. Thus each U_j is of the form

$$U_j(\varphi) = \int_{-\infty}^{\infty} \langle x\rangle^N b_j(x)\varphi(x)\, dx$$

where b_j is some function in $L^\infty(\mathbb{R})$. Define new functions $h_N(x) = b_N(x)$ and for $1 \le j \le N$

$$h_{N-j}(x) = \int_0^x \int_0^{x_2} \cdots \int_0^{x_j} \langle t\rangle^N b_{N-j}(t)\, dt\, dx_j \cdots dx_2.$$

Now each h_j is polynomially bounded, since

$$|h_{N-j}(x)| \le \|b_{N-j}\|_{L^\infty} \int \cdots \int \langle t\rangle^N$$

where the last integral is a polynomial (recall that N was even). Repeated integrations by parts give that

$$T(\varphi) = U_0(\varphi) + U_1(\varphi') + \ldots + U_N(\varphi^{(N)}) = \int_{-\infty}^{\infty} h(x)\varphi^{(N)}\, dx$$

where the function h is a linear combination of the h_j, hence it is polynomially bounded. One more integration by parts shows that h may be taken continuous if $\varphi^{(N)}$ is replaced by $\varphi^{(N+1)}$. □

3.4. The Fourier transform of tempered distributions

Parseval's identity (Theorem 3.2.7, part (2)) shows that the following definition extends the Fourier transform on $\mathscr{S}$.

Definition. The Fourier transform of any tempered distribution $T \in \mathscr{S}'$ is the tempered distribution $\hat{T} = \mathscr{F}T$ defined by

$$\hat{T}(\varphi) = T(\hat{\varphi}).$$

Similarly, the inverse Fourier transform of $T \in \mathscr{S}'$ is the distribution $\check{T} = \mathscr{F}^{-1}T$ for which $\check{T}(\varphi) = T(\check{\varphi})$.

The composition $\varphi \mapsto \hat{\varphi} \mapsto T(\hat{\varphi})$ is continuous so $\hat{T}$ and also $\check{T}$ are indeed tempered distributions.

EXAMPLE 3.4.1. (Dirac measure) The Fourier transform of the Dirac measure δ_{x_0} is the tempered distribution given by

$$\hat{\delta}_{x_0}(\varphi) = \delta_{x_0}(\hat{\varphi}) = \hat{\varphi}(x_0) = \int_{\mathbb{R}^n} e^{-ix_0\cdot y}\varphi(y)\,dy.$$

Thus $\hat{\delta}_{x_0}$ is the function $\xi \mapsto e^{-ix_0\cdot\xi}$. In particular, $\hat{\delta}_0 = 1$.

EXAMPLE 3.4.2. (Derivative of Dirac measure)

$$(D^\alpha\delta_0)\hat{}\,(\varphi) = D^\alpha\delta_0(\hat{\varphi}) = (-1)^{|\alpha|}\delta_0(D^\alpha\hat{\varphi}) = \delta_0((x^\alpha\varphi)\hat{}\,) = \int_{\mathbb{R}^n} x^\alpha\varphi(x)\,dx.$$

Thus $(D^\alpha\delta_0)\hat{}\,(\xi) = \xi^\alpha$.

EXAMPLE 3.4.3. (Dirac comb) If $a > 0$, define

$$\Delta_a = \sum_{k\in\mathbb{Z}^n} \delta_{ak}.$$

It is not difficult to check that Δ_a is a tempered distribution. The Poisson summation formula from the exercises,

$$\sum_{k\in\mathbb{Z}^n} \hat{\varphi}(ak) = (2\pi/a)^n \sum_{k\in\mathbb{Z}^n} \varphi(2\pi k/a), \qquad \varphi \in \mathscr{S}(\mathbb{R}^n),$$

implies that

$$\hat{\Delta}_a = (2\pi/a)^n\Delta_{2\pi/a}.$$

As for the Schwartz space, the Fourier transform is an isomorphism of the dual space $\mathscr{S}'$.

THEOREM 3.4.1. *(Fourier transform on tempered distributions) The Fourier transform is a bijective map from $\mathscr{S}'(\mathbb{R}^n)$ onto $\mathscr{S}'(\mathbb{R}^n)$. It is continuous in the sense that*

$$T_j \to T \text{ in } \mathscr{S}' \implies \hat{T}_j \to \hat{T} \text{ in } \mathscr{S}'.$$

One has the inversion formula

$$\mathscr{F}^{-1}\mathscr{F}T = \mathscr{F}\mathscr{F}^{-1}T = T, \qquad T \in \mathscr{S}'.$$

PROOF. Clearly $\mathscr{F} : \mathscr{S}' \to \mathscr{S}'$ is linear, and the continuity follows since $\mathscr{F}$ is continuous on Schwartz space. The inversion formula is a consequence of the corresponding formula on $\mathscr{S}$ since $\mathscr{F}^{-1}\mathscr{F}T(\varphi) = \hat{T}(\check{\varphi}) = T(\varphi)$. The proof that $\mathscr{F}\mathscr{F}^{-1}T = T$ is analogous, and thus $\mathscr{F}$ is bijective. $\square$

Note that the identities $\mathscr{F}^2 f = (2\pi)^n \tilde{f}$ and $\mathscr{F}^4 f = (2\pi)^{2n} f$ hold also on $\mathscr{S}'$.

THEOREM 3.4.2. *Let $x_0, \xi_0 \in \mathbb{R}^n$, let α and β be multi-indices, and let f be a function in $\mathscr{S}(\mathbb{R}^n)$. Then the Fourier transform on $\mathscr{S}'(\mathbb{R}^n)$ has the following properties.*

(1) $(\tau_{x_0} T)\hat{} = e^{-ix_0\cdot\xi}\hat{T}$ *(translation)*

(2) $(e^{i\xi_0\cdot x} T)\hat{} = \tau_{\xi_0}\hat{T}$ *(modulation)*

(3) $(D^\alpha T)\hat{} = \xi^\alpha \hat{T}$ *(derivative)*

(4) $((-x)^\beta T)\hat{} = D^\beta \hat{T}$ *(polynomial)*

PROOF. Follows from the definitions and the corresponding result on $\mathscr{S}$. □

We now give some classical theorems on the Fourier transform by restricting the Fourier transform on $\mathscr{S}'$ to certain special cases. For the first theorem, let

$$C_0(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{C}\,;\ f \text{ continuous and } f(x) \to 0 \text{ as } x \to \infty\}.$$

We equip $C_0(\mathbb{R}^n)$ with the $L^\infty(\mathbb{R}^n)$ norm, and then $C_0(\mathbb{R}^n)$ is a Banach space.

THEOREM 3.4.3. (Riemann-Lebesgue) *The Fourier transform is a continuous map from $L^1(\mathbb{R}^n)$ into $C_0(\mathbb{R}^n)$. For any $f \in L^1$ the Fourier transform is given by the usual formula*

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)\,dx, \qquad \xi \in \mathbb{R}^n. \tag{3.14}$$

THEOREM 3.4.4. (Plancherel) *The Fourier transform is an isomorphism from $L^2(\mathbb{R}^n)$ onto $L^2(\mathbb{R}^n)$. It is isometric in the sense that*

$$\|\hat{f}\|_{L^2} = (2\pi)^{n/2}\|f\|_{L^2}.$$

The transform is given by

$$\hat{f}(\xi) = \operatorname*{l.i.m}_{R\to\infty} \int_{|x|\le R} e^{-ix\cdot\xi} f(x)\,dx \tag{3.15}$$

where l.i.m *means that the limit is in L^2.*

THEOREM 3.4.5. *(Hausdorff-Young) If $1 \le p \le 2$, the Fourier transform is a continuous map from $L^p(\mathbb{R}^n)$ to $L^{p'}(\mathbb{R}^n)$ where $\frac{1}{p}+\frac{1}{p'} = 1$. Moreover,*

$$\|\hat{f}\|_{L^{p'}} \le (2\pi)^{n/p'}\|f\|_{L^p}, \qquad f \in L^p.$$

To complement the above results, we mention that the range of the Fourier transform on $L^p(\mathbb{R}^n)$ with $p > 2$ is a subset of $\mathscr{S}'(\mathbb{R}^n)$ which contains distributions that are not measures (see [Hö, Section 7.6]).

The proofs rely on two results. The first is an approximation result to be proved later in the section concerning convolution:

LEMMA 3.4.6. *$\mathscr{S}(\mathbb{R}^n)$ is a dense subspace of $L^p(\mathbb{R}^n)$ if $1 \leq p < \infty$.*

The other result is a basic functional analysis fact, sometimes known as the BLT (bounded linear transformation) theorem.

THEOREM 3.4.7. *(BLT theorem) Let X and Y be Banach spaces and let X_0 be a dense subspace of X. If $T : X_0 \to Y$ is a linear map that satisfies*

$$\|T(x)\|_Y \leq C\|x\|_X, \qquad x \in X_0,$$

then there is a unique bounded linear map $\bar{T} : X \to Y$ with $\bar{T}|_{X_0} = T$. Moreover,

$$\|\bar{T}(x)\|_Y \leq C\|x\|_X, \qquad x \in X,$$

and $\bar{T}(x) = \lim_{j\to\infty} T(x_j)$ whenever $(x_j) \subset X_0$ and $x_j \to x$ in X.

PROOF. If $x \in X$, we would like to define $\bar{T}(x)$ as in the last line of the statement of the theorem. If $(x_j) \subset X_0$ and $x_j \to x$ in X, then

$$\|T(x_j) - T(x_k)\|_Y \leq C\|x_j - x_k\|_X$$

and thus $T(x_j)$ is a Cauchy sequence in Y, hence it converges to some $y \in Y$ by completeness. We define $\bar{T}(x) = y$. The definition is independent of the choice of the sequence converging to x, since if $(x_j') \subset X_0$ is another sequence with $x_j' \to x$ in X, then $\|T(x_j) - T(x_j')\|_Y \leq C\|x_j - x_j'\|_X \to 0$ as $j \to \infty$ because both (x_j) and (x_j') converge to x. Thus also $T(x_j') \to y$.

It is easy to check that $\bar{T}$ is a bounded linear map $X \to Y$ with norm bounded by C, and it is the unique continuous extension of T since X_0 was dense. □

PROOF OF THEOREM 3.4.3. If $f \in \mathscr{S}$ then we already know that $\hat{f} \in C_0$ and $\|\hat{f}\|_{L^\infty} \leq \|f\|_{L^1}$. This means that $\mathscr{F} : \mathscr{S} \subset L^1 \to C_0$ is a bounded linear map from a dense subspace of L^1 to C_0, hence has a unique bounded extension $\Phi : L^1 \to C_0$ with $\|\Phi(f)\|_{L^\infty} \leq \|f\|_{L^1}$.

We wish to show that $\Phi = \mathscr{F}|_{L^1}$ where $\mathscr{F}$ is the Fourier transform on $\mathscr{S}'$. For this we take any $f \in L^1$ and choose a sequence $(f_j) \subset \mathscr{S}$ such that $f_j \to f$ in L^1. Then $\mathscr{F} f_j \to \Phi(f)$ in L^∞, hence also in $\mathscr{S}'$, but also $\mathscr{F} f_j \to \mathscr{F} f$ in $\mathscr{S}'$ by Theorem 3.4.1. Since limits in $\mathscr{S}'$ are unique, we have $\Phi(f) = \mathscr{F}(f)$ as distributions. The formula (3.14) is given by

$$\Phi(f)(\xi) = \lim_{j\to\infty} \hat{f}_j(\xi) = \lim_{j\to\infty} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f_j(x)\, dx = \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x)\, dx$$

where the last equality follows since $\|f_j - f\|_{L^1} \to 0$. $\square$

Proof of Theorem 3.4.4. If $f \in \mathscr{S}$ then $\hat{f} \in \mathscr{S}$ and $\|\hat{f}\|_{L^2} = (2\pi)^{n/2}\|f\|_{L^2}$ by Parseval's identity. Thus $\mathscr{F} : \mathscr{S} \subset L^2 \to L^2$ is an isometry from a dense subspace of L^2 to L^2 and extends uniquely into an isometry $\Phi : L^2 \to L^2$.

It follows from Schwarz's inequality and a similar argument as in the proof of the preceding theorem that Φ and $\mathscr{F}|_{L^2}$ coincide. For (3.15) let $B_R = B(0, R)$ and let χ_{B_R} be the characteristic function. Then for any $f \in L^2$ we have $\chi_{B_R} f \to f$ in L^2 as $R \to \infty$, thus $(\chi_{B_R} f)\hat{} \to \hat{f}$ in L^2 by what we have already proved. This gives

$$\hat{f}(\xi) = \underset{R\to\infty}{\text{l.i.m}}\, (\chi_{B_R} f)\hat{}(\xi) = \underset{R\to\infty}{\text{l.i.m.}} \int_{|x|\le R} e^{-ix\cdot\xi} f(x)\, dx,$$

the last equation coming from the preceding theorem since $\chi_{B_R} f$ is in L^1. $\square$

Proof of Theorem 3.4.5. The Riemann-Lebesgue and Plancherel theorems imply that $\mathscr{F}$ is a bounded linear map

$$\mathscr{F} : L^1 \to L^\infty, \qquad \|\mathscr{F} f\|_{L^\infty} \le \|f\|_{L^1},$$
$$\mathscr{F} : L^2 \to L^2, \qquad \|\mathscr{F} f\|_{L^2} \le (2\pi)^{n/2}\|f\|_{L^2}.$$

The result follows from these facts and the Riesz-Thorin interpolation theorem. $\square$

3.5. Compactly supported distributions

To study the local behaviour of tempered distributions we introduce the following concepts.

Definition. For any open set $V \subset \mathbb{R}^n$ the distribution $T \in \mathscr{S}'(\mathbb{R}^n)$ is said to *vanish* on V, written $T = 0$ on V, if $T(\varphi) = 0$ for any $\varphi \in C_c^\infty(V)$. Two distributions T_1 and T_2 are said to be *equal* on V if $T_1 - T_2 = 0$ on V.

LEMMA 3.5.1. *If $\{V_j\}_{j\in J}$ is a family of open sets in $\mathbb{R}^n$, and if T vanishes on each V_j, then T vanishes on $\bigcup_{j\in J} V_j$.*

PROOF. Let $V = \bigcup_{j\in J} V_j$. We use a locally finite partition of unity subordinate to $\{V_j\}_{j\in J}$ (see [Ru, Theorem 6.20]). This is a family of functions $\{\psi_j\}_{j\in J}$ with $\psi_j \in C^\infty(V_j)$, $0 \le \psi_j \le 1$, such that any compact set $K \subset V$ has a neighborhood U where only finitely many ψ_j are not identically zero, and

$$\sum_{j\in J} \psi_j(x) = 1 \qquad \text{for } x \in U.$$

Let $\varphi \in C_c^\infty(V)$, and write $K = \operatorname{supp}(\varphi)$. We can now write

$$\varphi = \sum_{j\in J} \psi_j \varphi$$

where only finitely many terms of the sum are nonzero. Thus

$$T(\varphi) = \sum_{j\in J} T(\psi_j \varphi) = 0,$$

using the fact that T vanishes on each V_j. □

DEFINITION. The *support* of a distribution $T \in \mathscr{S}'(\mathbb{R}^n)$, denoted by $\operatorname{supp}(T)$, is the complement of the largest open subset of $\mathbb{R}^n$ where T vanishes.

The definition makes sense since if a distribution vanishes on open sets $\{V_j\}$ then it vanishes on $\bigcup V_j$ by Lemma 3.5.1. It follows that $x \in \operatorname{supp}(T)$ if and only if T does not vanish on any neighborhood of x. Easy consequences of the definition are that $T(\varphi) = 0$ whenever $\varphi \in C_c^\infty(\mathbb{R}^n)$ and $\operatorname{supp}(\varphi) \cap \operatorname{supp}(T) = \emptyset$, and if $\psi \in \mathscr{O}_M(\mathbb{R}^n)$ is such that $\psi|_V = 1$ for some neighborhood V of $\operatorname{supp}(T)$ then $\psi T = T$.

It will be convenient to have a characterization of the tempered distributions with compact support. These have a natural connection with the space $\mathscr{E}(\mathbb{R}^n)$ which we now define.

DEFINITION. The space $\mathscr{E}(\mathbb{R}^n) = C^\infty(\mathbb{R}^n)$ is given the topology induced by the seminorms

$$\|f\|_N = \sum_{|\alpha|\le N} \|\partial^\alpha f\|_{L^\infty(B(0,N))}, \qquad N \in \mathbb{N}.$$

We denote by $\mathscr{E}'(\mathbb{R}^n)$ the set of continuous linear functionals on $\mathscr{E}$.

Theorem 3.1.2 implies that $\mathscr{E}$ is a metric space, and that $f_j \to f$ in $\mathscr{E}$ if and only if $\partial^\alpha f_j \to \partial^\alpha f$ uniformly on compact subsets of $\mathbb{R}^n$ for any $\alpha \in \mathbb{N}^n$.

LEMMA 3.5.2. *$\mathscr{E}(\mathbb{R}^n)$ is a complete metric space, and the identity map $i : \mathscr{S}(\mathbb{R}^n) \to \mathscr{E}(\mathbb{R}^n)$ is continuous.*

The continuity of the identity map $\mathscr{S} \to \mathscr{E}$ shows that any continuous linear functional S on $\mathscr{E}$ (that is, any $S \in \mathscr{E}'$) gives rise to a distribution $T \in \mathscr{S}'$ where $T = S \circ i$. On the other hand $\mathscr{S}$ is dense in $\mathscr{E}$ (for any $f \in \mathscr{E}$ just take a sequence (f_j) in $\mathscr{S}$ such that $f_j = f$ on $B(0, j)$), so two distinct elements of $\mathscr{E}'$ give different distributions in $\mathscr{S}'$. We may thus identify $\mathscr{E}'$ with a certain subspace of $\mathscr{S}'$; this subspace is exactly the set of distributions with compact support.

THEOREM 3.5.3. *(Compactly supported distributions) If $T \in \mathscr{S}'$, then T has compact support if and only if T can be extended into a continuous linear functional on $\mathscr{E}$.*

PROOF. Suppose T has compact support, and choose $\psi \in C_c^\infty(\mathbb{R}^n)$ so that $\psi = 1$ on some open set containing $\operatorname{supp}(T)$. Denote the support of ψ by K. Then $T(\varphi) = T(\psi\varphi)$ for all $\varphi \in \mathscr{S}$, and we can extend T into $\mathscr{E}$ by defining $T(f) = T(\psi f)$ for $f \in \mathscr{E}$. Since $T \in \mathscr{S}'$, there exist C and N such that

$$|T(\varphi)| \leq C \sum_{|\alpha| \leq N} \|\langle x\rangle^N \partial^\alpha \varphi\|_{L^\infty}, \qquad \varphi \in \mathscr{S}.$$

This implies that

$$|T(\varphi)| \leq C' \sum_{|\alpha| \leq N} \|\partial^\alpha \varphi\|_{L^\infty}, \qquad \varphi \in C_c^\infty(\mathbb{R}^n) \text{ with } \operatorname{supp}(\varphi) \subset K.$$

Now for any $f \in \mathscr{E}$ the function ψf is supported in K and we have

$$|T(f)| = |T(\psi f)| \leq C' \sum_{|\alpha| \leq N} \|\partial^\alpha f\|_{L^\infty}. \tag{3.16}$$

This implies that T is continuous on $\mathscr{E}$ and we have the desired extension.

For the converse we suppose that T is a continuous linear functional on $\mathscr{E}$, so there are C and N such that

$$|T(f)| \leq C \sum_{|\alpha| \leq N} \|\partial^\alpha f\|_{L^\infty(B(0,N))}, \qquad f \in \mathscr{E}.$$

If T does not have compact support, then for any M there is a function $\varphi \in C_c^\infty(\mathbb{R}^n \setminus \overline{B(0,M)})$ for which $T(\varphi) \neq 0$. This clearly contradicts the above inequality. $\square$

The next theorem characterizes all distributions with support consisting of one point.

THEOREM 3.5.4. *(Distributions supported at a point) If $T \in \mathscr{S}'$ and* $\operatorname{supp}(T) = \{x_0\}$*, then there are constants N and C_α such that*

$$T = \sum_{|\alpha| \leq N} C_\alpha \partial^\alpha \delta_{x_0}.$$

PROOF. [Ru], p. 165. $\square$

As a consequence, we obtain a generalization of the standard Liouville theorem which states that any bounded harmonic function is constant.

THEOREM 3.5.5. *(Liouville theorem for distributions) If $u \in \mathscr{S}'(\mathbb{R}^n)$ satisfies $\Delta u = 0$ in the sense of distributions, then u is a polynomial.*

PROOF. Taking Fourier transforms in the equation $\Delta u = 0$ implies that $|\xi|^2 \hat{u} = 0$. If $\varphi \in C_c^\infty(\mathbb{R}^n)$ vanishes near 0, also the function $|\xi|^{-2}\varphi(\xi)$ is in $C_c^\infty(\mathbb{R}^n)$ and

$$\langle \hat{u}, \varphi \rangle = \langle |\xi|^2 \hat{u}, |\xi|^{-2} \varphi \rangle = 0.$$

Thus $\operatorname{supp}(\hat{u}) = \{0\}$, and Theorem 3.5.4 implies that

$$\hat{u} = \sum_{|\alpha| \leq N} C_\alpha \partial^\alpha \delta_0.$$

Taking the inverse Fourier transform, we see that u is a polynomial. $\square$

It is natural that in the structure theorem for compactly supported distributions, compactly supported continuous functions appear.

THEOREM 3.5.6. *(Structure theorem for $\mathscr{E}'$) If $T \in \mathscr{E}'(\mathbb{R}^n)$ and if $V \subset \mathbb{R}^n$ is any open set containing* $\operatorname{supp}(T)$*, there exist $N \in \mathbb{N}$ and functions $f_\alpha \in C_c(V)$ such that*

$$T = \sum_{|\alpha| \leq N} \partial^\alpha f_\alpha.$$

PROOF. See [Sc, Section III.7]. $\square$

EXAMPLE 3.5.1. We illustrate the structure theorem in the case of the compactly supported distribution T_f, where $f(x) = \chi_{(0,1)}(x)$ is the characteristic function of the unit interval. Define

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

Then F is a tempered distribution, and $F' = f$ in the sense of distributions. Let $\psi \in C_c^\infty(\mathbb{R})$ satisfy $\psi = 1$ near $[0,1]$. Then for $\varphi \in \mathscr{E}(\mathbb{R})$ we have

$$\begin{aligned} T_f(\varphi) = T_f(\psi\varphi) &= \langle F', \psi\varphi \rangle = -\langle F, \psi'\varphi + \psi\varphi' \rangle \\ &= \langle -\psi' F + (\psi F)', \varphi \rangle. \end{aligned}$$

This shows that $T_f = f_0 + f_1'$ where $f_0 = -\psi' F$ and $f_1 = \psi F$ are continuous compactly supported functions in $\mathbb{R}$.

The structure theorem easily implies that the Fourier transform of any compactly supported distribution is actually a smooth function in $\mathbb{R}^n$. This illustrates the fact that the Fourier transform exchanges decay properties with smoothness.

THEOREM 3.5.7. *The Fourier transform of any $T \in \mathscr{E}'(\mathbb{R}^n)$ is the function*

$$\hat{T}(\xi) = T(e^{-ix\cdot\xi}), \qquad \xi \in \mathbb{R}^n.$$

More precisely, if $T \in \mathscr{E}'$ then $\hat{T} = T_F$ where F is the function in $\mathscr{O}_M$ defined by $F(\xi) = T(e^{-ix\cdot\xi})$.

PROOF. Let $T \in \mathscr{E}'$, and use the structure theorem in order to write $T = \sum_{|\alpha|\leq N} D^\alpha f_\alpha$ where $f_\alpha \in C_c(\mathbb{R}^n)$. Then by properties of the Fourier transform,

$$\begin{aligned} \hat{T}(\varphi) = T(\hat{\varphi}) &= \sum_{|\alpha|\leq N} \langle f_\alpha, (-D)^\alpha \hat{\varphi} \rangle = \sum_{|\alpha|\leq N} \langle f_\alpha, (\xi^\alpha \varphi)\hat{\ } \rangle \\ &= \langle \sum_{|\alpha|\leq N} \xi^\alpha \hat{f}_\alpha, \varphi \rangle. \end{aligned}$$

But also

$$F(\xi) = \langle \sum_{|\alpha|\leq N} D^\alpha f_\alpha, e^{-ix\cdot\xi} \rangle = \langle \sum_{|\alpha|\leq N} f_\alpha, \xi^\alpha e^{-ix\cdot\xi} \rangle = \sum_{|\alpha|\leq N} \xi^\alpha \hat{f}_\alpha(\xi).$$

This shows that $\hat{T}(\varphi) = \langle F, \varphi \rangle$, and it is not difficult to check that $F \in \mathscr{O}_M$. $\square$

Finally, we remark that the range of the Fourier transform on $C_c^\infty(\mathbb{R}^n)$ and $\mathscr{E}'(\mathbb{R}^n)$ can be completely characterized via the Paley-Wiener and Paley-Wiener-Schwartz theorems.

3.6. The test function space $\mathscr{D}$

The test function space $\mathscr{S}$, and the corresponding space of tempered distributions $\mathscr{S}'$, are objects that are naturally defined on the whole space $\mathbb{R}^n$. The requirement that the space $\mathscr{S}'$ should have a reasonable Fourier analysis is reflected in the decay properties of Schwartz functions at infinity. We will next consider a distribution space $\mathscr{D}'(\mathbb{R}^n)$ which is larger than $\mathscr{S}'(\mathbb{R}^n)$ and which is completely local: for elements in $\mathscr{D}'$ the behavior at infinity does not play any role, and if Ω is any open subset of $\mathbb{R}^n$ there is a natural corresponding space $\mathscr{D}'(\Omega)$, the set of distributions in Ω.

The test functions for $\mathscr{D}'$ have compact support so that any locally integrable function becomes a distribution, and they are infinitely differentiable to ensure that also the corresponding distributions will have derivatives of any order. The topology on this space will be taken so fine that it is not a harsh requirement for linear functionals to be continuous. However, the topology on the test function space will be more complicated than for $\mathscr{S}$ or $\mathscr{E}$ for instance. In particular, it will not be a metric space topology. We begin by defining the spaces $\mathscr{D}_K$.

Definition. If $K \subset \mathbb{R}^n$ is a compact set, we denote by $\mathscr{D}_K$ the set of all C^∞ complex functions on $\mathbb{R}^n$ with support contained in K. The topology on $\mathscr{D}_K$ is taken to be the one given by the norms

$$\|\varphi\|_N = \sum_{|\alpha| \leq N} \|\partial^\alpha \varphi\|_{L^\infty(\mathbb{R}^n)} \tag{3.17}$$

where $N \geq 0$ is an integer.

Lemma 3.6.1. *$\mathscr{D}_K$ is a complete metric space.*

Proof. This follows as before. □

Having established a topology for the spaces $\mathscr{D}_K$, the next step is to consider the space $\mathscr{D}(\Omega)$ of all compactly supported C^∞ functions on an open set $\Omega \subset \mathbb{R}^n$. Thus

$$\mathscr{D}(\Omega) = C_c^\infty(\Omega) = \bigcup_{K \subset \Omega \text{ compact}} \mathscr{D}_K.$$

In the following, it may be useful to consider an *exhaustion of* Ω *by compact subsets*. This means a family $\{K_m\}_{m=1}^{\infty}$ of compact subsets of Ω so that $K_m \subset K_{m+1}^{\circ}$ and $\bigcup K_m = \Omega$. One can take

$$K_m := \Omega \setminus \left(\{x\,;\,|x| > m\} \cup \bigcup_{z \in \mathbb{R}^n \setminus \Omega} B(z, 1/m) \right).$$

Imitating our previous arguments for the other test function spaces, one could try to give $\mathscr{D}(\Omega)$ the topology induced by the countable family of seminorms

$$\|\varphi\|_N = \sum_{|\alpha| \leq N} \|\partial^\alpha \varphi\|_{L^\infty(K_N)}, \qquad N \in \mathbb{N},$$

where $\{K_m\}$ is an exhaustion of Ω as above. This topology however has one immediate handicap: it is not complete. The problem is that although any Cauchy sequence converges with respect to all the seminorms to a C^∞ function, the limit function need not have compact support in Ω.

The situation can be remedied by considering $\mathscr{D}(\Omega)$ as a *strict inductive limit* of the spaces $\mathscr{D}_K$, where the compact sets K increase toward Ω. We will not give the details of this construction, but rather only state some of its properties. It is shown in [**Sc**] (in the case where $\Omega = \mathbb{R}^n$) that the topology on $\mathscr{D}(\Omega)$ is determined by the uncountable family of seminorms

$$\rho_{(\varepsilon_m),(r_m)}(\varphi) = \sup_{m \geq 0} \sup_{\substack{x \notin K_m^{\circ} \\ |\alpha| \leq r_m}} |D^\alpha \varphi(x)|/\varepsilon_m$$

where (ε_m) is a decreasing sequence of positive numbers with limit 0 and (r_m) is an increasing sequence of natural numbers converging to ∞. (We define $K_0 = \emptyset$.)

THEOREM 3.6.2. *There exists a topology on $\mathscr{D}(\Omega)$ which is a vector space topology (that is, addition and scalar multiplication are continuous operations) and has the following properties:*

(a) *A sequence (φ_j) in $\mathscr{D}(\Omega)$ converges if and only if $(\varphi_j) \subset \mathscr{D}_K$ for some fixed compact set $K \subset \Omega$ and (φ_j) converges in $\mathscr{D}_K$.*
(b) *$\mathscr{D}(\Omega)$ is complete (any Cauchy sequence or net in $\mathscr{D}(\Omega)$ converges).*

PROOF. See [**Ru**] or [**Sc**]. □

Although $\mathscr{D}(\Omega)$ is not metrizable (it is the countable union of the spaces $\mathscr{D}_{K_m}$ which are nowhere dense in $\mathscr{D}(\Omega)$) we have seen that the topology behaves well with respect to sequential convergence. Also continuous linear maps from $\mathscr{D}(\Omega)$ into other spaces are easily characterized: only sequential convergence needs to be considered.

THEOREM 3.6.3. *Let T be a linear map from $\mathscr{D}(\Omega)$ into some locally convex vector space Y. Then the following statements are equivalent.*

(a) *T is continuous.*
(b) *$T(\varphi_j) \to 0$ in Y whenever $\varphi_j \to 0$ in $\mathscr{D}(\Omega)$.*
(c) *$T|_{\mathscr{D}_K}$ is continuous for each K.*

PROOF. See [Ru] or [Sc]. □

We now introduce the usual operations on the space $\mathscr{D}(\Omega)$. The reflection and translation are only defined for certain sets (such as $\Omega = \mathbb{R}^n$), but complex conjugation and the derivative $\varphi \mapsto \partial^\alpha \varphi$ are well defined on $\mathscr{D}(\Omega)$. To define pointwise multiplication of functions in $\mathscr{D}(\Omega)$, it is clear that if $f \in C^\infty(\Omega)$ then $f\varphi$ will be in $\mathscr{D}(\Omega)$, and on the other hand multiplication by functions which are not infinitely differentiable need not give functions in $\mathscr{D}(\Omega)$. Hence $C^\infty(\Omega)$ is a natural space of multipliers on $\mathscr{D}(\Omega)$. We have the following theorem.

THEOREM 3.6.4. *Let $\Omega \subset \mathbb{R}^n$ be an open set. If $f \in C^\infty(\Omega)$, then the following operations are continuous maps from $\mathscr{D}(\Omega)$ into $\mathscr{D}(\Omega)$:*

$$\begin{aligned} &(1) \quad \varphi \mapsto \overline{\varphi} && \textit{(conjugation)} \\ &(2) \quad \varphi \mapsto \partial^\alpha \varphi && \textit{(derivative)} \\ &(3) \quad \varphi \mapsto f\varphi && \textit{(multiplication)} \end{aligned}$$

If $\Omega = \mathbb{R}^n$, then additionally the following operations are continuous from $\mathscr{D}(\mathbb{R}^n)$ into $\mathscr{D}(\mathbb{R}^n)$:

$$\begin{aligned} &(4) \quad \varphi \mapsto \tilde{\varphi} && \textit{(reflection)} \\ &(5) \quad \varphi \mapsto \tau_{x_0}\varphi && \textit{(translation)} \end{aligned}$$

PROOF. The proof is similar to the case of Schwartz functions (for continuity of multiplication, one needs to use the Leibniz rule for differentiation). □

3.7. The distribution space $\mathscr{D}'$

We are now ready to give a formal definition of distributions.

Definition. The set of continuous linear functionals on $\mathscr{D}(\Omega)$ is denoted by $\mathscr{D}'(\Omega)$ and its elements are called *distributions* on Ω.

It follows from Theorem 3.6.3 that a linear functional T on $\mathscr{D}(\Omega)$ is a distribution if for any sequence (φ_j) with $\varphi_j \to 0$ in $\mathscr{D}(\Omega)$ one has $T(\varphi_j) \to 0$, or equivalently if $T|_{\mathscr{D}_K}$ is continuous on $\mathscr{D}_K$ whenever $K \subset \Omega$ is compact. Combining the last fact with the same argument that was used to show that periodic or tempered distributions have finite order, we see that if $T \in \mathscr{D}'(\Omega)$ then for any compact set $K \subset \Omega$ there exist $C > 0$ and $N > 0$ such that

$$|T(\varphi)| \leq C \sum_{|\alpha| \leq N} \|\partial^\alpha \varphi\|_{L^\infty}, \qquad \varphi \in \mathscr{D}_K. \tag{3.18}$$

If there is a fixed N such that (3.18) is satisfied for any K then the distribution T is said to be of order $\leq N$, and if N is the least such integer then T is said to be of order N.

We will sometimes use the notation $\langle T, \varphi \rangle$ to indicate the action of a distribution on a test function.

Example 3.7.1. (Locally integrable functions) We denote by $L^1_{\text{loc}}(\Omega)$ the set of all measurable functions f on Ω such that $\int_K |f(x)|\, dx < \infty$ for any compact set $K \subset \Omega$. Any function $f \in L^1_{\text{loc}}(\Omega)$ gives rise to a distribution $T_f \in \mathscr{D}'(\Omega)$ defined by

$$T_f(\varphi) = \int_{\mathbb{R}^n} f(x)\varphi(x)\, dx. \tag{3.19}$$

Here T_f is continuous since for $\varphi \in \mathscr{D}_K$ we have

$$|T_f(\varphi)| \leq \int_K |f(x)\varphi(x)|\, dx \leq \|\varphi\|_\infty \int_K |f(x)|\, dx.$$

In particular any continuous function gives rise to a distribution. We will use the notation T_f for a distribution determined by the function f by (3.19). As before, different functions in L^1_{loc} determine different distributions, and we will identify the function f and the distribution T_f.

EXAMPLE 3.7.2. (Measures) Any positive or complex regular Borel measure μ in Ω gives rise to a distribution T_μ, where

$$T_\mu(\varphi) = \int_\Omega \varphi(x)\, d\mu(x).$$

This is continuous since for $\varphi \in \mathscr{D}_K$ one has $|T_\mu(\varphi)| \leq \|\varphi\|_{L^\infty} |\mu|(K)$ where $|\mu|$ is the total variation of μ. Conversely, if $T \in \mathscr{D}'$ has order 0, meaning that for any compact set K there is $C = C_K > 0$ such that

$$|T(\varphi)| \leq C\|\varphi\|_{L^\infty}, \qquad \varphi \in \mathscr{D}_K, \tag{3.20}$$

then T determines a unique measure (for the details see [**Sc**]). Hence distributions which satisfy (3.20) may be identified with measures.

EXAMPLE 3.7.3. (Tempered distributions) Any tempered distribution is in $\mathscr{D}'(\mathbb{R}^n)$. To see this, we need to show that if $T \in \mathscr{S}'(\mathbb{R}^n)$ and $\varphi_j \to 0$ in $\mathscr{D}(\mathbb{R}^n)$, then $T(\varphi_j) \to 0$. There is a compact set $K \subset \mathbb{R}^n$ such that $\mathrm{supp}(\varphi_j) \subset K$ for all j and $\partial^\alpha \varphi_j \to 0$ uniformly on K for any $\alpha \in \mathbb{N}^n$. Then also

$$\|\langle x\rangle^N \partial^\alpha \varphi_j\|_{L^\infty} \leq C_{K,N} \|\partial^\alpha \varphi_j\|_{L^\infty} \to 0$$

for any N and α, showing that $\varphi_j \to 0$ in $\mathscr{S}$. Thus $T(\varphi_j) \to 0$. This shows that we have the inclusions

$$\mathscr{E}'(\mathbb{R}^n) \subset \mathscr{S}'(\mathbb{R}^n) \subset \mathscr{D}'(\mathbb{R}^n).$$

The examples show that $\mathscr{D}'$ is a large space which contains many ordinary classes of functions and measures. The following step is to extend the operations from Theorem 3.6.4 to distributions. This proceeds exactly as before.

EXAMPLE 3.7.4. Consider the reflection operation on $\mathscr{D}$ which sends φ to $\tilde{\varphi}$. We wish to define the reflection of a distribution $T \in \mathscr{D}'$ as another distribution $\tilde{T}$. A reasonable requirement is that the operation should extend the reflection on $\mathscr{D}$, i.e. if $f \in \mathscr{D}$ then the reflection of T_f should be $T_{\tilde{f}}$. If this holds then we have

$$\tilde{T}_f(\varphi) = T_{\tilde{f}}(\varphi) = \int_{\mathbb{R}^n} f(-x)\varphi(x)\, dx = \int_{\mathbb{R}^n} f(x)\varphi(-x)\, dx = T_f(\tilde{\varphi}).$$

Motivated by this computation we *define* the reflection of $T \in \mathscr{D}'$ as the distribution $\tilde{T}$ given by

$$\tilde{T}(\varphi) = T(\tilde{\varphi}).$$

Here $\tilde{T}$ is continuous since the composition $\varphi \mapsto \tilde{\varphi} \mapsto T(\tilde{\varphi})$ is continuous from $\mathscr{D}$ to the scalars.

One may carry out similar computations as in the preceding example for the conjugation and translation to motivate the definitions $\overline{T}(\varphi) = \overline{T(\overline{\varphi})}$ and $(\tau_{x_0}T)(\varphi) = T(\tau_{-x_0}\varphi)$.

It is a remarkable fact that there is a natural notion of derivative on $\mathscr{D}'$. For $f \in \mathscr{D}$ the usual requirement that $\partial^\alpha T_f$ should be equal to $T_{\partial^\alpha f}$ leads to

$$\begin{aligned}(\partial^\alpha T_f)(\varphi) = T_{\partial^\alpha f}(\varphi) &= \int_{\mathbb{R}^n} (\partial^\alpha f)(x)\varphi(x)\,dx \\ &= (-1)^{|\alpha|}\int_{\mathbb{R}^n} f(x)(\partial^\alpha \varphi)(x)\,dx\end{aligned}$$

where we have integrated repeatedly by parts (the boundary terms vanish since the functions have compact support).

Definition. For any $T \in \mathscr{D}'$ we define the distribution $\partial^\alpha T$ by

$$(\partial^\alpha T)(\varphi) = (-1)^{|\alpha|} T(\partial^\alpha \varphi).$$

$(\partial^\alpha T)(\varphi)$ is called the *distribution derivative* or *weak derivative* of T.

Note that $\partial^\alpha T$ is continuous since differentiation is continuous on $\mathscr{D}$. It follows that any distribution has well defined derivatives of any order even if it arises from a function which is not differentiable in the classical sense. The definition of derivative also accommodates a form of integration by parts which is valid for distributions.

Example 3.7.5. As an example of weak derivatives consider the continuous function on $\mathbb{R}$ given by

$$f(x) = \begin{cases} 0, & x \le 0, \\ x, & x > 0. \end{cases}$$

Now f is not differentiable in the classical sense but determines a distribution

$$f : \varphi \mapsto \int_0^\infty x\varphi(x)\,dx,$$

and the distribution f has a derivative given by

$$f'(\varphi) = -f(\varphi') = -\int_0^\infty x\varphi'(x)\,dx = \int_0^\infty \varphi(x)\,dx$$

where we have used integration by parts. Hence f' can be identified with the *Heaviside unit step function*

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

Differentiation of H leads to the distribution H' with $H'(\varphi) = -H(\varphi') = -\int_0^\infty \varphi'(x)\,dx = \varphi(0)$. Hence we have arrived at the Dirac measure. The derivative of the Dirac measure is given by

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

EXAMPLE 3.7.6. Let $f \in C^1(\mathbb{R} \setminus \{x_0\})$ where f has a jump discontinuity at x_0, i.e. the limits $f(x_0-)$ and $f(x_0+)$ exist and are finite. We denote by $J(x_0) = f(x_0+) - f(x_0-)$ the jump of f at x_0. If $[f']$ is the classical derivative of f, defined everywhere except at x_0, and if Df is the distribution derivative then we have

$$\begin{aligned} \langle Df, \varphi\rangle = -\langle f, \varphi'\rangle &= -\int_{-\infty}^{x_0} f(x)\varphi'(x)\,dx - \int_{x_0}^{\infty} f(x)\varphi'(x)\,dx \\ &= (f(x_0+) - f(x_0-))\varphi(x_0) + \int_{-\infty}^{\infty} [f'(x)]\varphi(x)\,dx \end{aligned}$$

where integration by parts has been used. Thus we have the distributional relation

$$Df = [f'] + J(x_0)\delta_{x_0}.$$

EXAMPLE 3.7.7. The expression

$$T = \sum_{j=1}^{\infty} \delta_j^{(j)}$$

gives rise to a distribution in $\mathscr{D}'(\mathbb{R})$ that does not have finite order.

It is a striking fact that one can always differentiate pointwise convergent sequences of distributions.

THEOREM 3.7.1. *Let (T_j) be a sequence of distributions in $\mathscr{D}'$ so that $(T_j(\varphi))$ converges for all $\varphi \in \mathscr{D}$. Then there is a distribution $T \in \mathscr{D}'$ defined by $T(\varphi) = \lim_{j\to\infty} T_j(\varphi)$, and for any α we have*

$$\partial^\alpha T_j \to \partial^\alpha T \tag{3.21}$$

with convergence in $\mathscr{D}'$.

PROOF. See [Ru] or [Sc]. □

Besides differentiation one can also consider the inverse operation, which is the integration of distributions. We give two simple arguments in the one-dimensional case: the general case is treated in Schwartz [**Sc**], pp. 51-62. If $S \in \mathscr{D}'(\mathbb{R})$ then a distribution $T \in \mathscr{D}'(\mathbb{R})$ is called a *primitive* of S if $DT = S$.

THEOREM 3.7.2. *Any distribution $S \in \mathscr{D}'(\mathbb{R})$ has infinitely many primitives, any two of these differing by a constant.*

PROOF. We denote by H the space of those $\chi \in \mathscr{D}(\mathbb{R})$ which have integral zero over $\mathbb{R}$. It is easy to see that any $\chi \in H$ is of the form $\chi = \psi'$ for a unique $\psi \in \mathscr{D}(\mathbb{R})$. If φ_0 is a fixed function in $\mathscr{D}(\mathbb{R})$ which has integral one over $\mathbb{R}$, then any $\varphi \in \mathscr{D}(\mathbb{R})$ can be written uniquely in the form

$$\varphi = \lambda\varphi_0 + \chi \tag{3.22}$$

where $\lambda = \int_{-\infty}^{\infty} \varphi(t)\, dt$ and $\chi \in H$.

For $S \in \mathscr{D}'(\mathbb{R})$ define a linear form T on $\mathscr{D}(\mathbb{R})$ by

$$T(\varphi) = \lambda T(\varphi_0) - S(\chi) \tag{3.23}$$

where $\varphi \in \mathscr{D}(\mathbb{R})$ has been written in the form (3.22). If $\varphi_k \to 0$ in $\mathscr{D}(\mathbb{R})$ and $\varphi_k = \lambda_k\varphi_0 + \chi_k$, then $\lambda_k \to 0$ and also $\chi_k \to 0$ in $\mathscr{D}(\mathbb{R})$ since $\chi_k = \varphi_k - \lambda_k\varphi_0$. This shows that $T(\varphi_k) \to 0$ so T is a distribution, and by (3.23) we have

$$\langle DT, \varphi\rangle = -\langle T, \varphi'\rangle = \langle S, \varphi\rangle.$$

Thus T is a primitive of S. If T_1 and T_2 are two primitives of S then $\langle T_1 - T_2, \chi\rangle = 0$ for all $\chi \in H$ and

$$\langle T_1 - T_2, \varphi\rangle = \langle T_1 - T_2, \lambda\varphi_0 + \chi\rangle = C \int_{-\infty}^{\infty} \varphi(t)\, dt$$

where $C = \langle T_1 - T_2, \varphi_0\rangle$. Consequently $T_1 = T_2 + C$. ☐

THEOREM 3.7.3. *If $T \in \mathscr{D}'(\mathbb{R})$ is such that the distribution derivative D^kT is a continuous function $g(x)$, then T is a function in $C^k(\mathbb{R})$.*

PROOF. We may integrate k times to obtain a C^k function f so that $D^kf = g$ in the classical sense. Then $D^kT = D^kf$ which means that T and f differ by a polynomial of degree $\leq k - 1$. ☐

The final operation on distributions that we wish to introduce here is multiplication by functions. This is easy to define since if $f \in C^\infty$ then fT is a well-defined distribution if $(fT)(\varphi) = T(f\varphi)$, and the operation extends that on $\mathscr{D}$. We summarize what we have done.

THEOREM 3.7.4. *If $f \in C^\infty(\mathbb{R}^n)$ then the following operations are well defined maps from $\mathscr{D}'$ into $\mathscr{D}'$.*

(1) $\tilde{T}(\varphi) = T(\tilde{\varphi})$ *(reflection)*

(2) $\overline{T}(\varphi) = \overline{T(\overline{\varphi})}$ *(conjugation)*

(3) $(\tau_{x_0}T)(\varphi) = T(\tau_{-x_0}\varphi)$ *(translation)*

(4) $(D^\alpha T)(\varphi) = (-1)^{|\alpha|}T(D^\alpha\varphi)$ *(derivative)*

(5) $(fT)(\varphi) = T(f\varphi)$ *(multiplication)*

PROOF. Follows from the corresponding continuity properties on $\mathscr{D}$. □

To study the local behaviour of distributions we introduce the following concepts.

DEFINITION. For any open set $V \subset \Omega$ the distribution $T \in \mathscr{D}'(\Omega)$ is said to *vanish* on V, written $T = 0$ on V, if $T(\varphi) = 0$ for any $\varphi \in \mathscr{D}(V)$. Two distributions T_1 and T_2 are said to be *equal* on V if $T_1 - T_2 = 0$ on V.

It is an important fact that if the local behaviour of a distribution is known at each point then the distribution is uniquely determined globally. The proof uses a partition of unity.

THEOREM 3.7.5. *Let $\{V_i\}$ be an open cover of Ω and let $\{T_i\}$ be a family of distributions such that $T_i \in \mathscr{D}'(V_i)$, and suppose that for any V_i, V_j with $V_i \cap V_j \neq \emptyset$ one has*

$$T_i = T_j \text{ on } V_i \cap V_j.$$

Then there is a unique $T \in \mathscr{D}'(\Omega)$ for which $T = T_i$ on each V_i.

PROOF. Let $\{\psi_i\}$ be a C^∞ locally finite partition of unity subordinate to $\{V_i\}$. Define

$$T(\varphi) = \sum_i T_i(\psi_i\varphi) \qquad (\varphi \in \mathscr{D}(\Omega)). \tag{3.24}$$

If $K \subset \Omega$ is a compact set then the local finiteness of $\{\psi_i\}$ shows that K has some neighborhood where only finitely many of the ψ_i do not vanish. Consequently for $\varphi \in \mathscr{D}_K$ only finitely many of the functions $\psi_i\varphi$ will be nonzero, the sum in (3.24) will be finite, and T is a distribution.

If $\varphi \in \mathscr{D}_K$ for some compact $K \subset V_i$ then

$$T(\varphi) = \sum_j T_j(\psi_j\varphi) = T_i(\sum_j \psi_j\varphi) = T_i(\varphi)$$

since the sum is finite and for any j with $V_i \cap V_j \neq \emptyset$ one has $T_j(\psi_j\varphi) = T_i(\psi_j\varphi)$. This shows that $T = T_i$ on V_i, and also the uniqueness follows since any distribution T with $T = T_i$ on each V_i must be given by (3.24). □

Much of the justification for distribution theory comes from the fact that continuous functions possess infinitely many derivatives. On the other hand we have the following important theorem which states that any distribution is at least locally the derivative of a continuous function. This shows that $\mathscr{D}'$ is in a sense the smallest possible set where continuous functions can be differentiated at will.

THEOREM 3.7.6. *Let $T \in \mathscr{D}'(\Omega)$ and let $K \subset \Omega$ be a compact set. Then there is a continuous function f on Ω such that $T(\varphi) = (\partial^\alpha f)(\varphi)$ for all $\varphi \in \mathscr{D}_K$.*

3.8. Convolution of functions

We now define the important convolution operation first for certain classes of functions. This operation arises naturally in Fourier analysis since the Fourier transform takes convolutions into products.

DEFINITION. The *convolution* of two measurable functions $f, g : \mathbb{R}^n \to \mathbb{C}$ is the function $f * g : \mathbb{R}^n \to \mathbb{C}$ given by

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)\,dy \qquad (3.25)$$

provided that the integral exists almost everywhere.

Clearly the convolution of arbitrary functions need not be defined. In order for the integral (3.25) to converge the functions must satisfy certain growth restrictions at infinity, in particular the rapid growth of one function must be compensated by the rapid decrease at infinity of the other. Theorem 3.8.1 below illustrates this.

A change of variables in (3.25) gives that

$$(f * g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)\,dy, \tag{3.26}$$

which shows that the convolution is commutative ($f * g = g * f$). It is also associative by Fubini's theorem if the functions involved satisfy certain decay conditions. The identity (3.26) allows one to interpret $f * g$ as a weighted sum of translates of f. Since averaging translates of a function is a smoothing operation it should be no surprise that convolution turns irregular functions into smoother ones.

We introduce some new notation for the following theorem, which is stated in a fairly general form but which should clarify the relationship of regularity and growth properties in convolution.

Definition. We denote by $L_{\text{pol}}(\mathbb{R}^n)$ the space of measurable complex functions on $\mathbb{R}^n$ which are polynomially bounded. In other words, a measurable function f is in L_{pol} if and only if there are $C > 0$ and $N \in \mathbb{N}$ such that $|f(x)| \le C\langle x\rangle^N$ for almost all $x \in \mathbb{R}^n$. The set of continuous functions in L_{pol} is denoted by C_{pol}.

The space $C_\infty(\mathbb{R}^n)$ of *rapidly decreasing continuous functions* on $\mathbb{R}^n$ consists of those $f \in C(\mathbb{R}^n)$ for which $\langle x\rangle^N f(x)$ is a bounded function for all $N \in \mathbb{N}$. Differentiability is indicated by a superscript; a function f is in C^k_∞ (in C^k_{pol}) if $\partial^\alpha f$ is in C_∞ (in C_{pol}) whenever $|\alpha| \le k$.

Theorem 3.8.1. *The convolution is a map*

$$\begin{array}{lcll}
(1) & L^1_{\text{loc}} \times C^k_c & \to & C^k \\
(2) & C^j \times C^k_c & \to & C^{j+k} \\
(3) & C^j_c \times C^k_c & \to & C^{j+k}_c \\
(4) & L_{\text{pol}} \times C^k_\infty & \to & C^k_{\text{pol}} \\
(5) & C^j_{\text{pol}} \times C^k_\infty & \to & C^{j+k}_{\text{pol}} \\
(6) & C^j_\infty \times C^k_\infty & \to & C^{j+k}_\infty
\end{array}$$

One has the identity

$$\partial^{\alpha+\beta}(f * g) = (\partial^\alpha f) * (\partial^\beta g)$$

whenever $|\alpha| \le j, |\beta| \le k$ (where $j = 0$ in (1) and (4)).

LEMMA 3.8.2. *If $\langle x\rangle^N f \in L^\infty(\mathbb{R}^n)$ and $K \subset \mathbb{R}^n$ is compact, then there is $C = C_{K,N} > 0$ such that*

$$\sup_{h\in K}\sup_{x\in\mathbb{R}^n} |\langle x\rangle^N f(x+h)| \le C\|\langle x\rangle^N f\|_{L^\infty}. \tag{3.27}$$

PROOF. We first take $N = 2m$ to be an even integer, and we may also assume that $K = \overline{B}(0,R)$. Now

$$\sup_{h\in K}\sup_{x\in\mathbb{R}^n} |\langle x\rangle^{2m} f(x+h)| = \sup_{h\in K}\sup_{x\in\mathbb{R}^n} |\langle x-h\rangle^{2m} f(x)|.$$

The expression $\langle x-h\rangle^{2m} = (1+|x-h|^2)^m = (1+|x|^2 - 2x\cdot h + |h|^2)^m$ may be expanded into

$$\langle x-h\rangle^{2m} = \sum_{j=0}^{m} \binom{m}{j}(1+|x|^2)^{m-j}(-2x\cdot h + |h|^2)^j.$$

The condition $|h| \le R$ implies that $|-2x\cdot h + |h|^2| \le C_R\langle x\rangle$. We thus have the estimate

$$\langle x-h\rangle^{2m} \le C\langle x\rangle^{2m}, \qquad h \in K.$$

If $N = 2m+1$ is an odd integer, we write $\langle x-h\rangle^{2m+1} = (\langle x-h\rangle^{2m})^{\frac{2m+1}{2m}}$. The estimate above implies that for any $N \in \mathbb{N}$,

$$\langle x-h\rangle^{N} \le C\langle x\rangle^{N}, \qquad h \in K.$$

This proves the result. □

PROOF OF THEOREM 3.8.1. (1) Let $f \in L^1_{\text{loc}}$ and $g \in C^k_c$. If $x \in \mathbb{R}^n$ is fixed then the integral in (3.25) reduces to one over the compact set $\operatorname{supp}(\tau_x\tilde{g})$, so $(f*g)(x)$ exists. For differentiability consider

$$\frac{(f*g)(x+he_j) - (f*g)(x)}{h} = \int_{\mathbb{R}^n} f(y)\frac{g(x-y+he_j) - g(x-y)}{h}dy. \tag{3.28}$$

If $|h| \le 1$ the integral reduces to one over some compact set K, and Taylor's theorem gives

$$\frac{g(x-y+he_j) - g(x-y)}{h} = \frac{\partial g}{\partial x_j}(x-y+\theta e_j)$$

where $|\theta| \le 1$. The integrand in (3.28) is now the product of $f(y)$ and a bounded function on K, hence is bounded by a function in $L^1(K)$, and we may apply dominated convergence to obtain

$$\frac{\partial(f*g)}{\partial x_j}(x) = \left(f * \frac{\partial g}{\partial x_j}\right)(x). \tag{3.29}$$

Iterating this argument gives that $f*g$ is in C^k and $\partial^\beta(f*g) = f*(\partial^\beta g)$ for $|\beta| \leq k$.

(2) The same argument as in (1) shows that $\partial^\alpha(f*g) = (\partial^\alpha f)*g$ for $|\alpha| \leq j$.

(3) Differentiability follows from (2), and the support condition follows from the inclusion $\mathrm{supp}(f*g) \subset \mathrm{supp}(f) + \mathrm{supp}(g)$. This last fact is shown by noting that if $x \notin \mathrm{supp}(f)+\mathrm{supp}(g)$, then $y \in \mathrm{supp}(f)$ implies that $x-y \notin \mathrm{supp}(g)$, and then $(f*g)(x)$ must be zero by the definition (3.25). Since $\mathrm{supp}(f)+\mathrm{supp}(g)$ is closed the given inclusion must hold.

(4) Let $f \in L_{\mathrm{pol}}$ and $g \in C^k_\infty$. Then $\langle y\rangle^{-N} f(y) \in L^1(\mathbb{R}^n)$ for some large enough N, and for any fixed x

$$\begin{aligned} |(f*g)(x)| &\leq \int_{\mathbb{R}^n} |f(y)g(x-y)|\,dy \\ &\leq \|\langle y\rangle^{-N} f(y)\|_{L^1} \|\langle y\rangle^N \tau_x \tilde{g}(y)\|_{L^\infty}. \end{aligned}$$

This shows that $f*g$ exists. If $|h| \leq 1$ then the integrand in (3.28) satisfies

$$\begin{aligned} \left| f(y)\frac{g(x-y+he_j)-g(x-y)}{h} \right| &\leq |\langle y\rangle^{-N} f(y)| \\ &\quad\times \left| \langle y\rangle^N \frac{\partial g}{\partial x_j}(x-y+\theta e_j) \right| \qquad (|\theta| \leq 1). \end{aligned}$$

If N is large enough then the first factor is in $L^1(\mathbb{R}^n)$ and the second is bounded by Lemma 3.8.2, hence dominated convergence gives (3.29) in this case. It follows that $f*g \in C^k$.

The identity (3.29) also shows that $f*g \in C^k_{\mathrm{pol}}$ if we can prove that $f*g \in C_{\mathrm{pol}}$. Choosing N as above we have

$$\begin{aligned} |\langle x\rangle^{-N}(f*g)(x)| &= \left| \langle x\rangle^{-N} \int_{\mathbb{R}^n} f(x-y)g(y)dy \right| \\ &\leq \int_{\mathbb{R}^n} \left| \langle x-y\rangle^{-N} f(x-y) \frac{\langle x-y\rangle^N}{\langle x\rangle^N} g(y) \right| dy \\ &\leq C\, \|\langle y\rangle^{-N} f(y)\|_{L^1} \cdot \sup_{y\in\mathbb{R}^n} \frac{\langle x-y\rangle^N}{\langle x\rangle^N \langle y\rangle^N}. \end{aligned}$$

The last expression is finite since $1+|x-y|^2 \leq 1+2(|x|^2+|y|^2) \leq 2(1+|x|^2)(1+|y|^2)$. Hence $f*g$ is polynomially bounded.

(5) This follows similarly as in (4).

(6) By (5) it is enough to show that $f*g \in C_\infty$ whenever $f,g \in C_\infty$. The binomial expansion gives $(x-y+y)^\alpha = \sum_{i=1}^k c_i(x-y)^{\alpha_i}y^{\beta_i}$ for some constants c_i and some multi-indices α_i and β_i, so we have

$$\begin{aligned}|x^\alpha(f*g)(x)| &\leq \int_{\mathbb{R}^n}\Big|(x-y+y)^\alpha f(y)g(x-y)\Big|\,dy\\ &\leq \sum_{i=1}^k |c_i| \int_{\mathbb{R}^n}\Big|y^{\beta_i}f(y)(x-y)^{\alpha_i}g(x-y)\Big|\,dy\\ &\leq \sum_{i=1}^k |c_i|\ \|z^{\beta_i}f(z)\|_{L^1(\mathbb{R}^n)}\ \|z^{\alpha_i}g(z)\|_\infty. \qquad (3.30)\end{aligned}$$

This implies that $\langle x\rangle^N(f*g)(x)$ is a bounded function for any $N\in\mathbb{N}$, so the claim follows. □

THEOREM 3.8.3. *The convolution is a separately continuous map*

$$\begin{array}{llll} (1) & \mathscr{D}\times\mathscr{D} & \to & \mathscr{D}, \\ (2) & \mathscr{E}\times\mathscr{D} & \to & \mathscr{E}, \\ (3) & \mathscr{S}\times\mathscr{S} & \to & \mathscr{S}. \end{array}$$

PROOF. Theorem 3.8.1 immediately gives that the ranges in (1) – (3) are correct. It remains to show continuity. If $f\in\mathscr{E}$ and $\varphi\in\mathscr{D}_K$, then for any compact subset K_0 of $\mathbb{R}^n$,

$$\begin{aligned}\sup_{x\in K_0}|\partial^\alpha(f*\varphi)(x)| = \sup_{x\in K_0}|(f*\partial^\alpha\varphi)(x)| &\leq \sup_{x\in K_0}\int_K |f(x-y)(\partial^\alpha\varphi)(y)|\,dy\\ &\leq \sup_{y\in K_1}|f(y)|\cdot\sup_{y\in K}|(\partial^\alpha\varphi)(y)|\cdot\mu(K)\end{aligned}$$

where $K_1 = K_0 - K$ is compact. Taking $\varphi=\varphi_k$ where $\varphi_k\to 0$ in $\mathscr{D}_K$ gives (1) and one half of (2). Also the other half of (2) follows if one takes the derivative of f instead of φ in the above.

Part (3) is a consequence of (3.30) which states that for $f,g\in\mathscr{S}$ one has $\|x^\alpha(f*g)\|_{L^\infty}\leq\rho(g)$ where ρ is a continuous seminorm on $\mathscr{S}$; this implies that

$$\|f*g\|_{\alpha,\beta} = \|x^\alpha f*(\partial^\beta g)\|_{L^\infty}\leq\rho(\partial^\beta g),$$

the right side being another continuous seminorm on $\mathscr{S}$. □

The convolution is a tool which can be used to prove approximation theorems. The idea, which is classical, is that convolving a function with a regular function looking like the Dirac delta gives a regular function close to the original one. In fact we will later define convolution of a function and a distribution, and then $f * \delta$ will be exactly equal to f.

DEFINITION. Suppose $j \in \mathscr{D}(\mathbb{R}^n)$ is such that $j \geq 0$, the support of j is contained in the closed unit ball of $\mathbb{R}^n$, and $\int_{\mathbb{R}^n} j(x)\,dx = 1$. Then the family of functions $\{j_\varepsilon\}$, where $j_\varepsilon(x) = \varepsilon^{-n} j(x/\varepsilon)$ and $\varepsilon > 0$, is called an *approximate identity*.

It is clear that approximate identities exist on $\mathbb{R}^n$. The function j_ε has support contained in $\overline{B}(0,\varepsilon)$ and its integral over $\mathbb{R}^n$ is equal to one, so the functions j_ε converge (in a sense which is made precise later) to the Dirac delta. For a locally integrable function f, the convolutions $f * j_\varepsilon$ are called *regularizations* of f.

THEOREM 3.8.4. *Let $\{j_\varepsilon\}$ be an approximate identity on $\mathbb{R}^n$.*

(a) *If f is a continuous function on $\mathbb{R}^n$ then $f * j_\varepsilon \to f$ uniformly on compact subsets of $\mathbb{R}^n$.*
(b) *If f is in $L^p(\mathbb{R}^n)$ then $f * j_\varepsilon \to f$ in $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$.*
(c) *If f is in $\mathscr{D}$ (in $\mathscr{E}$, $\mathscr{S}$) then $f * j_\varepsilon \to f$ in $\mathscr{D}$ (in $\mathscr{E}$, $\mathscr{S}$).*

PROOF. (a) Let $K \subset \mathbb{R}^n$ be compact and let $\varepsilon' > 0$. One has

$$\begin{aligned}|(f * j_\varepsilon)(x) - f(x)| &= \Big| \int_{\mathbb{R}^n} f(x-y) j_\varepsilon(y)\,dy - f(x) \int_{\mathbb{R}^n} j_\varepsilon(y)\,dy \Big| \\ &\leq \int_{\mathbb{R}^n} j_\varepsilon(y) \Big| f(x-y) - f(x) \Big|\,dy.\end{aligned}$$

The last integral can be taken over $\overline{B}(0,\varepsilon)$, so choosing ε so small that $|f(x-y) - f(x)| < \varepsilon'$ on $K + \overline{B}(0,\varepsilon)$ for $|y| \leq \varepsilon$ gives the claim.

(b) This follows from Minkowski's inequality in integral form and the continuity of translation on L^p similarly as in the periodic case (see Lemma 2.1.5).

(c) The claim follows for $\mathscr{D}$ and $\mathscr{E}$ directly from (a) and the fact that derivatives commute with convolution. For $\mathscr{S}$ we have

$$\begin{aligned}|x^\alpha[(f * j_\varepsilon)(x) - f(x)]| &= \Big| x^\alpha \Big\{ \int_{\mathbb{R}^n} f(x-y) j_\varepsilon(y)\,dy - f(x) \int_{\mathbb{R}^n} j_\varepsilon(y)\,dy \Big\} \Big| \\ &\leq \int_{\overline{B}(0,\varepsilon)} j_\varepsilon(y) \Big| x^\alpha \Big\{ f(x-y) - f(x) \Big\} \Big|\,dy.\end{aligned}$$

Using Taylor's theorem we may write

$$x^\alpha(f(x-y)-f(x)) = -\sum_{i=1}^n x^\alpha \frac{\partial f}{\partial x_i}(x+\theta y)y_i \quad (|\theta|<\varepsilon). \tag{3.31}$$

Lemma 3.8.2 now shows that the expressions $x^\alpha(\partial f/\partial x_i)(x+h)$ are bounded for $x \in \mathbb{R}^n$ and $|h| \le 1$, so the absolute value of (3.31) goes to zero as $\varepsilon \to 0$. We have shown that $\|f * j_\varepsilon - f\|_{\alpha,0} \to 0$ as $\varepsilon \to 0$, and the convergence with respect to $\|\cdot\|_{\alpha,\beta}$ follows just because we may replace f in the above by $\partial^\beta f$. □

LEMMA 3.8.5. *$\mathscr{D}(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \le p < \infty$ and uniformly dense in $C_0(\mathbb{R}^n)$.*

PROOF. Since C_c is dense in L^p the first claim follows from (b) in the theorem. The second claim is given by (a) which says that $\mathscr{D}$ is uniformly dense in C_c and hence in C_0. □

3.9. Convolution of distributions

We first define convolution between distributions and functions. The usual requirement that the operation should extend the convolution of functions leads to the following: if g is a function and T_g the corresponding distribution, and if f is a function, then

$$\begin{aligned}(T_g * f)(\varphi) &= \int_{\mathbb{R}^n} (g*f)(x)\varphi(x)\,dx = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} g(y)f(x-y)\varphi(x)\,dy\,dx \\ &= \int_{\mathbb{R}^n} g(y)\Big(\int_{\mathbb{R}^n} \tilde f(y-x)\varphi(x)\,dx\Big)\,dy \\ &= T_g(\tilde f * \varphi).\end{aligned}$$

The formula

$$(T*f)(\varphi) = T(\tilde f * \varphi)$$

can be used in conjunction with Theorem 3.8.3 to define $T * f$ on $\mathscr{D}' \times \mathscr{D}$, $\mathscr{E}' \times \mathscr{E}$ and $\mathscr{S}' \times \mathscr{S}$ (for instance if $T \in \mathscr{D}'$ and $f \in \mathscr{D}$ then the composition $\varphi \mapsto \tilde f * \varphi \mapsto T(\tilde f * \varphi)$ is continuous $\mathscr{D} \to \mathbb{C}$). This definition gives that $T * f$ will be either in $\mathscr{D}'$ or $\mathscr{S}'$; it is due to the smoothing nature of convolution that $T * f$ will in fact be a function.

THEOREM 3.9.1. *The convolution is a separately continuous map*

$$\begin{array}{llcl} (1) & \mathscr{D}' \times \mathscr{D} & \to & \mathscr{E}, \\ (2) & \mathscr{E}' \times \mathscr{D} & \to & \mathscr{D}, \\ (3) & \mathscr{E}' \times \mathscr{E} & \to & \mathscr{E}, \\ (4) & \mathscr{S}' \times \mathscr{S} & \to & \mathscr{O}_M. \end{array}$$

In each case the function $T * f$ *is given by*

$$(T * f)(x) = T(\tau_x \tilde{f}).$$

Furthermore, the following identities are valid.

$$\begin{array}{ll} (a) & \partial^\beta (T * f) = \partial^\beta T * f = T * \partial^\beta f \\ (b) & \quad (T * f) * g = T * (f * g) \end{array}$$

PROOF. This theorem follows from the structure theorems since the distributions can be written as derivatives of continuous functions. We provide the details for (4).

Let $T \in \mathscr{S}'$ and $f \in \mathscr{S}$. By Theorem 3.3.5 there is a polynomially bounded continuous function h such that $T = \partial^\alpha T_h$, where

$$T_h(\varphi) = \int_{\mathbb{R}^n} h(x)\varphi(x)\,dx, \quad T(\varphi) = (-1)^{|\alpha|} \int_{\mathbb{R}^n} h(x)\partial^\alpha \varphi(x)\,dx.$$

We now have

$$\begin{aligned} (T * f)(\varphi) &= T(\tilde{f} * \varphi) = (\partial^\alpha T_h)(\tilde{f} * \varphi) = (-1)^\alpha T_h((\partial^\alpha \tilde{f}) * \varphi) \\ &= T_h((\partial^\alpha f)\tilde{\ } * \varphi) = \int_{\mathbb{R}^n} h(x) \Big\{ \int_{\mathbb{R}^n} \partial^\alpha f(y-x)\varphi(y)\,dy \Big\}\,dx \\ &= \int_{\mathbb{R}^n} (h * \partial^\alpha f)(y)\varphi(y)\,dy. \end{aligned}$$

This shows that $T * f$ is the distribution arising from the function $h * \partial^\alpha f$, which is in $\mathscr{O}_M$ by Theorem 3.8.1. This function has the form

$$\begin{aligned} (h * \partial^\alpha f)(x) &= \int_{\mathbb{R}^n} h(y)(\partial^\alpha f)(x-y)\,dy \\ &= (-1)^{|\alpha|} \int_{\mathbb{R}^n} h(y)\partial^\alpha (\tau_x \tilde{f})(y)\,dy = T(\tau_x \tilde{f}). \end{aligned}$$

The continuity proof, which would require us to define a topology on $\mathscr{O}_M$, is omitted (see [Sc]). The proofs of (a) and (b) are just manipulations of the definitions. □

Where discussing convolution on $\mathscr{S}'$ there is another space of distributions which is useful, namely the space $\mathscr{O}'_C$ of *rapidly decreasing distributions*, which we set out to define.

Definition. The space $\mathscr{D}_{L^1}(\mathbb{R}^n)$ consists of those functions $f \in C^\infty(\mathbb{R}^n)$ such that f is in L^1 along with all its derivatives. We give $\mathscr{D}_{L^1}$ a topology by the countable family of norms

$$\|f\|_\alpha = \|\partial^\alpha f\|_{L^1}, \qquad \alpha \in \mathbb{N}^n.$$

The space of continuous linear functionals on $\mathscr{D}_{L^1}(\mathbb{R}^n)$ is denoted by $\mathscr{B}'(\mathbb{R}^n)$ and its members are said to be distributions *bounded* on $\mathbb{R}^n$.

The space $\mathscr{O}'_C(\mathbb{R}^n)$ is now taken to be the set of those $T \in \mathscr{D}'$ for which $\langle x\rangle^N T$ is a bounded distribution (i.e. belongs to $\mathscr{B}'$) for any $N \in \mathbb{N}$.

The test function space $\mathscr{D}_{L^1}$ is a complete metric space, and we have $\mathscr{D} \subset \mathscr{D}_{L^1} \subset \mathscr{E}$ with continuous embeddings. Also $\mathscr{D}$ is dense in $\mathscr{D}_{L^1}$ since for $f \in \mathscr{D}_{L^1}$ there is a sequence $\varphi_k(x) = \psi(x/k)f(x)$ in $\mathscr{D}$ where $\psi \in \mathscr{D}$ is equal to one on the closed unit ball of $\mathbb{R}^n$, and it is easy to check that

$$\|\partial^\alpha(\varphi_k - f)\|_{L^1} = \int_{|x|\geq k} |\partial^\alpha(\varphi_k(x) - f(x))|\, dx \to 0$$

Thus $\mathscr{B}'$ can be identified with a subspace of $\mathscr{D}'$ and it contains all compactly supported distributions. The structure theorem for the spaces $\mathscr{B}'$ and $\mathscr{O}'_C$ has the following form.

Theorem 3.9.2. *Let T be a distribution in $\mathscr{D}'$.*

(1) *T is in $\mathscr{B}'$ if and only if $T = \sum_{|\alpha|\leq N} \partial^\alpha g_\alpha$ where the g_α are in L^∞.*
(2) *T is in $\mathscr{O}'_C$ if and only if for any $N \in \mathbb{N}$ there exist $M(N) \in \mathbb{N}$ and continuous functions g_α such that $T = \sum_{|\alpha|\leq M(N)} \partial^\alpha g_\alpha$, where $\langle x\rangle^N g_\alpha$ is a bounded function for each α.*

Proof. Modifications of the proof of Theorem 3.3.5 give the first claim. The second claim follows from the first upon integrating by parts. ☐

Note that the preceding theorem implies that $\mathscr{E}' \subset \mathscr{O}'_C \subset \mathscr{S}'$, and functions in $\mathscr{S}$ and also C_∞, for instance $x \mapsto e^{-|x|}$ on $\mathbb{R}$, lie in $\mathscr{O}'_C$.

THEOREM 3.9.3. *The convolution is a map $\mathscr{O}'_C \times \mathscr{S} \to \mathscr{S}$ continuous in the second argument.*

PROOF. We know from Theorem 3.9.1 that $T * f$ is a function in $\mathscr{O}_M$ and $(T*f)(x) = T(\tau_x \tilde{f})$ if $T \in \mathscr{O}'_C$ and $f \in \mathscr{S}$. To show that $T*f$ is in $\mathscr{S}$ take any $\beta \in \mathbb{N}^n$ and choose m such that $|x^\beta| \le \langle x \rangle^{2m}$ for all $x \in \mathbb{R}^n$. Let g_α be the functions in Theorem 3.9.2, part (b); then

$$\begin{aligned} x^\beta \partial^\gamma (T*f)(x) = x^\beta T(\tau_x (\partial^\gamma f)\tilde{}) &= \sum_{|\alpha| \le N} x^\beta (\partial^\alpha g_\alpha)(\tau_x (\partial^\gamma f)\tilde{}) \\ &= \sum_{|\alpha| \le N} \int_{\mathbb{R}^n} x^\beta g_\alpha(y) (\partial^{\alpha+\gamma} f)(x-y)\, dy. \end{aligned}$$

The method used in the proof of Theorem 3.8.1, part (6) now gives that $T*f \in \mathscr{S}$ and that $f \mapsto T*f$ is continuous $\mathscr{S} \to \mathscr{S}$. □

As for functions, also distributions T can be approximated with the regularizations $T * j_\varepsilon$. It is remarkable here that the regularizations are functions by Theorem 3.9.1, so any distribution is in fact the limit of a sequence of C^∞ functions.

THEOREM 3.9.4. *Let $\{j_\varepsilon\}$ be an approximate identity on $\mathbb{R}^n$.*

(a) *If $T \in \mathscr{D}'$ then $T * j_\varepsilon \to T$ in $\mathscr{D}'$.*
(b) *If $T \in \mathscr{E}'$ then $T * j_\varepsilon \to T$ in $\mathscr{E}'$.*
(c) *If $T \in \mathscr{S}'$ then $T * j_\varepsilon \to T$ in $\mathscr{S}'$.*

PROOF. The proofs of (a) – (c) are identical. For (c) let $T \in \mathscr{S}'$ and choose any $\varphi \in \mathscr{S}$. Now $\{\tilde{j}_\varepsilon\}$ is an approximate identity, so $\tilde{j}_\varepsilon * \varphi \to \varphi$ in $\mathscr{S}$ by Theorem 3.8.4, part (b). Then $T(\tilde{j}_\varepsilon * \varphi) \to T(\varphi)$ by the continuity of T, which means that $T * j_\varepsilon \to T$ in the topology of $\mathscr{S}'$ by the definition of convolution. □

We have discussed convolution for functions and for a function and a distribution. It is natural to define the convolution of two distributions S and T by

$$(S*T)(\varphi) = S(\tilde{T} * \varphi)$$

provided that the expression on the right makes sense. This is the case for instance when $S \in \mathscr{D}'$ and T has compact support; in general the growth of S must be compensated by decay of T for $S*T$ to be defined, exactly as for functions. The analogy between the following theorem and Theorem 3.8.1 is evident.

THEOREM 3.9.5. *The convolution is a separately continuous map*

$$\begin{array}{llll} (1) & \mathscr{D}' \times \mathscr{E}' & \to & \mathscr{D}', \\ (2) & \mathscr{E}' \times \mathscr{E}' & \to & \mathscr{E}', \\ (3) & \mathscr{S}' \times \mathscr{O}'_C & \to & \mathscr{S}'. \end{array}$$

PROOF. Let S be in $\mathscr{D}'$ (in $\mathscr{E}'$, $\mathscr{S}'$) and T in $\mathscr{E}'$ (in $\mathscr{E}'$, $\mathscr{O}'_C$). If φ is any test function in $\mathscr{D}$ (in $\mathscr{E}$, $\mathscr{S}$), then

$$\varphi \mapsto \tilde{T} * \varphi \mapsto S(\tilde{T} * \varphi)$$

maps $\mathscr{D}$ ($\mathscr{E}, \mathscr{S}$) continuously and linearly into $\mathbb{C}$ by Theorem 3.9.1. This shows that convolution is indeed well defined in the settings of (1) – (3). □

Having defined the convolution on fairly general spaces, we now summarize some of the properties of the operation. In the following the distributions are assumed to be chosen so that all the convolutions are defined (for instance in the first part $T_1 \in \mathscr{D}'$ and $T_2 \in \mathscr{E}'$, or $T_1 \in \mathscr{S}'$ and $T_2 \in \mathscr{O}'_C$ etc.).

(1) *Commutativity.* For two distributions T_1 and T_2 one has

$$T_1 * T_2 = T_2 * T_1.$$

For functions this is valid by a change of variables, and for distributions commutativity is essentially a matter of definition.

(2) *Associativity.* If T_1, T_2, T_3 are distributions in $\mathscr{D}'$ and at least two have compact support, then

$$T_1 * (T_2 * T_3) = (T_1 * T_2) * T_3.$$

This is not valid without the condition on supports even if all the convolutions were defined: take for instance $T_1 = 1$, $T_2 = \delta'$, and $T_3 = H$ (the Heaviside unit step function). If two of the distributions have compact support then the statement follows by manipulating the definitions.

(3) *Translation invariance.* If $x \in \mathbb{R}^n$ then

$$\tau_x(T_1 * T_2) = (\tau_x T_1) * T_2 = T_1 * (\tau_x T_2).$$

This clearly holds for functions, and the extension to distributions follows from the definition.

(4) *Differentiation.* If α is a multi-index then

$$\partial^\alpha(T_1 * T_2) = (\partial^\alpha T_1) * T_2 = T_1 * (\partial^\alpha T_2).$$

We saw in Theorem 3.8.1 that if a function is convolved with a second one which is differentiable in the classical sense, then the convolution is differentiable and the derivatives are obtained by differentiating the second function. Distribution theory generalizes the classical setting and the above identity is always valid if all the derivatives are taken in the distributional sense.

(5) *Identity.* The Dirac measure δ is an identity element for the convolution operation: if $T \in \mathscr{D}'$ then

$$T * \delta = \delta * T = T.$$

To show this take $\varphi \in \mathscr{D}$ and note that $(\delta * \varphi)(x) = \delta(\tau_x \tilde{\varphi}) = \varphi(x)$, which gives the general case since $(T * \delta)(\varphi) = T(\delta * \varphi)$.

(6) *Translation.* If $T \in \mathscr{D}'$ and $x \in \mathbb{R}^n$ then

$$T * \delta_x = \delta_x * T = \tau_x T.$$

This is a consequence of part 5 and translation invariance.

(7) *Differentiation.* If $T \in \mathscr{D}'$ and α is a multi-index then

$$T * (\partial^\alpha \delta) = (\partial^\alpha \delta) * T = \partial^\alpha T.$$

Use part 5 and part 4.

A map $L : \mathscr{D} \to \mathscr{D}'$ is said to be translation-invariant if

$$\tau_x \circ L = L \circ \tau_x$$

for all $x \in \mathbb{R}^n$. The above discussion shows that convolution with a given $T \in \mathscr{D}'$ is a continuous translation-invariant linear map $\mathscr{D} \to \mathscr{E}$; we prove below that these properties also characterize convolution maps. In the following we denote by $\mathbb{C}^{\mathbb{R}^n}$ the space of all maps $\mathbb{R}^n \to \mathbb{C}$ with the product topology (i.e. the weakest topology which makes all projections $f \mapsto f(x)$ continuous).

THEOREM 3.9.6. *If L is any continuous translation-invariant linear map $\mathscr{D} \to \mathbb{C}^{\mathbb{R}^n}$, then there is a unique distribution $T \in \mathscr{D}'$ so that $L(\varphi) = T * \varphi$ for all $\varphi \in \mathscr{D}$. Particularly, the range of L is in $\mathscr{E}$.*

PROOF. We define T as a linear functional on $\mathscr{D}$ by $T(\varphi) = L(\tilde{\varphi})(0)$. The continuity assumption gives that T is continuous, hence it is a distribution. If $\varphi \in \mathscr{D}$ then by translation invariance

$$L(\varphi)(x) = (\tau_{-x}L(\varphi))(0) = L(\tau_{-x}\varphi)(0) = T(\tau_x\tilde{\varphi}) = (T * \varphi)(x).$$

This gives existence, and uniqueness follows since if $T * \varphi = 0$ for all $\varphi \in \mathscr{D}$ then also $T * j_\varepsilon = 0$ for any ε, and we have $T = 0$ by taking the limit. □

There is a similar theorem of Schwartz ([Sc], p. 53) which states that any continuous linear map $\mathscr{D} \to \mathscr{D}'$ which commutes with translations is a convolution map. All in all one can conclude that linearity and translation invariance combined with fairly weak continuity requirements force a map on $\mathscr{D}$ to come from convolution.

There is a much stronger theorem that is valid for almost any linear operator (not necessarily translation invariant). To motivate this, note that if Ω_1 and Ω_2 are open sets and $K \in C(\Omega_1 \times \Omega_2)$, then there is a corresponding integral operator

$$L : C_c(\Omega_2) \to C(\Omega_1), \quad Lf(x) = \int_{\Omega_2} K(x,y)f(y)\,dy.$$

The function $K(x,y)$ is called the integral kernel of L. Typically one might not think that arbitrary linear operators can be written as integral operators with respect to some kernel. However, the Schwartz kernel theorem says that this is in fact true if one allows the integral kernel to be a distribution (and if the linear operator satisfies a mild continuity assumption).

THEOREM 3.9.7. *(Schwartz kernel theorem) Assume that $\Omega_1 \subset \mathbb{R}^{n_1}$, $\Omega_2 \subset \mathbb{R}^{n_2}$ are open. If L is a continuous linear operator $\mathscr{D}(\Omega_2) \to \mathscr{D}'(\Omega_1)$, there is a unique $K \in \mathscr{D}'(\Omega_1 \times \Omega_2)$ such that*

$$\langle L(\varphi), \psi\rangle = \langle K, \psi \otimes \varphi\rangle, \qquad \psi \in \mathscr{D}(\Omega_1), \quad \varphi \in \mathscr{D}(\Omega_2).$$

Here the tensor product is defined by

$$(\psi \otimes \varphi)(x,y) = \psi(x)\varphi(y), \qquad x \in \Omega_1, \quad y \in \Omega_2.$$

Conversely, any $K \in \mathscr{D}'(\Omega_1 \times \Omega_2)$ gives rise to a unique continuous linear operator $L : \mathscr{D}(\Omega_1) \to \mathscr{D}'(\Omega_2)$ that satisfies the above formula.

We conclude the section with a general convolution-multiplication theorem for the Fourier transform of tempered distributions. For the proof of the first theorem note that the Laplace operator

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_n^2}$$

satisfies $((1-\Delta)T)\hat{} = \langle x \rangle^2 \hat{T}$ and by iteration $((1-\Delta)^k T)\hat{} = \langle x \rangle^{2k} \hat{T}$.

Theorem 3.9.8. *The Fourier transform is a bijective map from* $\mathscr{O}_M(\mathbb{R}^n)$ *onto* $\mathscr{O}'_C(\mathbb{R}^n)$.

Proof. It is enough to show that $\mathscr{F}$ takes $\mathscr{O}_M$ into $\mathscr{O}'_C$ and vice versa. Let first $f \in \mathscr{O}_M$. For any $m \geq 0$ there is by definition some $k > 0$ such that $|(1-\Delta)^m f(x)| \leq C\langle x \rangle^{2k}$. We can increase k so that the function

$$h(x) = \langle x \rangle^{-2k}(1-\Delta)^m f(x)$$

will be in $L^1(\mathbb{R}^n)$. Taking Fourier transforms gives

$$(1-\Delta)^k \hat{h} = \langle x \rangle^{2m} \hat{f}.$$

Now by the Riemann-Lebesgue lemma (Theorem 3.4.3) the function $\hat{h}$ is continuous and bounded, hence $\langle x \rangle^{2m} \hat{f}$ is a bounded distribution by Theorem 3.9.2. This shows that $\hat{f} \in \mathscr{O}'_C$.

On the other hand if $T \in \mathscr{O}'_C$, then for any β also $(-x)^\beta T$ is in $\mathscr{O}'_C$ and Theorem 3.9.2 implies that we may write $(-x)^\beta T = \sum_{|\alpha| \leq N} D^\alpha g_\alpha$ where the g_α are functions in L^1. The Fourier transform immediately gives

$$D^\beta \hat{T} = \sum_{|\alpha| \leq N} x^\alpha \hat{g}_\alpha.$$

An application of the Riemann-Lebesgue lemma shows that $D^\beta \hat{T}$ is a continuous polynomially bounded function for each β, thus $\hat{T} \in \mathscr{O}_M$. □

The next result is a very general convolution-multiplication theorem for the Fourier transform.

Theorem 3.9.9. *If* $S \in \mathscr{S}'$ *and* $T \in \mathscr{O}'_C$, *then*

$$(S * T)\hat{} = \hat{T}\hat{S}. \tag{3.32}$$

If $f \in \mathscr{O}_M$ *and* T *in* $\mathscr{S}'$, *then*

$$(fT)\hat{} = (2\pi)^{-n} \hat{f} * \hat{T}.$$

PROOF. First note that if $f, g \in \mathscr{S}$, then $f * g \in \mathscr{S}$ and

$$(f * g)\hat{}(\xi) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} f(x-y) g(y)\, dy\, dx$$
$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{-i(x-y)\cdot\xi} f(x-y) e^{-iy\cdot\xi} g(y)\, dx\, dy.$$

Changing variables $x \mapsto x + y$ gives

$$(f * g)\hat{}(\xi) = \hat{f}(\xi)\hat{g}(\xi).$$

Applying this to $\check{f}$ and $\check{g}$ implies

$$(fg)\hat{}(\xi) = \mathscr{F}^2(\check{f} * \check{g})(\xi) = (2\pi)^n (\check{f} * \check{g})(-\xi)$$
$$= (2\pi)^{-n} \int_{\mathbb{R}^n} \hat{f}(-\eta)\hat{g}(\xi+\eta) = (2\pi)^{-n}(\hat{f} * \hat{g})(\xi).$$

Let now $S \in \mathscr{S}'$ and $g \in \mathscr{S}$. We compute

$$(S * g)\hat{}(\varphi) = (S * g)(\hat{\varphi}) = S(\tilde{g} * \hat{\varphi}) = (2\pi)^n S((\check{\tilde{g}}\varphi)\hat{})$$
$$= \hat{S}(\hat{g}\varphi) = (\hat{g}\hat{S})(\varphi)$$

and

$$(gS)\hat{}(\varphi) = S(g\hat{\varphi}) = S((\check{g} * \varphi)\hat{}) = \hat{S}((2\pi)^{-n}\tilde{\check{g}} * \varphi) = (2\pi)^{-n}(\hat{S} * \hat{g})(\varphi).$$

Finally, if $S \in \mathscr{S}'$ and $T \in \mathscr{O}_C'$, then $S * T \in \mathscr{S}'$ and we have

$$(S * T)\hat{}(\varphi) = (S * T)(\hat{\varphi}) = S(\tilde{T} * \hat{\varphi}) = (2\pi)^n S((\varphi\check{\tilde{T}})\hat{})$$
$$= \hat{S}(\hat{T}\varphi) = (\hat{T}\hat{S})(\varphi).$$

□

3.10. Fundamental solutions

In this section we discuss how convolution can be used for solving partial differential equations. We only consider constant coefficient partial differential operators in $\mathbb{R}^n$, that is, operators of the form

$$P(D) = \sum_{|\alpha| \leq N} a_\alpha D^\alpha$$

where a_α are complex numbers. Note that if $u \in \mathscr{D}'(\mathbb{R}^n)$, then $P(D)u$ makes sense as an element of $\mathscr{D}'(\mathbb{R}^n)$.

DEFINITION. Let $P(D)$ be a constant coefficient differential operator in $\mathbb{R}^n$. A distribution $E \in \mathscr{D}'(\mathbb{R}^n)$ is called a *fundamental solution* for $P(D)$ if

$$P(D)E = \delta_0.$$

We note that fundamental solutions are not unique, since if E is a fundamental solution of $P(D)$ then so is $E+v$ for any $v \in \mathscr{D}'(\mathbb{R}^n)$ satisfying $P(D)v = 0$. The next result shows how fundamental solutions can be used in PDE theory in two ways: in producing solutions to $P(D)u = f$ for $f \in \mathscr{E}'$, and in studying properties of solutions $u \in \mathscr{E}'$ of $P(D)u = f$.

Theorem 3.10.1. *Let $P(D)$ be a constant coefficient partial differential operator, and let E be a fundamental solution for $P(D)$. Then for any $f \in \mathscr{E}'(\mathbb{R}^n)$â the equation*

$$P(D)u = f \qquad \text{in } \mathbb{R}^n$$

has a solution $u = E * f \in \mathscr{D}'(\mathbb{R}^n)$. Moreover, if $u \in \mathscr{E}'$ satisfies $P(D)u = f$ for some $f \in \mathscr{E}'$, then u can be represented as

$$u = E * f.$$

Proof. The first fact follows since the convolution of distributions $E \in \mathscr{D}'$ and $f \in \mathscr{E}'$ is in $\mathscr{D}'$, and

$$P(D)(E * f) = (P(D)E) * f = \delta_0 * f = f.$$

Also, if $u \in \mathscr{E}'$ satisfies $P(D)u = f$, then

$$u = \delta_0 * u = (P(D)E) * u = E * (P(D)u) = E * f. \qquad \square$$

The following basic result shows that fundamental solutions always exist.

Theorem 3.10.2. *(Malgrange-Ehrenpreis) Any constant coefficient partial differential operator has a fundamental solution.*

The previous theorem does not give much information on the properties of fundamental solutions. In the remainder of this section we will discuss briefly the fundamental solutions of four classical linear PDE:

$$\begin{aligned} \Delta u &= 0 && \text{(Laplace equation)} \\ (\partial_t - \Delta)u &= 0 && \text{(heat equation)} \\ (\partial_t^2 - \Delta)u &= 0 && \text{(wave equation)} \\ (i\partial_t + \Delta)u &= 0 && \text{(Schrödinger equation)} \end{aligned}$$

For the Laplacian we try to find a fundamental solution $E \in \mathscr{S}'(\mathbb{R}^n)$, and note that the Fourier transform implies

$$-\Delta E = \delta_0 \quad \Longleftrightarrow \quad |\xi|^2 \hat{E} = 1.$$

Thus formally (at least if $n \geq 3$)

$$E(x) = \mathscr{F}^{-1}\left\{\frac{1}{|\xi|^2}\right\}(x) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\frac{1}{|\xi|^2}\,d\xi, \qquad x \in \mathbb{R}^n.$$

The function $\frac{1}{|\xi|^2}$ is radial and homogeneous of degree -2, thus by properties of the Fourier transform E should be radial and homogeneous of degree $2-n$. Thus we guess that $E(x)$ would be given by

$$E(x) = \frac{c_n}{|x|^{n-2}}.$$

This function satisfies $\Delta E = 0$ in $\mathbb{R}^n \setminus \{0\}$, since we may express the Laplacian in polar coordinates (r,ω) as

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{n-1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_\omega$$

where Δ_ω is the Laplacian on S^{n-1} and only acts in the ω variable. Then it is easy to check that $E(r) = c_n r^{2-n}$ satisfies $\Delta E = 0$ for $r > 0$. If $n = 2$, we try a radial solution $E(r)$ and compute for $r > 0$

$$\Delta E = \partial_r^2 E + \frac{1}{r}\partial_r E = (\partial_r + \frac{1}{r})(\partial_r E).$$

The equation $(\partial_r + \frac{1}{r})v = 0$ has the solution $v = c\frac{1}{r}$, thus we guess that

$$E(x) = c_2 \log|x|.$$

The following theorem makes these formal computations precise.

THEOREM 3.10.3. *(Fundamental solution of the Laplace equation) Define*

$$E(x) = \begin{cases} -\frac{1}{2\pi}\log|x|, & n = 2, \\ \frac{1}{(n-2)\beta(n)}|x|^{2-n}, & n \geq 3, \end{cases}$$

where $\beta(n) = |S^{n-1}|$. *Then* $E \in \mathscr{S}'(\mathbb{R}^n)$ *and* $-\Delta E = \delta_0$.

PROOF. We only do the case $n \geq 3$. First let χ_B be the characteristic function of the unit ball, and write

$$|x|^{2-n} = \chi_B|x|^{2-n} + (1-\chi_B)|x|^{2-n}$$

where $\chi_B|x|^{2-n} \in L^1(\mathbb{R}^n)$ and $(1-\chi_B)|x|^{2-n} \in L^\infty(\mathbb{R}^n)$. Thus $E \in L^1 + L^\infty$, and in particular $E \in \mathscr{S}'$.

The identity $-\Delta E = \delta_0$ means that

$$\langle E, \Delta\varphi\rangle = -\varphi(0), \qquad \varphi \in C_c^\infty(\mathbb{R}^n).$$

Let $\mathrm{supp}(\varphi) \subset B(0,R)$. Since $E \in L^1_{loc}$, we have

$$(n-2)\beta(n)\langle E, \Delta\varphi\rangle = \lim_{\varepsilon\to 0}\int_{\varepsilon<|x|<R} |x|^{2-n}\Delta\varphi(x)\,dx.$$

We use the integration by parts formula

$$\int_\Omega u\partial_j v\,dx = \int_{\partial\Omega} uv\nu_j\,dS - \int_\Omega (\partial_j u)v\,dx, \qquad u,v \in C^1(\overline{\Omega}).$$

This implies

$$\begin{aligned}&(n-2)\beta(n)\langle E, \Delta\varphi\rangle \\ &\quad= \lim_{\varepsilon\to 0}\left(-\int_{\partial B(0,\varepsilon)} |x|^{2-n}\Delta\varphi(x)\,dS - \int_{\varepsilon<|x|<R} \nabla(|x|^{2-n})\cdot\nabla\varphi\,dx\right).\end{aligned}$$

The boundary integral goes to 0 as $\varepsilon \to 0$. Integrating by parts again, and using that $\Delta(|x|^{2-n}) = 0$ in $\mathbb{R}^n \setminus \{0\}$, gives that

$$(n-2)\beta(n)\langle E, \Delta\varphi\rangle = \lim_{\varepsilon\to 0}\int_{\partial B(0,\varepsilon)} \partial_\nu(|x|^{2-n})\varphi\,dS.$$

Here $\partial_\nu(|x|^{2-n}) = (2-n)|x|^{1-n}x/|x|\cdot\nu = (2-n)\varepsilon^{1-n}$ on $\partial B(0,\varepsilon)$. Therefore

$$(n-2)\beta(n)\langle E, \Delta\varphi\rangle = (2-n)\lim_{\varepsilon\to 0}\frac{1}{\varepsilon^{n-1}}\int_{\partial B(0,\varepsilon)} \varphi\,dS = (2-n)\beta(n)\varphi(0).$$

This is the required result. □

Now consider the heat equation,

$$(\partial_t - \Delta)u(t,x) = 0, \qquad u(0,x) = f(x).$$

If we denote by ˆ the partial Fourier transform with respect to the x variable, Fourier transforming the equation gives

$$(\partial_t + |\xi|^2)\hat{u}(t,\xi) = 0, \qquad \hat{u}(0,\xi) = \hat{f}(\xi).$$

This is a first order ODE in the t variable, and it has the solution

$$\hat{u}(t,\xi) = e^{-t|\xi|^2}\hat{f}(\xi).$$

Thus u should be given by

$$u(t,x) = (\mathscr{F}_\xi^{-1}\{e^{-t|\xi|^2}\} * f)(x).$$

We have computed earlier that $\mathscr{F}^{-1}\{e^{-\frac{1}{2}|x|^2}\} = (2\pi)^{-n/2}e^{-\frac{1}{2}|x|^2}$. Using the scaling property of Fourier transform, the function $\mathscr{F}_\xi^{-1}\{e^{-t|\xi|^2}\}$ is equal to

$$K(t,x) = (4\pi t)^{-n/2}e^{-\frac{1}{4t}|x|^2}.$$

The function $K(t,x)$ is called the *heat kernel* in $\mathbb{R}^n$, and the solution of the heat equation is given by

$$u(t,x) = \int_{\mathbb{R}^n} K(t,x-y)f(y)\,dy.$$

THEOREM 3.10.4. *(a) Let $f \in \mathscr{S}'(\mathbb{R}^n)$, and consider the problem*

$$(\partial_t - \Delta)u = 0 \text{ in } (0,\infty)\times\mathbb{R}^n, \qquad u(0) = f.$$

There is a unique solution $u \in C^\infty((0,\infty)\times\mathbb{R}^n) \cap C^\infty([0,\infty),\mathscr{S}'(\mathbb{R}^n))$ given by

$$u(t,\,\cdot\,) = K(t,\,\cdot\,) * f, \qquad t > 0.$$

(b) The function

$$E(t,x) = \begin{cases} K(t,x), & t > 0, \\ 0, & t \leq 0 \end{cases}$$

is in $L^1_{loc}(\mathbb{R}^{n+1}) \cap C^\infty(\mathbb{R}^{n+1}\setminus\{0\})$, and it is a fundamental solution of the heat operator in the sense that

$$(\partial_t - \Delta)E = \delta_0 \qquad \text{in } \mathbb{R}^{n+1}.$$

Now consider the wave equation

$$(\partial_t^2 - \Delta)u = 0, \qquad u(0) = f, \quad \partial_t u(0) = g.$$

As for the heat equation, we take Fourier transforms in x:

$$(\partial_t^2 + |\xi|^2)\hat{u}(t,\xi) = 0, \qquad \hat{u}(0) = \hat{f}, \quad \partial_t\hat{u}(0) = \hat{g}.$$

If ξ is fixed this is an ODE, and its solution is given by

$$\hat{u}(t,\xi) = c_1(\xi)\cos(|\xi|t) + c_2(\xi)\sin(|\xi|t)$$

for some constants $c_j(\xi)$. By using the initial conditions we get

$$c_1(\xi) = \hat{f}(\xi), \qquad |\xi|c_2(\xi) = \hat{g}(\xi).$$

THEOREM 3.10.5. *(a) Let $f, g \in \mathscr{S}'(\mathbb{R}^n)$, and consider*

$$(\partial_t^2 - \Delta)u = 0, \qquad u(0) = f, \quad \partial_t u(0) = g.$$

This problem has a unique solution $u \in C^\infty(\mathbb{R},\mathscr{S}'(\mathbb{R}^n))$ given by

$$u(t) = C(t)f + S(t)g$$

where $C(t)$ *and* $S(t)$ *are the* cosine *and* sine propagators

$$C(t)f = \mathscr{F}^{-1}\{\cos(t|\xi|)\hat{f}\}, \qquad S(t)f = \mathscr{F}^{-1}\{\frac{\sin(t|\xi|)}{|\xi|}\hat{f}\}.$$

(b) Let

$$E(t,\,\cdot\,) = \mathscr{F}^{-1}\{\frac{\sin(t|\xi|)}{|\xi|}\}.$$

This gives rise to a distribution in $\mathbb{R}^{n+1}$ *which is a fundamental solution of the wave operator in the sense that*

$$(\partial_t^2 - \Delta)E = \delta.$$

One has for $t > 0$

$$E(t,x) = \begin{cases} \frac{1}{2}\chi_{(-t,t)}(x), & n = 1, \\ \frac{1}{2\pi}\frac{1}{\sqrt{t^2-|x|^2}}\chi_{\{|x|<t\}}, & n = 2, \\ \frac{1}{4\pi t}\delta(t-|x|), & n = 3. \end{cases}$$

Bibliography

[Du] J. Duoandikoetxea: *Fourier analysis.* Graduate Studies in Mathematics **29**. AMS, Providence Rhode Island, 2001.

[Hö] L. Hörmander: *The analysis of linear partial differential operators I.* Grundlehren der mathematischen Wissenschaften **256**. Second edition, Springer-Verlag, Berlin Heidelberg, 1990.

[Ru] W. Rudin: *Functional analysis.* McGraw-Hill, New York, 1973.

[Sc] L. Schwartz: *Théorie des distributions.* Hermann, Paris, 1966.

[SS] E. Stein, R. Sharkarchi: *Fourier analysis: an introduction.* Princeton Lectures in Analysis **I**. Princeton University Press, Princeton, 2003.

[St] R. Strichartz: *A guide to distribution theory.* Studies in Advanced Mathematics. CRC Press, Boca Raton, 1994.