LEBESGUE, FOURIER AND SOBOLEV (NOTES FOR FUNCTIONAL ANALYSIS)

CHRIS WENDL

ABSTRACT. The goal of these notes is to establish the basic properties of a sufficiently wide range of function spaces so as to have a wealth of interesting examples on hand for results in abstract functional analysis. The reader is assumed to be familiar with the essentials of measure theory (including dominated convergence and Fubini's theorem, the change of variables formula, the definition and completeness of the L^p-spaces, the Hölder and Minkowski inequalities), and some basic facts about Banach and Hilbert spaces and bounded linear operators (e.g. the fact that dual spaces of Banach spaces are also Banach spaces, but not any of the deeper results such as the Baire category or open mapping theorem).

CONTENTS

0. Preliminaries 3
0.1. Integrals of vector-valued functions 3
0.2. Differentiation under the integral sign 4
0.3. Some standard function spaces 5
1. Uniform convexity 9
1.1. Convexity in Banach spaces 9
1.2. Orthogonal complements in Hilbert space 12
2. Duality in L^p-spaces 14
2.1. The pairing of L^p and L^q 14
2.2. Differentiability of the norm 16
2.3. Uniform convexity of L^p 17
2.4. Proof of the representation theorem 19
3. Separability of L^p 22
4. Weak convergence 25
5. Mollification 30
5.1. Continuity under translation 30
5.2. Convolution and regularity 31
5.3. Young's inequality 32
5.4. Approximate identities 33
6. Absolute continuity 37
6.1. The fundamental theorem of calculus 37
6.2. The Lebesgue differentiation theorem 39
6.3. Maximal functions and weak L^1 41
6.4. The Radon-Nikodým theorem 43
6.5. Absolutely continuous functions are integrals 46
7. Fourier series 51
7.1. Fully periodic functions 51
7.2. Function spaces on the torus and the lattice 51
7.3. The transformations $\mathscr{F}$ and $\mathscr{F}^*$ 55
7.4. Fourier series and derivatives 57
7.5. Approximate identities 58
7.6. Completeness 60
7.7. Parseval's identity 61

8. The Fourier transform 62
8.1. The Fourier transform on the Schwartz space 62
8.2. Fourier transforms and derivatives 64
8.3. The Gaussian 64
8.4. Approximate identities revisited 65
8.5. The Fourier inversion formula 66
8.6. Plancherel's theorem 67
8.7. Convolutions 67
8.8. Nowhere differentiable functions 69
9. Sobolev spaces via Fourier analysis 72
9.1. The general idea of Sobolev spaces 72
9.2. The spaces $H^m(\mathbb{R}^n)$ and $H^m(\mathbb{T}^n)$ 73
9.3. The Sobolev embedding theorem 75
9.4. Compact inclusions 76
9.5. Approximation by smooth functions 78
9.6. Hölder estimates 79
9.7. Elliptic regularity 83
10. Distributions 86
10.1. Weak derivatives 86
10.2. Test functions and the space of distributions 88
10.3. Smoothness of distributions 91
10.4. Product distributions 97
10.5. The Sobolev spaces $W^{m,p}(\Omega)$ 99
10.6. Tempered distributions and Fourier transforms 103
10.7. Distributions with compact support 106
10.8. Appendix: The topology of the space of test functions 108
References 119

0. Preliminaries

0.1. **Integrals of vector-valued functions.** In most of the following, we choose a field $\mathbb{K}$ to be either $\mathbb{R}$ or $\mathbb{C}$ and consider functions with values in a fixed finite-dimensional inner product space $(V, \langle\ ,\ \rangle)$ over $\mathbb{K}$, with norm denoted by

$$|\cdot| := \sqrt{\langle\cdot,\cdot\rangle}.$$

The discussion of Fourier analysis starting in §7 will require choosing $\mathbb{K} = \mathbb{C}$, but in most other places, the differences between the real and complex cases will be negligible, e.g. we will sometimes need to use the relation

$$\langle v + w, v + w\rangle = |v|^2 + 2\,\mathrm{Re}\langle v, w\rangle + |w|^2,$$

which is true in both cases, the difference being only that in the real case, the symbol "Re" is redundant. We adopt the convention that a complex inner product is antilinear in its first argument and linear in its second:

$$\langle iv, w\rangle = -i\langle v, w\rangle, \qquad \langle v, iw\rangle = i\langle v, w\rangle.$$

We will sometimes make use of the fact that a complex vector space is also a real vector space (of twice the dimension).

Convention. By the standard definition, a **measure space** $(X, \mathcal{A}, \mu)$ consists of three pieces of data: a set X, a σ-algebra $\mathcal{A} \subset 2^X$ and a measure $\mu : \mathcal{A} \to [0, \infty]$. Since we will almost never have occasion to talk about the σ-algebra itself, we shall typically omit it from the notation and simply call (X, μ) a measure space, referring when necessary to the elements of $\mathcal{A}$ as the **measurable** (or **μ-measurable**) **sets**.

Given a measure space (X, μ), a function $f : X \to V$ is considered **measurable** if it is Borel measurable, meaning the preimage of every open subset of V is μ-measurable in X. It is easy to show that if we choose any real basis $e_1, \dots, e_n$ of V and write $f = \sum_{j=1}^n f_j e_j$ for functions $f_j : X \to \mathbb{R}$, then f is measurable if and only if all of the f_j are measurable. Similarly, if f is measurable then $|f| : X \to [0, \infty)$ is also measurable, and in this case the component functions f_j are μ-integrable if and only if $\int_X |f|\, d\mu < \infty$. One can then define the vector-valued integral

$$\int_X f\, d\mu = \sum_{j=1}^n \left(\int_X f_j\, d\mu \right) e_j \in V. \tag{0.1}$$

We will sometimes also write $\int_X f(x)\, d\mu(x) := \int_X f\, d\mu$ when we want to specify the name of the variable $x \in X$.

Exercise 0.1. Show that for μ-integrable functions $f : X \to V$, the integral $\int_X f\, d\mu \in V$ defined above is independent of the choice of real basis $e_1, \dots, e_n \in V$.

Exercise 0.2. Show that for every μ-integrable function $f : X \to V$, $\left|\int_X f\, d\mu\right| \leqslant \int_X |f|\, d\mu$.

The simplest example beyond $V = \mathbb{R}$ is $V = \mathbb{C}$ with the standard inner product $\langle v, w\rangle := \bar{v}w$. Here we can take $e_1 := 1$ and $e_2 := i$ as a real basis of $\mathbb{C}$, so $f : X \to \mathbb{C}$ is measurable/integrable if and only if its real and imaginary parts are both measurable/integrable, and (0.1) becomes

$$\int_X f\, d\mu = \int_X (\mathrm{Re}\, f)\, d\mu + i \int_X (\mathrm{Im}\, f)\, d\mu \in \mathbb{C}.$$

Remark 0.3. The assumption $\dim V < \infty$ is inessential for much of what follows, though obviously the definition of $\int_X f\, d\mu \in V$ requires some modification if V has no finite basis. A definition (using approximation by step functions) for the case where V is an arbitrary Banach space may be found in [Lan93]. Since many details become more complicated in this more general setting, we will stick to the case $\dim V < \infty$ but give occasional remarks on what needs to be modified in order to lift this assumption.

0.2. **Differentiation under the integral sign.** The following standard consequence of the dominated convergence theorem will be an essential tool to have at our disposal.

Theorem 0.4. *Suppose (Y, ν) is a measure space, M is a metric space, and $\varphi : M \times Y \to V$ is a function with the following properties:*

(1) For every $x \in M$, the function $\varphi(x, \cdot) : Y \to V$ is measurable and satisfies $|\varphi(x, \cdot)| \leqslant \psi$ for some fixed ν-integrable function $\psi : Y \to [0, \infty]$ independent of x;
(2) For every $y \in Y$, the function $\varphi(\cdot, y) : M \to V$ is continuous.

Then the function $F : M \to V$ given by

$$F(x) := \int_Y \varphi(x, \cdot)\, d\nu$$

is continuous. If additionally M is an open subset of $\mathbb{R}^m$ with coordinates $x = (x_1, \ldots, x_m)$ and the partial derivatives $\frac{\partial \varphi}{\partial x_j} : M \times Y \to V$ exist for every $j = 1, \ldots, m$ and also satisfy the two conditions above, then F is continuously differentiable and satisfies

$$\partial_j F(x) = \int_Y \frac{\partial \varphi}{\partial x_j}(x, \cdot)\, d\nu$$

for every $x \in M$ and $j = 1, \ldots, m$.

Proof. To prove $F : M \to V$ is continuous at a point $x \in M$, consider a sequence $x_n \in M$ with $x_n \to x$. Since $\varphi(\cdot, y) : M \to V$ is continuous for every $y \in Y$, the sequence of functions $\varphi(x_n, \cdot) : Y \to \mathbb{R}$ converges pointwise to $\varphi(x, \cdot) : Y \to \mathbb{R}$, and by assumption it also satisfies

$$|\varphi(x_n, \cdot)| \leqslant \psi \quad \text{for all } n$$

for a fixed ν-integrable function $\psi : Y \to [0, \infty]$. The dominated convergence theorem thus implies $F(x_n) \to F(x)$.

Now suppose additionally that $M = \mathcal{U} \subset \mathbb{R}^m$ is open and $\frac{\partial \varphi}{\partial x_j}(x, y)$ exists for all $(x, y) \in \mathcal{U} \times Y$ and defines a function that is (for each fixed $y \in Y$) continuous with respect to $x \in \mathcal{U}$ and (for each fixed $x \in \mathcal{U}$) measurable with respect to $y \in Y$, additionally satisfying the bound $\left| \frac{\partial \varphi}{\partial x_j}(x, \cdot) \right| \leqslant \psi$ for all $x \in \mathcal{U}$. Let $e_1, \ldots, e_m$ denote the standard basis of $\mathbb{R}^m$. The partial derivative $\frac{\partial \varphi}{\partial x_j}(x, y)$ is then the limit as $h \to 0$ of the difference quotients

$$D_j^h \varphi(x, y) := \frac{\varphi(x + h e_j, y) - \varphi(x, y)}{h} \in V,$$

where for each $x \in \mathcal{U}$, the function $D_j^h \varphi(x, \cdot) : Y \to V$ is defined for all $h \in \mathbb{R} \backslash \{0\}$ sufficiently close to 0. For any sequence $h_n \in \mathbb{R} \backslash \{0\}$ with $h_n \to 0$, we therefore have

$$D_j^{h_n} \varphi(x, \cdot) \to \frac{\partial \varphi}{\partial x_j}(x, \cdot) \quad \text{pointwise on } Y. \tag{0.2}$$

For every $y \in Y$ and $h \in \mathbb{R}$ sufficiently close to 0, the fact that $\varphi(\cdot, y)$ is continuously differentiable with respect to x_j allows us to derive a formula for $D_j^h \varphi(x, y)$ using the fundamental theorem of calculus: we have

$$\varphi(x + h e_j, y) = \varphi(x, y) + \int_0^1 \frac{d}{dt} \varphi(x + t h e_j, y)\, dt = \varphi(x, y) + h \int_0^1 \frac{\partial \varphi}{\partial x_j}(x + t h e_j, y)\, dt,$$

and thus

$$D_j^h \varphi(x, y) = \int_0^1 \frac{\partial \varphi}{\partial x_j}(x + t h e_j, y)\, dt, \tag{0.3}$$

giving rise to the bound

$$\left| D_j^h \varphi(x, y) \right| \leqslant \int_0^1 \psi(y)\, dt = \psi(y).$$

Since ψ is integrable, one can again apply the dominated convergence theorem and obtain a convergence result for the corresponding difference quotients of F: for any sequence $h_n \in \mathbb{R}\backslash\{0\}$ with $h_n \to 0$, we have

$$D_j^{h_n} F(x) := \frac{F(x + h_n e_j) - F(x)}{h_n} = \int_Y D_j^{h_n} \varphi(x, \cdot)\, d\nu \to \int_Y \frac{\partial \varphi}{\partial x_j}(x, \cdot)\, d\nu$$

Since the sequence h_n was arbitrary, this proves

$$\frac{\partial F}{\partial x_j}(x) = \lim_{h \to 0} D_j^h F(x) = \int_Y \frac{\partial \varphi}{\partial x_j}(x, \cdot)\, d\nu,$$

and the continuity of $\frac{\partial F}{\partial x_j}$ now follows from the same argument as the continuity of F. $\square$

Remark 0.5. The hypotheses of Theorem 0.4 can be weakened (at the cost of more cumbersome notation) in various ways that are sometimes useful. Most importantly, since the continuity and differentiability of F are purely local conditions, the bounds $|\varphi(x, \cdot)| \leqslant \psi$ and $\left|\frac{\partial \varphi}{\partial x_j}(x, \cdot)\right| \leqslant \psi$ do not really need to hold with a single function ψ for every $x \in M$; it suffices if every $x_0 \in M$ has a neighborhood $\mathcal{U} \subset M$ and an associated integrable function $\psi_{x_0} : Y \to [0, \infty]$ that bounds these functions for all $x \in \mathcal{U}$. One can also insert the words "almost everywhere" in various places among the hypotheses, so that certain steps in the proof make sense only after deleting sets of measure zero from Y, which is harmless. For more elaborate versions of the statement, see e.g. [AE01, Theorems 3.17 and 3.18]) or [Wen19].

0.3. **Some standard function spaces.** We shall assume basic knowledge of the spaces of L^p-functions on measure spaces and C^m-functions on domains in Euclidean space. Let us clarify the essential definitions.

Assume (X, μ) is an arbitrary measure space, and $(V, \langle\ ,\ \rangle)$ is again a finite-dimensional inner product space over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ with norm $|\cdot| := \sqrt{\langle \cdot, \cdot \rangle}$. The L^p-norm of a measurable function $f : X \to V$ is defined for each $p \in [1, \infty)$ by

$$\|f\|_{L^p} := \|f\|_{L^p(X)} := \left(\int_X |f(x)|^p \, d\mu(x) \right)^{1/p} \in [0, \infty],$$

and for the case $p = \infty$,

$$\|f\|_{L^\infty} := \|f\|_{L^\infty(X)} := \operatorname{ess\,sup}_{x \in X} |f(x)| := \inf \left\{ c \geqslant 0 \;\middle|\; |f| \leqslant c \text{ almost everywhere} \right\} \in [0, \infty].$$

We assume the reader is familiar with the standard Minkowski and Hölder inequalities, and the fact that the space $L^p(X, \mu)$ of equivalence classes of measurable functions (defined almost everywhere) with finite L^p-norms is a Banach space. We will typically abbreviate

$$L^p(X) := L^p(X, \mu)$$

when the measure is clear from context. Here is a precise statement of the completeness theorem:

Theorem 0.6 (see e.g. [Sal16, §4.2])**.** *For $1 \leqslant p \leqslant \infty$, every L^p-Cauchy sequence $f_n \in L^p(X)$ is L^p-convergent and also has a pointwise almost everywhere convergent subsequence. In the case $p = \infty$, the original sequence also converges pointwise almost everywhere.* $\square$

The usual Hölder inequality for real-valued functions combines with the Cauchy-Schwarz inequality on $(V, \langle\ ,\ \rangle)$ and Exercise 0.2 to give the relation

$$\left| \int_X \langle f(x), g(x) \rangle \, d\mu(x) \right| \leqslant \int_X |\langle f(x), g(x) \rangle| \, d\mu(x) \leqslant \|f\|_{L^p} \cdot \|g\|_{L^q}$$

for $f \in L^p(X)$ and $g \in L^q(X)$ with $\frac{1}{p} + \frac{1}{q} = 1$.

Exercise 0.7. In case you have only seen $L^p(X)$ defined for real-valued functions before, convince yourself that the proof of Theorem 0.6 still goes through when the functions in $L^p(X)$ take values in an arbitrary (real or complex) finite-dimensional vector space.

Exercise 0.8. Show that for every measurable function $f : X \to V$, $\|f\|_{L^\infty} \leqslant \liminf_{p\to\infty} \|f\|_{L^p}$, and if additionally either $\mu(X) < \infty$ or $f \in L^r(X)$ for some $r \in [1,\infty)$, then $\|f\|_{L^\infty} = \lim_{p\to\infty} \|f\|_{L^p}$.
Hint: For the case with $f \in L^r(X)$ for some $r < \infty$, show that $\|f\|_{L^p} \leqslant \|f\|_{L^r}^{r/p} \cdot \|f\|_{L^\infty}^{1-r/p}$ holds for every $p > r$. (Note that this is not a version of Hölder's inequality—it is easier.) Use this to bound $\limsup_{p\to\infty} \|f\|_{L^p}$.

When X is an open subset of Euclidean space

$$X := \Omega \subset \mathbb{R}^n \qquad \text{with} \qquad \mu := m \text{ (Lebesgue measure)},$$

it is often useful to consider functions that need not be in $L^p(\Omega)$ but restrict to L^p-functions on all compact subsets. Since compact subsets of $\mathbb{R}^n$ are bounded and therefore have finite measure, this includes for instance the nontrivial constant functions, which are not in $L^p(\Omega)$ unless $m(\Omega) < \infty$. We define the vector space

$$L^p_{\text{loc}}(\Omega) := \left\{ f : \Omega \to V \;\middle|\; f|_K \in L^p(K) \text{ for all } K \subset \Omega \text{ compact} \right\} / \sim,$$

where as usual the equivalence relation $f \sim g$ means $f = g$ almost everywhere on Ω. The functions in $L^p_{\text{loc}}(\Omega)$ are said to be **locally of class** L^p, and in the case $p = 1$, a function $f \in L^1_{\text{loc}}(\Omega)$ is called **locally integrable** on Ω. The space $L^p_{\text{loc}}(\Omega)$ is strictly larger than $L^p(\Omega)$, and it is not a Banach space since there is no single norm to determine whether or not a given function is of class L^p_{loc}. It does however have a natural topology as a locally convex space, defined via the family of seminorms

$$\|f\|_{L^p(K)} = \|f|_K\|_{L^p}, \tag{0.4}$$

where K ranges over the set of all compact subsets $K \subset \Omega$. Note that these are seminorms rather than norms, because a function $f \in L^p_{\text{loc}}(\Omega)$ may be nontrivial but satisfy $\|f\|_{L^p(K)} = 0$ because it vanishes almost everywhere on K. Convergence of a sequence $f_j \to f$ in $L^p_{\text{loc}}(\Omega)$ means that $\|f - f_j\|_{L^p(K)} \to 0$ is satisfied for all of these seminorms, which is equivalent to saying that the restrictions of f_j to every compact subset $K \subset \Omega$ are convergent in $L^p(K)$ to $f|_K$.

It is possible to derive the topology of $L^p_{\text{loc}}(\Omega)$ from a countable subfamily of the seminorms in (0.4). Indeed, Ω can always be covered by a nested sequence

$$\Omega_1 \subset \bar{\Omega}_1 \subset \Omega_2 \subset \bar{\Omega}_2 \subset \ldots \subset \bigcup_{m\in\mathbb{N}} \Omega_m = \Omega$$

of open subsets $\Omega_m \subset \Omega$ with compact closures $K_m := \bar{\Omega}_m$, so that any compact subset $K \subset \Omega$ is contained in Ω_m for $m \in \mathbb{N}$ sufficiently large. For a concrete construction of Ω_m, one can for instance define $\Omega_m := \left\{ x \in \Omega \;\middle|\; |x| < m \text{ and } \operatorname{dist}(x, \mathbb{R}^n \backslash \Omega) > 1/m \right\}$, where for two subsets $A, B \subset \mathbb{R}^n$, we denote

$$\operatorname{dist}(A, B) := \inf \left\{ |x - y| \;\middle|\; x \in A,\ y \in B \right\}.$$

A sequence $f_j \in L^p_{\text{loc}}(\Omega)$ is then L^p_{loc}-convergent if and only if it converges in each of the seminorms $\|\cdot\|_{L^p(K_m)}$ for $m \in \mathbb{N}$, and similarly, every open subset of $L^p_{\text{loc}}(\Omega)$ is a union of sets of the form $\left\{ f \in L^p_{\text{loc}}(\Omega) \;\middle|\; \|f - f_0\|_{L^p(K_m)} < \epsilon \right\}$ for $f_0 \in L^p_{\text{loc}}(\Omega)$, $m \in \mathbb{N}$ and $\epsilon > 0$. It follows (see e.g. [RS80, Theorem V.5]) that $L^p_{\text{loc}}(\Omega)$ is metrizable, with open subsets defined via the metric

$$d(f, g) := \sum_{m=1}^{\infty} \frac{1}{2^m} \frac{\|f - g\|_{L^p(K_m)}}{1 + \|f - g\|_{L^p(K_m)}}.$$

In fact, $L^p_{\text{loc}}(\Omega)$ is a Fréchet space: completeness follows from the completeness of the Banach space $L^p(K_m)$ for every m, as a sequence $f_j \in L^p_{\text{loc}}(\Omega)$ is Cauchy if and only if $f_j|_{K_m}$ is Cauchy in $L^p(K_m)$ for every m.

Continuing under the assumption that $\Omega \subset \mathbb{R}^n$ is an open subset, we shall denote

$$C^m(\Omega) := \{f : \Omega \to V \ m \text{ times continuously differentiable}\}$$

for integers $m \geqslant 0$. This is not a Banach space, but it can be made into one by imposing an extra boundedness condition. To express this properly, recall that a **multi-index** for functions on $\mathbb{R}^n$ is an n-tuple $\alpha = (\alpha_1, \dots, \alpha_n)$ of nonnegative integers, which can be used to define the differential operator

$$\partial^\alpha := \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n}, \qquad \text{where} \qquad \partial_j := \frac{\partial}{\partial x_j} \text{ for } j = 1, \dots, n.$$

The **order** of this operator (also called the order or *degree* of the multi-index) is $|\alpha| := \alpha_1 + \dots + \alpha_n$. We now define the C^m**-norm** by

$$\|f\|_{C^m} := \|f\|_{C^m(\Omega)} := \sum_{|\alpha| \leqslant m} \sup_{x \in \Omega} |\partial^\alpha f(x)|,$$

and let

$$C_b^m(\Omega) := \left\{ f \in C^m(\Omega) \;\middle|\; \|f\|_{C^m} < \infty \right\}.$$

Convergence of a sequence f_j in the C^m-norm means uniform convergence of f_j and all its derivatives up to order m. By standard results of first-year analysis, $C_b^m(\Omega)$ with this norm is a Banach space for every integer $m \geqslant 0$. A useful subspace of $C_b^m(\Omega)$ can be defined by[1]

$$C^m(\bar{\Omega}) := \left\{ f \in C_b^m(\Omega) \;\middle|\; \partial^\alpha f \text{ is uniformly continuous for all multi-indices } \alpha \text{ with } |\alpha| \leqslant m \right\}.$$

The following exercise explains the motivation for this notation.

Exercise 0.9. Let $\bar{\Omega} \subset \mathbb{R}^n$ denote the closure of the open subset $\Omega \subset \mathbb{R}^n$.

(a) Show that if $f : \Omega \to \mathbb{R}$ is uniformly continuous, then it admits a (necessarily unique) continuous extension over $\bar{\Omega}$. (Note that the converse is also true if Ω is bounded, since continuous functions on compact sets are always uniformly continuous.)

(b) Show that $C^m(\bar{\Omega})$ is a closed subspace of $C_b^m(\Omega)$, hence it is a Banach space with the C^m-norm.

In particular, $C^m(\bar{\Omega})$ can be characterized as the space of C^m-functions on Ω whose derivatives up to order m all admit bounded continuous extensions to $\bar{\Omega}$. (The word "bounded" is redundant here if Ω itself is bounded, since $\bar{\Omega}$ is then compact.)[2]

For smooth (i.e. infinitely differentiable) functions, we define

$$C^\infty(\Omega) := \bigcap_{m \geqslant 0} C^m(\Omega), \qquad C_b^\infty(\Omega) := \bigcap_{m \geqslant 0} C_b^m(\Omega),$$

and endow the latter with the locally convex topology defined via the entire sequence of norms $\|\cdot\|_{C^m}$ for $m \geqslant 0$, hence a sequence $f_j \in C_b^\infty(\Omega)$ is C^∞-convergent if and only if its derivatives of all orders are uniformly convergent. One could similarly define $C^\infty(\bar{\Omega})$, but this turns out to be the same space as $C_b^\infty(\Omega)$ since the boundedness of the derivatives of order $m+1$ implies uniform continuity for derivatives of order m. Since the family of C^m-norms for $m \geqslant 0$ is countable, one can define a metric on $C_b^\infty(\Omega)$ in the same manner that we did so for $L^p_{\text{loc}}(\Omega)$, and the completeness of $C_b^m(\Omega)$ for each $m \geqslant 0$ implies that $C_b^\infty(\Omega)$ is a Fréchet space.

The C^m-topologies also have local variants, which are defined on $C^m(\Omega)$ without requiring any boundedness condition: we say that a sequence $f_j \in C^m(\Omega)$ is C^m_{loc}**-convergent** to $f \in C^m(\Omega)$ if

$$\|f - f_j\|_{C^m(K)} := \sum_{|\alpha| \leqslant m} \max_{x \in K} |\partial^\alpha f(x) - \partial^\alpha f_j(x)| \to 0$$

for every compact subset $K \subset \Omega$. As with $L^p_{\text{loc}}(\Omega)$, one can use an exhaustion of Ω by a nested sequence of open subsets with compact closure to characterize this notion of convergence via

[1] There is potential ambiguity in the notation when $\Omega = \mathbb{R}^n$ since $\mathbb{R}^n$ is its own closure, but $C^m(\bar{\mathbb{R}}^n)$ is nonetheless a smaller space than $C^m(\mathbb{R}^n)$.

[2] If $\bar{\Omega}$ is compact and has a sufficiently "nice" boundary, meaning for instance that the boundary is a C^m-smooth submanifold of $\mathbb{R}^n$, then one can show with somewhat more effort that $C^m(\bar{\Omega})$ is the space of C^m-functions on Ω that admit extensions of class C^m over some open neighborhood of $\bar{\Omega}$; for details, see [AF03, §5.19–§5.21].

a countable family of seminorms, making $C^m(\Omega)$ into a Fréchet space with the C^m_{loc}-topology. There is similarly a C^∞_{loc}-topology on $C^\infty(\Omega)$, in which sequences converge if and only if their derivatives of *all* orders converge on compact subsets, and this endows $C^\infty(\Omega)$ with a natural Fréchet space structure. Note that for each $m \in \mathbb{N} \cup \{0, \infty\}$, C^m_{loc}-convergence is a much weaker notion than C^m-convergence, i.e. many sequences converge in C^m_{loc} but not in C^m, and the behavior of a C^m_{loc}-convergent sequence "near infinity" can be arbitrarily wild.

The **support** $\operatorname{supp}(f) \subset \Omega$ of a function $f : \Omega \to V$ is the closure of the set $\{x \in \Omega \mid f(x) \neq 0\}$. We will denote

$$C^m_0(\Omega) := \left\{ f \in C^m(\Omega) \;\middle|\; \operatorname{supp}(\Omega) \subset \Omega \text{ is compact} \right\}.$$

This is a subspace of both of the Banach spaces $C^m_b(\Omega)$ and $C^m(\bar{\Omega})$, though not a closed subspace in either case, as a sequence of functions with growing compact supports can easily be C^m-convergent to one whose support is not compact.

1. Uniform convexity

1.1. **Convexity in Banach spaces.** A subset K in a vector space X is called **convex** if K contains the line segment joining any two of its points (see Figure 1), i.e.

$$x, y \in K \quad \Rightarrow \quad tx + (1-t)y \in K \text{ for every } t \in [0,1].$$

Similarly, a function $f : K \to \mathbb{R}$ on a convex set $K \subset X$ is called **convex** if for every pair of points in its domain, the values of f along the line segment between those points are bounded by the corresponding "convex combinations" of its values at the end points (Figure 2); concretely,

$$\forall x, y \in K \text{ and } t \in [0,1], \qquad f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y). \tag{1.1}$$

It is straightforward to show that if f is convex, then $f^{-1}((-\infty, a))$ and $f^{-1}((-\infty, a])$ are convex subsets for every $a \in \mathbb{R}$. We say additionally that f is **strictly convex** if the inequality in (1.1) is strict for all $t \in (0,1)$ whenever $x \neq y$.

Example 1.1. By a standard exercise in first-year analysis, if $\mathcal{U} \subset \mathbb{R}^n$ is an open convex set, then a C^2-function $f : \mathcal{U} \to \mathbb{R}$ is convex (or strictly convex) if and only if its Hessian at every point is positive semidefinite (or positive definite, respectively).

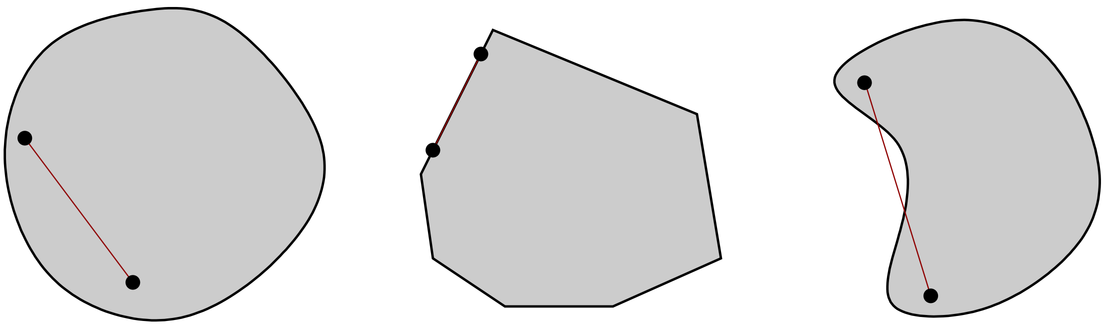

FIGURE 1. The two sets on the left are convex, while the set on the right is not. The set in the middle is convex but not *strictly convex*, i.e. it contains a segment connecting boundary points that does not stay in the interior. In particular, if this set occurs as the closed unit ball in some normed vector space, it implies that that space is not strictly (and therefore not uniformly) convex.

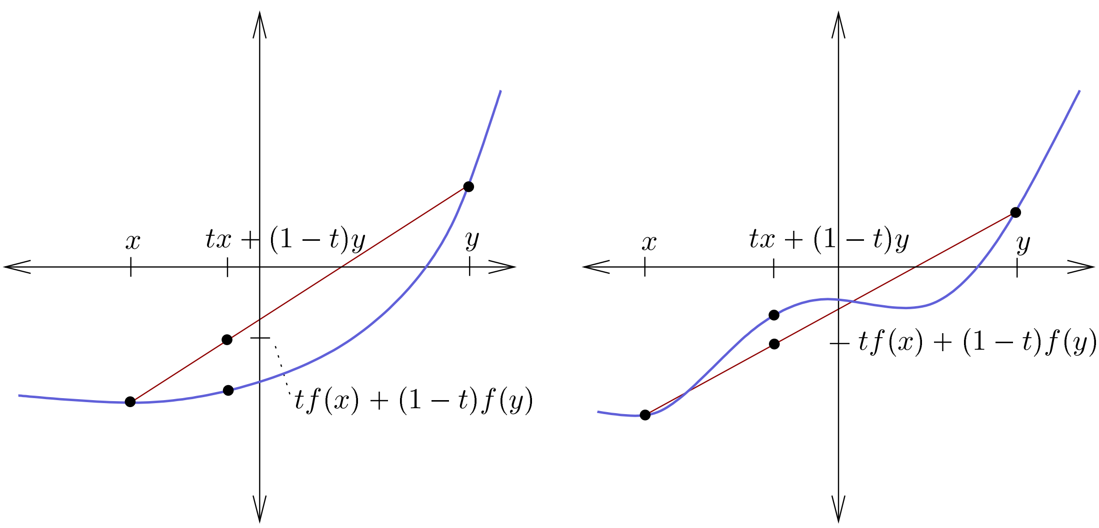

FIGURE 2. The function $f : \mathbb{R} \to \mathbb{R}$ on the left is convex, and the function on the right is not.

Example 1.2. For any normed vector space $(X, \|\cdot\|)$ and $x_0 \in X$, the triangle inequality implies that the function $X \to [0, \infty) : x \mapsto \|x - x_0\|$ is convex. As a consequence, (closed or open) balls about points in a normed vector space are always convex sets. This remains true if the norm is replaced by a seminorm, and is the reason why a topological vector space with topology generated by a family of seminorms is called a *locally convex* space.

Strict and *uniform* convexity are geometric properties of normed vector spaces that strengthen the observation in Example 1.2 about balls $B \subset X$ being convex—the idea is to require that the segment joining any two points in the ball stays in the *interior* of the ball. This is a nontrivial condition on the "shape" of the unit ball as determined by the norm, and it is not satisfied by every norm (see Exercise 1.6 below). In the following, we denote the closed unit ball and unit sphere in a normed vector space $(X, \|\cdot\|)$ by

$$\bar{B} := \{x \in X \mid \|x\| \leqslant 1\}, \quad \text{and} \quad \partial\bar{B} := \{x \in X \mid \|x\| = 1\}$$

respectively, and denote the distance between two subsets $\mathcal{U}, \mathcal{V} \subset X$ by

$$\operatorname{dist}(\mathcal{U}, \mathcal{V}) := \inf\{\|x - y\| \mid x \in \mathcal{U},\ x \in \mathcal{V}\}.$$

Definition 1.3. A normed vector space $(X, \|\cdot\|)$ is called **strictly convex** if

$$x, y \in \bar{B} \text{ with } x \neq y \quad \Rightarrow \quad tx + (1-t)y \in \bar{B}\backslash\partial\bar{B} \quad \forall t \in (0,1).$$

The middle picture in Figure 1 gives an example of something one might imagine the unit ball looking like in a normed vector space that is not strictly convex. The next definition amounts to a quantitative version of strict convexity, in which the distance of the midpoint between x and y to the boundary cannot become arbitrarily small unless x and y are close.

Definition 1.4. A normed vector space $(X, \|\cdot\|)$ is called **uniformly convex** if for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$x, y \in \bar{B} \text{ with } \|x - y\| \geqslant \epsilon \qquad \Rightarrow \qquad \operatorname{dist}\left(\frac{x+y}{2}, \partial\bar{B}\right) \geqslant \delta.$$

Observe that every uniformly convex space is clearly also strictly convex.

Remark 1.5. The definition of uniform convexity appears in many references with a weaker condition on x and y, namely that they lie in $\partial\bar{B}$ instead of $\bar{B}$. The resulting notion is equivalent to our definition; for a proof of this, see [Tes, Lemma 5.20]. This detail will not concern us since, for all uniformly convex spaces that we actually encounter, the apparently stronger condition is not any more difficult to prove than the weaker one. On the other hand, our main application of uniform convexity, Theorem 1.8 below, only uses the weaker condition.

Exercise 1.6. For vectors $x = (x_1, \ldots, x_n)$ in $\mathbb{R}^n$, consider the norms

$$|x|_p := \left(\sum_{j=1}^n x_j^p\right)^{1/p} \quad \text{for } 1 \leqslant p < \infty, \qquad |x|_\infty := \max\{|x_1|, \ldots, |x_n|\}.$$

(a) Show (by drawing pictures of the unit ball) that $(\mathbb{R}^n, |\cdot|_1)$ and $(\mathbb{R}^n, |\cdot|_\infty)$ are not strictly convex.
(b) Show that the spaces of real-valued functions of class L^1 or L^∞ on any measure space are not strictly convex. (Note that this implies part (a) if you take the measure space to be $\{1, \ldots, n\}$ with the counting measure.)

We will see in §2.3 that all L^p-spaces for $1 < p < \infty$ are uniformly convex; this of course includes the examples $(\mathbb{R}^n, |\cdot|_p)$ defined in Exercise 1.6. Notice that uniform convexity is not a property of the *equivalence class* of a norm but rather of the norm itself—indeed, all norms on $\mathbb{R}^n$ are equivalent, but some are uniformly convex and some are not.

Proposition 1.7. *Every inner product space $(X, \langle\cdot,\cdot\rangle)$ is uniformly convex.*

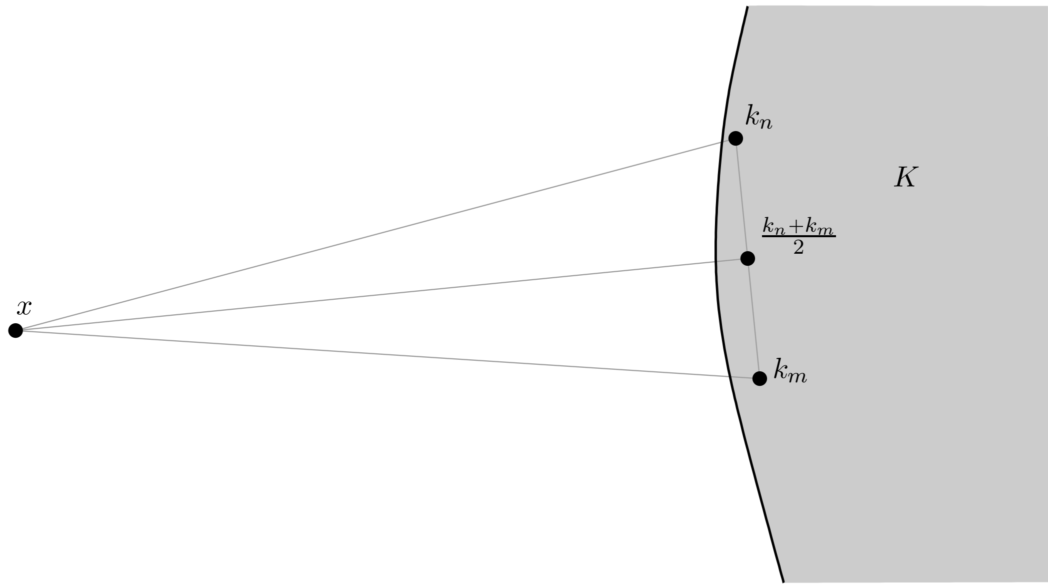

FIGURE 3. The geometric setup behind the proof of Theorem 1.8.

Proof. Denoting the norm by $\|\cdot\| := \sqrt{\langle\cdot,\cdot\rangle}$, a straightforward computation yields the **parallelogram identity**,

$$\|v+w\|^2 + \|v-w\|^2 = 2\|v\|^2 + 2\|w\|^2 \qquad \forall v, w \in X, \tag{1.2}$$

which for $\|v\| \leqslant 1$ and $\|w\| \leqslant 1$ implies the relation

$$\frac{1}{4}\|v-w\|^2 \leqslant 1 - \left\|\frac{v+w}{2}\right\|^2.$$

This gives a concrete upper bound on $\|v-w\|$ in terms of the distance from $\frac{v+w}{2}$ to the boundary of the unit ball. ☐

The following theorem on uniformly convex Banach spaces is a useful source of existence results, and will play a key role in characterzing the duals of Hilbert spaces and L^p-spaces.

Theorem 1.8. *Assume $(X, \|\cdot\|)$ is a uniformly convex Banach space, $K \subset X$ is a closed convex subset and $x \in X \backslash K$. Then the function $K \to (0, \infty) : k \mapsto \|k-x\|$ attains a unique global minimum.*

If $\dim X < \infty$, then Theorem 1.8 follows easily from the fact that since closed and bounded subsets of X are compact, $K \to (0,\infty) : k \mapsto \|k-x\|$ is a proper function: one only has to take a sequence $k_n \in K$ with $\|k_n - x\| \to \inf\{\|k-x\| \mid k \in K\}$ and use compactness to extract a convergent subsequence, whose limit is the desired minimum. This argument completely falls apart if $\dim X = \infty$, because closed bounded subsets are no longer compact. One must instead appeal to the completeness of X, using the idea represented in Figure 3: suppose $k_n, k_m \in K$ both have distances to x that are close to the infimum. After rescaling the whole picture, we can assume without loss of generality that $k_n - x$ and $k_m - x$ are both in the unit ball, in which case so is the midpoint $\frac{(k_n-x)+(k_m-x)}{2} = \frac{k_n+k_m}{2} - x$, where $\frac{k_n+k_m}{2}$ also lies in K since K is convex. By assumption, $\left\|\frac{k_n+k_m}{2} - x\right\|$ cannot be that much smaller than $\|k_n - x\|$ and $\|k_m - x\|$, since both of the latter were already close to the infimum, hence $\frac{k_n+k_m}{2} - x$ cannot be too far away from the boundary of the unit ball. But in that case, uniform convexity implies that $k_n - x$ and $k_m - x$ must be close, or equivalently, k_n and k_m must be close. We will use a version of this argument in the following to show that k_n is a Cauchy sequence, and thus converges to an element that attains the minimum.

Proof of Theorem 1.8. Let $I := \inf\{\|k - x\| \mid k \in K\}$, choose a sequence $k_n \in K$ with $I_n := \|k_n - x\| \to I$, and let

$$z_n := \frac{k_n - x}{I_n} = \frac{k_n - x}{I} + \frac{I - I_n}{I} z_n,$$

which defines a sequence in the unit sphere of X. If $\epsilon > 0$ is given, we can choose $N \in \mathbb{N}$ such that $I_n < I + \epsilon$ for all $n \geqslant N$. For any $m, n \geqslant N$, the fact that K is convex implies $\frac{k_m + k_n}{2} \in K$, thus it satisfies

$$\left\| \frac{(k_m - x) + (k_n - x)}{2} \right\| = \left\| \frac{k_m + k_n}{2} - x \right\| \geqslant I,$$

which implies

$$\left\| \frac{z_m + z_n}{2} \right\| = \left\| \frac{1}{I} \frac{(k_m - x) + (k_n - x)}{2} + \frac{I - I_m}{2I} z_m + \frac{I - I_n}{2I} z_n \right\| \geqslant 1 - \frac{\epsilon}{I}.$$

Since the latter can be made arbitrarily close to 1 by choosing $\epsilon > 0$ small, uniform convexity now implies that $\|z_m - z_n\|$ can be assumed arbitrarily small for N large, so z_n is a Cauchy sequence and therefore converges to some $z_\infty \in X$. It follows that k_n converges to $k_\infty := x + I z_\infty$, and since K is a closed set, $k_\infty \in K$. Clearly $\|x - k_\infty\| = I$.

The uniqueness of the minimum follows almost immediately since, if $k_0, k_1 \in K$ are two minimums, then the argument above shows that $k_0, k_1, k_0, k_1, \ldots$ is a Cauchy sequence, implying $k_0 = k_1$. ☐

1.2. **Orthogonal complements in Hilbert space.** Our first concrete application of uniform convexity is to prove a fundamental geometric fact about Hilbert spaces. Assume in this section that $(\mathcal{H}, \langle\ ,\ \rangle)$ is a Hilbert space over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, and denote its norm by $\|\cdot\| := \sqrt{\langle \cdot, \cdot \rangle}$.

Given a linear subspace $V \subset \mathcal{H}$, the **orthogonal complement** of V is defined as

$$V^\perp := \{x \in \mathcal{H} \mid \langle x, v \rangle = 0 \text{ for all } v \in V\}.$$

Theorem 1.9. *If V is a closed linear subspace of the Hilbert space $\mathcal{H}$, then every $x \in \mathcal{H}$ can be written as $v + w$ for unique elements $v \in V$ and $w \in V^\perp$; symbolically, we write*

$$\mathcal{H} = V \oplus V^\perp.$$

Theorem 1.9 is a classic example of a result that is very familiar in finite dimensions and sounds obvious, but is actually quite nontrivial in the general case. In particular, it depends in essential ways on the completeness of $\mathcal{H}$ and the assumption that $V \subset \mathcal{H}$ is closed. To see the latter, recall that while many functions $f : \mathbb{R}^n \to V$ of class L^2 are not continuous, the space of continuous functions of class L^2 is dense in the Hilbert space $L^2(\mathbb{R}^n)$; we will review this fact in §5. One can therefore use the Cauchy-Schwarz inequality to argue that if g is any function L^2-orthogonal to every continuous function in $L^2(\mathbb{R}^n)$, then g is in fact orthogonal to *everything* in $L^2(\mathbb{R}^n)$, implying $g = 0$. In other words, $C^0(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ is a proper subspace of $L^2(\mathbb{R}^n)$ whose orthogonal complement is the *trivial* subspace, thus not every L^2-function can be written as the sum of one that is continuous plus one that is orthogonal to the continuous functions. Viewing $C^0(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ itself as an inner product space with the L^2-inner product, one can also find closed proper subspaces of $C^0(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ whose orthogonal complements are trivial, showing that the completeness of $\mathcal{H}$ is also an indispensable assumption.

Proof of Theorem 1.9. The uniqueness of the decomposition $x = v + w$ with $v \in V$ and $w \in V^\perp$ is immediate from the nondegeneracy of the inner product: if it were not unique, then two distinct decompositions $x = v + w = v' + w'$ would give rise to a nontrivial vector $v - v' = w' - w \in V \cap V^\perp$, which is impossible since every nonzero $y \in V$ satisfies $\langle y, y \rangle > 0$.

For existence, observe that there is nothing to prove if $x \in V$, so assume $x \in \mathcal{H} \backslash V$. Since $\mathcal{H}$ is a complete inner product space, Proposition 1.7 implies that it is also a uniformly convex Banach space; moreover, the subspace $V \subset \mathcal{H}$ is a convex set that is closed by assumption. Theorem 1.8 thus implies the existence of an element $v \in V$ that is nearest to x, and we claim that $w := x - v$ then lies in $V^\perp$. Indeed, for any $h \in V$, the fact that $\|x - v\|^2 = \langle x - v, x - v\rangle$ minimizes the distance from x to V implies

$$0 = \frac{d}{dt}\|x - (v + th)\|^2\Big|_{t=0} = \frac{d}{dt}\langle x - (v + th), x - (v + th)\rangle\Big|_{t=0} = \frac{d}{dt}\langle w - th, w - th\rangle\Big|_{t=0}$$
$$= \frac{d}{dt}\left(\|w\|^2 - 2t\,\mathrm{Re}\langle w, h\rangle + t^2\|h\|^2\right)\Big|_{t=0} = -2\,\mathrm{Re}\langle w, h\rangle,$$

where the symbol "Re" is redundant in the case $\mathbb{K} = \mathbb{R}$, and the result is then simply $\langle w, h\rangle = 0$. In the complex case, we can plug in $ih \in V$ instead of h, so that the same computation also gives

$$0 = -2\,\mathrm{Re}\langle w, ih\rangle = -2\,\mathrm{Re}\,(i\langle w, h\rangle) = 2\,\mathrm{Im}\langle w, h\rangle,$$

and the conclusion is again $\langle w, h\rangle = 0$ for all $h \in V$, as claimed. ☐

Recall that the **dual space** $\mathcal{H}^*$ of $\mathcal{H}$ is the space of all bounded linear functionals $\Lambda : \mathcal{H} \to \mathbb{K}$, endowed with the operator norm

$$\|\Lambda\| := \sup_{v \in \mathcal{H}\setminus\{0\}} \frac{|\Lambda(v)|}{\|v\|}.$$

The Cauchy-Schwarz inequality $|\langle v, w\rangle| \leqslant \|v\| \cdot \|w\|$ implies that every $v \in \mathcal{H}$ gives rise to a bounded linear functional $\Lambda_v : \mathcal{H} \to \mathbb{K}$ defined by $\Lambda_v(x) := \langle v, x\rangle$, which satisfies $\|\Lambda_v\| = \|v\|$ since the maximum of $\frac{\Lambda_v(x)}{\|x\|}$ is attained by $x := v$. The following is one of at least three results that are often called the *Riesz representation theorem*, all of which give concrete characterizations of the dual spaces of certain classes of Banach spaces. Its content in the present setting is that all bounded linear functionals on $\mathcal{H}$ are of the type described above.

Theorem 1.10 (Riesz representation theorem for Hilbert spaces)**.** *The real-linear*[3] *map $\mathcal{H} \to \mathcal{H}^* : v \mapsto \Lambda_v := \langle v, \cdot\rangle$ is a bijection.*

Proof. The injectivity of the map $\mathcal{H} \to \mathcal{H}^* : v \mapsto \Lambda_v = \langle v, \ \rangle$ is clear since $\Lambda_v(v) = \|v\|^2 > 0$ for all $v \neq 0$. The main step is thus to prove surjectivity, i.e. given any $\Lambda \in \mathcal{H}^*$, we need to find $v \in \mathcal{H}$ such that $\langle v, \ \rangle = \Lambda$. The idea is to look for v in the orthogonal complement of the subspace

$$K := \ker \Lambda \subset \mathcal{H}.$$

The latter is a closed subspace since, by the continuity of Λ, any convergent sequence $x_n \to x$ in $\mathcal{H}$ with $\Lambda(x_n) = 0$ for all n implies $\Lambda(x) = 0$. Since the problem is trivial if $\Lambda = 0$, suppose there exists $x \in \mathcal{H}$ with $\Lambda(x) \neq 0$, and after multiplication with a scalar, assume without loss of generality $\Lambda(x) = 1$. By Theorem 1.9, we can write $x = k + w$ for unique elements $k \in K$ and $w \in K^\perp$, which satisfy $\Lambda(w) = \Lambda(k) + \Lambda(w) = \Lambda(k + w) = \Lambda(x) = 1$. We claim that

$$v := \frac{w}{\|w\|^2} \in \mathcal{H}$$

is the element we are looking for. Indeed, $\langle v, k\rangle = 0 = \Lambda(k) = 0$ for all $k \in K$, and $\langle v, w\rangle = 1 = \Lambda(w)$, so the result now follows from the purely algebraic observation that K is a subspace of codimension 1 which does not contain w, implying that every $x \in \mathcal{H}$ can be written uniquely as $cw + k$ for some $c \in \mathbb{K}$ and $k \in K$. ☐

[3]Due to the conventions of complex inner product spaces, the map $\mathcal{H} \to \mathcal{H}^* : v \mapsto \Lambda_v$ in the case $\mathbb{K} = \mathbb{C}$ is not complex linear, but is instead complex *antilinear*.

2. Duality in L^p-spaces

2.1. **The pairing of L^p and L^q.** For this section, assume (X, μ) is an arbitrary measure space, and $(V, \langle\ ,\ \rangle)$ is (as in §0) a finite-dimensional inner product space over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ with norm $|\cdot| := \sqrt{\langle\cdot,\cdot\rangle}$. Our aim is to prove a characterization of the space $(L^p(X))^*$ of bounded linear functionals $L^p(X) \to \mathbb{K}$ which, like Theorem 1.10, is also sometimes called the *Riesz representation theorem.* To prepare the statement, notice that whenever $1 \leqslant p, q \leqslant \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, Hölder's inequality gives rise to a real-linear map

$$L^q(X) \to (L^p(X))^* : g \mapsto \Lambda_g := \int_X \langle g, \cdot\rangle \, d\mu \tag{2.1}$$

satisfying $\|\Lambda_g\|_{(L^p)^*} \leqslant \|g\|_{L^q}$, where $\|\cdot\|_{(L^p)^*}$ denotes the operator norm on bounded linear operators $L^p(X) \to \mathbb{K}$.

Lemma 2.1. *Assume $1 \leqslant p, q \leqslant \infty$, $\frac{1}{p} + \frac{1}{q} = 1$ and, additionally, either $p < \infty$ or X is σ-finite.*[4] *Then for every $f \in L^p(X)$,*

$$\sup_{g \in L^q(X) \backslash \{0\}} \frac{\left|\int_X \langle g, f\rangle \, d\mu\right|}{\|g\|_{L^q}} = \|f\|_{L^p},$$

and the ratio on the left hand side attains its maximum in the case $p < \infty$.

Proof. Hölder's inequality implies that the ratio in question can never be greater than $\|f\|_{L^p}$. There is nothing to prove if $f = 0$, so assume $f \in L^p(X)$ is nontrivial. If $p < \infty$, we define $g : X \to V$ by $g := |f|^{p-2} f$ at points where $f \neq 0$ and $g := 0$ otherwise. Then g satisfies $|g| = 1$ almost everywhere if $p = 1$, and in the other cases, $|g|^q = |f|^{q(p-1)} = |f|^p$, thus $g \in L^q(X)$ and

$$\int_X \langle g, f\rangle \, d\mu = \int_X |f|^p \, d\mu = \|f\|_{L^p}^p = \left(\|f\|_{L^p}^p\right)^{\frac{p-1}{p}} \cdot \|f\|_{L^p} = \left(\|g\|_{L^q}^q\right)^{1-\frac{1}{p}} \cdot \|f\|_{L^p} = \|g\|_{L^q} \cdot \|f\|_{L^p},$$

so this choice of $g \in L^q(X)$ maximizes the ratio in question.

In the case $p = \infty$ and $q = 1$, we argue by contradiction and suppose that $\|f\|_{L^\infty}$ is strictly greater than the supremum of $\left|\int_X \langle g, f\rangle \, d\mu\right| / \|g\|_{L^1}$ over all $g \in L^1(X) \backslash \{0\}$. Then there exists a constant c strictly greater than this supremum such that the set $A' := \left\{x \in X \;\middle|\; |f(x)| \geqslant c\right\}$ has positive measure. Assuming X is σ-finite, there also exists a subset $A \subset A'$ with $0 < \mu(A) < \infty$, and the function g defined as $f/|f|$ on A and 0 everywhere else is then in $L^1(X)$, with $\|g\|_{L^1} = \mu(A)$. Since $|f| \geqslant c > \left|\int_X \langle g, f\rangle \, d\mu\right| / \|g\|_{L^1}$ on A, we now find the contradiction,

$$\left|\int_X \langle g, f\rangle \, d\mu\right| = \int_A |f| \, d\mu \geqslant \mu(A) \cdot c = \|g\|_{L^1} \cdot c > \left|\int_X \langle g, f\rangle \, d\mu\right|.$$

□

Corollary 2.2. *For every $1 \leqslant p, q \leqslant \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, if either $p > 1$ or X is σ-finite, then the bounded real-linear map* (2.1) *is injective and satisfies $\|\Lambda_g\|_{(L^p)^*} = \|g\|_{L^q}$ for all $g \in L^q(X)$.* □

Exercise 2.3. Show that for any $f \in L^\infty(X)$ satisfying $|f| < \|f\|_{L^\infty}$ almost everywhere, the inequality $\left|\int_X \langle g, f\rangle \, d\mu\right| \leqslant \|g\|_{L^1} \cdot \|f\|_{L^\infty}$ is strict for every $g \in L^1(X) \backslash \{0\}$.

Here is the hard part:

Theorem 2.4 (Riesz representation theorem for L^p). *The map* (2.1) *is bijective for all $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, and also for $p = 1$ and $q = \infty$ if X is σ-finite.*

[4]Certain measure-theoretic pathologies can arise in the case $p = \infty$ that are excluded if we assume X is σ-finite. This is not the most general assumption possible, but it suffices for all applications we will want to consider. For more general versions of the results in this section involving duality between $L^1(X)$ and $L^\infty(X)$, see [Sal16, §4.5].

Remark 2.5. In the case $\mathbb{K} = \mathbb{C}$, the map $L^q(X) \to (L^p(X))^*$ in (2.1) is complex antilinear and thus is not, strictly speaking, an isomorphism of complex Banach spaces. However, one can also define a space $(L^p(X))'$ consisting of all bounded complex-antilinear functionals $\Lambda : L^p(X) \to \mathbb{C}$ and consider a complex-linear map defined by

$$L^q(X) \to (L^p(X))' : g \mapsto \Lambda'_g := \int_X \langle \cdot, g \rangle \, d\mu. \tag{2.2}$$

It is an easy exercise to check that this map is bijective whenever (2.1) is, so under the same hypotheses as Theorem 2.4, it is a complex Banach space isomorphism.

The proof of Theorem 2.4 given below follows the same strategy as our proof of the corresponding statement about Hilbert spaces in Theorem 1.10. The crucial idea in the latter was that given a nontrivial dual vector $\Lambda \in \mathcal{H}^*$ for a Hilbert space $\mathcal{H}$, the right place to search for elements x with $\Lambda = \langle x, \cdot \rangle$ is in the *orthogonal complement* of the closed hyperplane $\ker \Lambda \subset \mathcal{H}$. While the notion of orthogonality does not make sense in $L^p(X)$ for $p \neq 2$, Hölder's inequality furnishes us with a reasonable substitute in the form of the natural pairing of L^p with L^q for $\frac{1}{p} + \frac{1}{q} = 1$; informally, we can thus regard the orthogonal complement of a subspace in $L^p(X)$ as a subspace of $L^q(X)$. With this notion in mind, the main task is then to prove, as we did for Hilbert spaces in §1.2, that a proper closed subspace $K \subset L^p(X)$ always has a *nontrivial* orthogonal complement. Our proof of this in the Hilbert space setting required two fundamental ingredients:

(1) The uniform convexity of every Hilbert space $\mathcal{H}$;
(2) The differentiability of the function $t \mapsto \|x + tv\|^2$ for any $x, v \in \mathcal{H}$.

Both were easy to prove using the characterization of the Hilbert space norm via an inner product, but since the latter is not available in $L^p(X)$ for $p \neq 2$, we will have to work a bit harder.

Recall that every Banach space $(E, \|\cdot\|)$ has a canonical continuous inclusion into the dual of its dual space, defined by

$$\Phi : E \to E^{**}, \qquad \Phi(v)\Lambda := \Lambda(v) \text{ for } v \in E,\ \Lambda \in E^*.$$

The injectivity of this map for general Banach spaces is not so obvious, though for $E = L^p(X)$ with $p < \infty$, it is an easy consequence of the following corollary of Lemma 2.1. Outside of these special cases, it follows immediately from the Hahn-Banach theorem (see [RS80, §III.3]), whose standard proof uses the axiom of choice.

Lemma 2.6. *For every normed vector space $(E, \|\cdot\|)$ and every $x \in E$, there exists a dual vector $\Lambda \in E^*$ with $\|\Lambda\| = 1$ and $\Lambda(x) = \|x\|$.*

Proof for $E = L^p(X)$ with $p < \infty$. Given $f \in L^p(X)$, choose $\Lambda := \Lambda_g \in (L^p(X))^*$ for $g \in L^q(X)$ as in Lemma 2.1, then normalize g. □

Corollary 2.7. *For every Banach space $(E, \|\cdot\|)$, the canonical map $\Phi : E \to E^{**}$ is an injective isometry, i.e. it satisfies $\|\Phi(x)\| = \|x\|$ for every $x \in E$.* □

One calls $(E, \|\cdot\|)$ **reflexive** if the inclusion $\Phi : E \hookrightarrow E^{**}$ is also surjective. For $E = L^p(X)$ with $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, Theorem 2.4 identifies E^* with $L^q(X)$ and then identifies E^{**} in turn with $L^p(X)$, so that under these identifications, $\Phi : E \to E^{**}$ becomes a map $L^p(X) \to L^p(X)$ uniquely determined by[5]

$$\int_X \langle \Phi(f), g \rangle \, d\mu = \int_X \langle f, g \rangle \, d\mu \quad \text{for all} \quad g \in L^q(X).$$

[5]One needs to be a bit careful with this argumentation in the case $\mathbb{K} = \mathbb{C}$, because the bijection $E^* \cong L^q(X)$ is then complex antilinear rather than linear, so substituting $L^q(X)$ for E^* identifies E^{**} with the space $(L^q(X))'$ of bounded complex-*antilinear* maps $L^q(X) \to \mathbb{C}$ instead of the actual dual space of $L^q(X)$. As mentioned in Remark 2.5, however, the Riesz representation identifies the latter complex-linearly with $L^p(X)$.

This implies that $\int_X \langle \Phi(f) - f, g\rangle \, d\mu$ vanishes for all $g \in L^q(X)$, proving that the function $\Phi(f) - f \in L^p(X)$ is identified with the trivial element of $(L^q(X))^*$, which makes $\Phi : L^p(X) \to L^p(X)$ the identity map.

Corollary 2.8. *For $1 < p < \infty$, $L^p(X)$ is reflexive.* □

Remark 2.9. Reflexivity is in fact a general property of uniformly convex Banach spaces, by the *Milman-Pettis theorem*; see e.g. [RS80, Problem V.15].

Theorem 2.4 is false for $p = \infty$ and $q = 1$; the dual of $L^\infty(X)$ is generally a larger space than can be described via such a pairing. One can see this by comparing Lemma 2.6 with Exercise 2.3: there exist nontrivial functions $f \in L^\infty(X)$ for which an element $\Lambda \in (L^\infty(X))^*$ with $\|\Lambda\|_{(L^\infty)^*} = 1$ satisfying $|\Lambda(f)| = \|f\|_{L^\infty}$ must exist, but the strictness of the inequality in Exercise 2.3 implies that Λ cannot be represented by any function in $L^1(X)$.[6] For more counterexamples, see also [Rud87, Chapter 6, Exercise 13] or [Sal16, Example 4.36]. It follows that $L^1(X)$ is not reflexive, and by the next exercise, neither is $L^\infty(X)$.

Exercise 2.10. For a Banach space E, let $\Phi_E : E \hookrightarrow E^{**}$ and $\Phi_{E^*} : E^* \hookrightarrow E^{***}$ denote the canonical inclusions, and denote by $\Phi_E^* : E^{***} \to E^*$ the transpose of Φ_E.

(a) Show that $\Phi_E^* \circ \Phi_{E^*}$ is the identity map on E^*.
(b) Show that the image of Φ_E is always a closed subspace of E^{**}.
(c) Deduce that E^* is reflexive if and only if E is reflexive.
Hint: Another easy consequence of the Hahn-Banach theorem is that if $A : X \to Y$ is a bounded linear operator between Banach spaces such that $\operatorname{im} A \subset Y$ is closed and $A^* : Y^* \to X^*$ is injective, then A is surjective.

2.2. **Differentiability of the norm.** Let us examine whether the function $\|f + tg\|_{L^p}^p$ can be differentiated with respect to $t \in \mathbb{R}$ for $f, g \in L^p(X)$. Assume in the following

$$1 < p < \infty.$$

For $v, w \in V$ and $t \in \mathbb{R}$ with $v + tw \neq 0$, the differentiability of the function $x \mapsto x^{p/2}$ for $x \neq 0$ implies

$$\begin{aligned} \frac{d}{dt}|v + tw|^p &= \frac{d}{dt}\langle v + tw, v + tw\rangle^{p/2} \\ &= \frac{p}{2}\langle v + tw, v + tw\rangle^{\frac{p}{2}-1} \cdot \frac{d}{dt}\left(|v|^2 + 2t \operatorname{Re}\langle v, w\rangle + t^2|w|^2\right) \\ &= p|v + tw|^{p-2}\left(\operatorname{Re}\langle v, w\rangle + t|w|^2\right) = p|v + tw|^{p-2} \cdot \operatorname{Re}\langle v + tw, w\rangle. \end{aligned} \tag{2.3}$$

Notice that by the Cauchy-Schwarz inequality on $(V, \langle\ ,\ \rangle)$, the right hand side of this expression satisfies

$$\left| p|v + tw|^{p-2} \cdot \operatorname{Re}\langle v + tw, w\rangle \right| \leqslant p|v + tw|^{p-1} \cdot |w|,$$

whenever $v + tw \neq 0$. Since $p > 1$, one can therefore sensibly define the right hand side of (2.3) to be 0 when $v + tw = 0$, and the relation remains correct since in this case

$$\frac{d}{dt}|v + tw|^p = \frac{d}{ds}|(v + tw) + sw|^p\Big|_{s=0} = \frac{d}{ds}|s|^p|w|^p\Big|_{s=0} = |w|^p \lim_{s\to 0}\frac{|s|^p}{s} = 0.$$

[6] Quoting Lemma 2.6 for $L^\infty(X)$ means we are relying on the Hahn-Banach theorem, which is inherently non-constructive, i.e. it guarantees the existence of an element in $(L^\infty(X))^* \setminus L^1(X)$ as an artefact of the axioms of set theory, but gives no hint how one could ever write one down. In fact, *all* proofs that $(L^\infty(X))^* \setminus L^1(X) \neq \varnothing$ are non-constructive in this sense. Readers who wish to explore this particular set-theoretic rabbit hole may consult [Sch99, Chapter 14]; see also `https://mathoverflow.net/questions/5351/whats-an-example-of-a-space-that-needs-the-hahn-banach-theorem`.

With this understood, for any given $f, g \in L^p(X)$, differentiation under the integral sign now suggests the formula

$$\begin{aligned} \frac{d}{dt}\|f+tg\|_{L^p}^p &= \frac{d}{dt}\int_X |f(x)+tg(x)|^p \, d\mu(x) = \int_X \frac{d}{dt}|f(x)+tg(x)|^p \, d\mu(x) \\ &= \int_X p|f(x)+tg(x)|^{p-2} \cdot \operatorname{Re}\langle f(x)+tg(x), g(x)\rangle \, d\mu(x), \end{aligned} \tag{2.4}$$

where the same application of the Cauchy-Schwarz inequality interprets the integrand on the right as 0 whenever $f(x) + tg(x) = 0$. Let us use Theorem 0.4 to justify this formula at $t = 0$. Set $(Y, \nu) := (X, \mu)$ and $M := (-1, 1) \subset \mathbb{R}$ and define $\varphi : (-1, 1) \times X \to V$ by $\varphi(t, x) := |f(x) + tg(x)|^p$, so $\frac{\partial \varphi}{\partial t}(t, x)$ is given by the integrand on the right hand side of (2.4). Both φ and $\frac{\partial \varphi}{\partial t}$ are then continuous functions of $t \in (-1, 1)$ for every fixed $x \in X$. For every fixed $t \in (-1, 1)$, they also satisfy

$$|\varphi(t,x)| \leqslant (|f(x)| + |g(x)|)^p \tag{2.5}$$

and

$$\left|\frac{\partial \varphi}{\partial t}(t,x)\right| \leqslant p\,(|f(x)| + |g(x)|)^{p-1} \cdot |g(x)|. \tag{2.6}$$

By Minkowski's inequality,

$$\int_X (|f(x)| + |g(x)|)^p \; d\mu(x) = \big\||f| + |g|\big\|_{L^p}^p \leqslant (\|f\|_{L^p} + \|g\|_{L^p})^p < \infty,$$

thus the right hand side of (2.5) defines a μ-integrable function on X. It follows in turn that the function $(|f| + |g|)^{p-1}$ is of class $L^{p/(p-1)}$ on X, and since $\frac{p-1}{p} + \frac{1}{p} = 1$, Hölder's inequality implies that the right hand side of (2.6) is also μ-integrable. The hypotheses of Theorem 0.4 are thus satisfied, and we conclude:

Lemma 2.11. *For any $f, g \in L^p(X)$ with $1 < p < \infty$, the function $\mathbb{R} \to [0, \infty) : t \mapsto \|f + tg\|_{L^p}^p$ is differentiable and satisfies*

$$\left.\frac{d}{dt}\|f+tg\|_{L^p}^p\right|_{t=0} = p\int_X |f|^{p-2} \cdot \operatorname{Re}\langle f, g\rangle \, d\mu.$$

□

2.3. **Uniform convexity of L^p.** In order to prove that $L^p(X)$ is uniformly convex for $1 < p < \infty$, we begin with the observation that the function

$$V \to \mathbb{R} : v \mapsto |v|^p$$

is strictly convex for all $p \in (1, \infty)$. One can show this by computing that its Hessian is positive definite everywhere outside of the origin; at the origin it may fail to have second derivatives, but it is then easy enough to check the convexity condition along segments connecting 0 to any other point. It follows that the function

$$\psi : V \times V \to \mathbb{R} : (v, w) \mapsto \frac{|v|^p + |w|^p}{2} - \left|\frac{v+w}{2}\right|^p \tag{2.7}$$

is nonnegative everywhere, and strictly positive whenever $v \neq w$. For any constant $\epsilon > 0$, its restriction to the compact subset

$$K := \left\{(v, w) \in V \times V \;\middle|\; |v - w|^p \geqslant \epsilon \text{ and } |v|^p + |w|^p \leqslant 1\right\}$$

therefore satisfies $\psi|_K \geqslant \delta$ for some constant $\delta > 0$.[7] Now if $v, w \in V$ are any elements with $v \neq w$, set $\tau := (|v|^p + |w|^p)^{1/p} > 0$, $v' := v/\tau$ and $w' := w/\tau$, so $|v'|^p + |w'|^p = 1$, and the condition $|v' - w'|^p \geqslant \epsilon$ is equivalent to $|v - w|^p \geqslant \epsilon\tau^p$. Under this condition, $\psi(v', w') \geqslant \delta$ becomes $\psi(v, w) \geqslant \delta\tau^p$, which proves:

Lemma 2.12. *Given any $p \in (1, \infty)$ and $\epsilon > 0$, there exists $\delta > 0$ such that the function ψ in* (2.7) *satisfies*

$$|v - w|^p \geqslant \epsilon\,(|v|^p + |w|^p) \quad \Rightarrow \quad \psi(v, w) \geqslant \delta\,(|v|^p + |w|^p) \qquad \forall v, w \in V.$$

□

Exercise 2.13. Extract from Lemma 2.12 a new proof that $(V, \langle\ ,\ \rangle)$ is uniformly convex.

The uniform convexity of $L^p(X)$ is an easy application of the following estimate.

Theorem 2.14. *Given any $p \in (1, \infty)$ and $\epsilon > 0$, there exists $\delta > 0$ such that for all $f, g \in L^p(X)$,*

$$\frac{\|f\|_{L^p}^p + \|g\|_{L^p}^p}{2} - \left\|\frac{f+g}{2}\right\|_{L^p}^p \geqslant \delta\left[\|f - g\|_{L^p}^p - \epsilon\left(\|f\|_{L^p}^p + \|g\|_{L^p}^p\right)\right].$$

Proof. Given $f, g \in L^p(X)$ and $\epsilon > 0$, decompose X into the subsets

$$A := \left\{x \in X \;\middle|\; |f(x) - g(x)|^p \geqslant \epsilon\,(|f(x)|^p + |g(x)|^p)\right\}, \qquad A^c = X \backslash A.$$

For $x \in A$, we have $\psi(f(x), g(x)) \geqslant \delta_0\,(|f(x)|^p + |g(x)|^p)$ for some constant $\delta_0 > 0$ provided by Lemma 2.12. Now using the fact that $|f - g|^p < \epsilon\,(|f|^p + |g|^p)$ on A^c, while $\psi(f, g) \geqslant 0$ and $\left|\frac{f-g}{2}\right|^p = \left|\frac{f+(-g)}{2}\right|^p \leqslant \frac{|f|^p + |g|^p}{2}$ hold everywhere, we estimate

$$\begin{aligned}
\frac{\|f\|_{L^p}^p + \|g\|_{L^p}^p}{2} - \left\|\frac{f+g}{2}\right\|_{L^p}^p &\geqslant \int_A \psi(f, g)\,d\mu \geqslant \delta_0 \int_A (|f|^p + |g|^p)\,d\mu \geqslant \frac{\delta_0}{2^{p-1}} \int_A |f - g|^p\,d\mu \\
&= \frac{\delta_0}{2^{p-1}}\left(\|f - g\|_{L^p}^p - \int_{A^c} |f - g|^p\,d\mu\right) \\
&\geqslant \frac{\delta_0}{2^{p-1}}\left(\|f - g\|_{L^p}^p - \epsilon\int_{A^c} (|f|^p + |g|^p)\,d\mu\right) \\
&\geqslant \frac{\delta_0}{2^{p-1}}\left(\|f - g\|_{L^p}^p - \epsilon\left(\|f\|_{L^p}^p + \|g\|_{L^p}^p\right)\right).
\end{aligned}$$

Set $\delta := \delta_0/2^{p-1}$. □

Corollary 2.15. *For $1 < p < \infty$, $L^p(X)$ is uniformly convex.* □

Remark 2.16. The notion of uniform convexity and Corollary 2.15 are originally due to Clarkson [Cla36], and the literature contains many other proofs based on more powerful inequalities than in Theorem 2.14; see for instance [LL01, §2.5], which uses *Hanner's inequality.* Our proof has been adapted from [Shi18].

[7]Recall from Remark 0.3 that we are assuming $\dim V < \infty$, and we are using that assumption here in order to say that K is compact. However, if V is an infinite-dimensional Hilbert space, then one can fix an orthonormal basis, single out two basis vectors $e_1, e_2 \in V$ and then argue as follows: if $(v_n, w_n) \in K$ is a sequence such that $\psi(v_n, w_n) \to 0$, then by choosing suitable new orthonormal bases for each n, we can transform each (v_n, w_n) by isometries of $(V, \langle\ ,\ \rangle)$ (which leave both K and ψ invariant) so that without loss of generality, each v_n and w_n lies in the span of e_1 and e_2. It follows now that the sequence (v_n, w_n) lives in a compact subset of V, so a subsequence converges to some $(v, w) \in K$ with $\psi(v, w) = 0$, which cannot exist. The estimate $\psi|_K \geqslant \delta > 0$ therefore also holds in this case.

2.4. **Proof of the representation theorem.** As in the Hilbert space case, the idea for finding a function $g \in L^q(X)$ to represent any given $\Lambda \in (L^p(X))^*$ is to look for nontrivial functions whose pairing with $L^p(X)$ annihilates $\ker \Lambda$. We do this by finding the closest point in $\ker \Lambda$ to some $h \in L^p(X) \backslash \ker \Lambda$.

Proof of Theorem 2.4 for $1 < p < \infty$. Assume $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Given $\Lambda \in (L^p(X))^*$, we need to find $g \in L^q(X)$ such that $\int_X \langle g, f \rangle \, d\mu = \Lambda(f)$ for all $f \in L^p(X)$. Assume $\Lambda \neq 0$ since the problem is otherwise trivial, let $K := \ker \Lambda \subset L^p(X)$ and choose $h \in L^p(X) \backslash K$; after multiplication by a scalar, we may assume $\Lambda(h) = 1$. Then K is a closed convex subset, and since $L^p(X)$ is uniformly convex, Theorem 1.8 provides an element $k_0 \in K$ minimizing the distance to h. For any $k \in K$, Lemma 2.11 then gives

$$0 = \frac{d}{dt} \| h - (k_0 - tk) \|_{L^p}^p \bigg|_{t=0} = p \int_X |h - k_0|^{p-2} \cdot \operatorname{Re} \langle h - k_0, k \rangle \, d\mu,$$

where the integral on the right hand side is well defined due to Hölder's inequality. The symbol "Re" in this formula is redundant in the case $\mathbb{K} = \mathbb{R}$, while if $\mathbb{K} = \mathbb{C}$, replacing $k \in K$ with $ik \in K$ in this relation shows that the same thing holds with the imaginary part instead of the real part, implying that the function $\tilde{g} := |h - k_0|^{p-2}(h - k_0)$ satisfies

$$\int_X \langle \tilde{g}, k \rangle \, d\mu = 0 \quad \text{ for all } \quad k \in K.$$

Observe that since $h - k_0 \in L^p(X)$ and $|\tilde{g}| \leqslant |h - k_0|^{p-1}$, $\tilde{g} \in L^q(X)$. Now let $g := c\tilde{g} \in L^q(X)$ for a constant $c > 0$ to be determined momentarily. The relation above implies $\int_X \langle g, f \rangle \, d\mu = \Lambda(f)$ holds for all $f \in K$, and moreover,

$$\int_X \langle g, h - k_0 \rangle \, d\mu = c \int_X |h - k_0|^{p-2} \langle h - k_0, h - k_0 \rangle \, d\mu = c \| h - k_0 \|_{L^p}^p > 0,$$

so the latter matches $\Lambda(h - k_0) = \Lambda(h) = 1$ if we set $c := 1/\|h - k_0\|_{L^p}^p$. Clearly $h - k_0 \notin K$, so $L^p(X)$ is spanned by K and $h - k_0$, thus we have proved that $\int_X \langle g, f \rangle \, d\mu = \Lambda(f)$ holds for all $f \in L^p(X)$. $\square$

The case $p = 1$ is easily derived from the case $p > 1$ if X has finite measure, and we will then use σ-finiteness to extend to the case $\mu(X) = \infty$. We will need to know that L^1-functions can be approximated by L^p-functions for $p > 1$.

Lemma 2.17. *For every* $p \in (1, \infty]$, $L^p(X) \cap L^1(X)$ *is dense in* $L^1(X)$.

Proof. Given $f \in L^1(X)$ and $n \in \mathbb{N}$, denote

$$A_n := \{ x \in X \mid |f(x)| \leqslant n \}$$

and define $f_n : X \to V$ as the product of f with the characteristic function of A_n. Since $f \in L^1(X)$ and $|f| > 1$ on $X \backslash A_1$, we have $\mu(X \backslash A_1) \leqslant \int_{X \backslash A_1} |f| \, d\mu \leqslant \int_X |f| \, d\mu < \infty$, i.e. $X \backslash A_1$ has finite measure. Clearly $|f_n| \leqslant n$ everywhere for each $n \in \mathbb{N}$, and since $|f|^p \leqslant |f|$ on A_1,

$$\|f_n\|_{L^p}^p = \int_{X \backslash A_1} |f_n|^p \, d\mu + \int_{A_1} |f|^p \, d\mu \leqslant n^p \mu(X \backslash A_1) + \int_{A_1} |f| \, d\mu \leqslant n^p \mu(X \backslash A_1) + \|f\|_{L^1} < \infty,$$

so $f_n \in L^p(X)$ for all p. Since the intersection of the sets $X \backslash A_n$ for all $n \in \mathbb{N}$ is empty, we find

$$\|f - f_n\|_{L^1} = \int_{X \backslash A_n} |f| \, d\mu \to 0 \quad \text{ as } \quad n \to \infty,$$

proving $f_n \to f$ in $L^1(X)$. $\square$

Proof of Theorem 2.4 for $p = 1$ *and* $\mu(X) < \infty$. The advantage of having finite measure is that for every $p' > p \geqslant 1$, $L^{p'}(X)$ is contained in $L^p(X)$, and the inclusion $L^{p'}(X) \hookrightarrow L^p(X)$ is a continuous linear map. This follows from Hölder's inequality, which for $r \geqslant p$ with $\frac{1}{p'} + \frac{1}{r} = \frac{1}{p}$ gives

$$\|f\|_{L^p} \leqslant \|1\|_{L^r} \cdot \|f\|_{L^{p'}} = \mu(X)^{1/r} \cdot \|f\|_{L^{p'}}.$$

Now if $\Lambda \in (L^1(X))^*$, then for $f \in L^p(X)$ with $1 < p < \infty$,

$$|\Lambda(f)| \leqslant \|\Lambda\|_{(L^1)^*} \cdot \|f\|_{L^1} \leqslant \mu(X)^{1/q} \cdot \|\Lambda\|_{(L^1)^*} \cdot \|f\|_{L^p}, \tag{2.8}$$

where $q \in (1, \infty)$ is determined by $\frac{1}{p} + \frac{1}{q} = 1$. This means Λ also belongs to $(L^p(X))^*$, so by the $p > 1$ case of Theorem 2.4, there exists a function $g_p \in L^q(X)$ such that $\Lambda(f) = \int_X \langle g_p, f\rangle \, d\mu$ for all $f \in L^p(X)$. Notice that if $p < p' < \infty$, then $g_{p'} \in L^{q'}(X)$ with $\frac{1}{p'} + \frac{1}{q'} = 1$, where $q' < q$, thus $L^{p'}(X) \subset L^p(X)$ and $L^q(X) \subset L^{q'}(X)$. It follows that g_p is also in $L^{q'}(X)$ and satisfies

$$\int_X \langle g_p - g_{p'}, f\rangle \, d\mu = \Lambda(f) - \Lambda(f) = 0 \quad \text{for all} \quad f \in L^{p'}(X),$$

$g_p - g_{p'} \in L^{q'}(X)$ defines the trivial element of $(L^{p'}(X))^*$, implying $g_p - g_{p'} = 0$ almost everywhere. For this reason we will now drop p from the notation and write g_p for every $p \in (1, \infty)$ as a single function g, which belongs to $L^q(X)$ for every $q \in (1, \infty)$. By (2.8) and Corollary 2.2, it satisfies

$$\|g\|_{L^q} = \|\Lambda\|_{(L^p)^*} \leqslant \mu(X)^{1/q} \cdot \|\Lambda\|_{(L^1)^*} \quad \text{for every} \quad q \in (1, \infty).$$

We claim that this implies $g \in L^\infty(X)$ with $\|g\|_{L^\infty} \leqslant \|\Lambda\|_{(L^1)^*}$. Indeed, for each $c > 0$, let $A_c := \{x \in X \mid |g(x)| \geqslant c\}$; then fixing $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$c\mu(A_c)^{1/q} \leqslant \|g\|_{L^q} \leqslant \mu(X)^{1/q} \cdot \|\Lambda\|_{(L^1)^*}.$$

Taking the limit $q \to \infty$ then yields $c \leqslant \|\Lambda\|_{(L^1)^*}$ unless $\mu(A_c) = 0$, thus proving the claim.

We have now found a function $g \in L^\infty(X)$ such that $\Lambda(f) = \int_X \langle g, f\rangle \, d\mu$ holds for all $f \in L^p(X)$ with $1 < p < \infty$. For an arbitrary $f \in L^1(X)$, Lemma 2.17 then provides a sequence $f_n \in L^p(X)$ with $f_n \to f_1$ in L^1, and Hölder's inequality implies

$$\left| \int_X \langle g, f\rangle \, d\mu - \int_X \langle g, f_n\rangle \, d\mu \right| \leqslant \int_X |\langle g, f - f_n\rangle| \, d\mu \leqslant \|g\|_{L^\infty} \cdot \|f - f_n\|_{L^1} \to 0,$$

thus

$$\Lambda(f) = \lim_{n\to\infty} \Lambda(f_n) = \lim_{n\to\infty} \int_X \langle g, f_n\rangle \, d\mu = \int_X \langle g, f\rangle \, d\mu.$$

□

Proof of Theorem 2.4 for $p = 1$ and $\mu(X) = \infty$. We assume X is σ-finite, so $X = \bigcup_{n\in\mathbb{N}} X_n$ for subsets $X_n \subset X$ with $\mu(X_n) < \infty$, and without loss of generality

$$X_1 \subset X_2 \subset X_3 \subset \ldots.$$

Any $\Lambda \in (L^1(X))^*$ gives rise to functionals $\Lambda_n \in (L^1(X_n))^*$ for every $n \in \mathbb{N}$, defined by

$$\Lambda_n(f) := \Lambda(f_n), \quad \text{where} \quad f_n := \begin{cases} f & \text{on } X_n, \\ 0 & \text{on } X \backslash X_n, \end{cases}$$

and they satisfy

$$\|\Lambda_n\|_{(L^1)^*} = \sup_{f \in L^1(X_n)\backslash\{0\}} \frac{|\Lambda(f_n)|}{\|f_n\|_{L^1}} \leqslant \sup_{f \in L^1(X)\backslash\{0\}} \frac{|\Lambda(f)|}{\|f\|_{L^1}} = \|\Lambda\|_{(L^1)^*}.$$

Applying the theorem for the case of finite measure, we obtain functions $g_n \in L^\infty(X_n)$ such that $\Lambda(f) = \int_{X_n} \langle g_n, f\rangle \, d\mu$ for every $f \in L^1(X)$ that vanishes outside of X_n, with norms satisfying $\|g_n\|_{L^\infty} \leqslant \|\Lambda\|_{(L^1)^*}$ for all n. Notice that for $n > m \geqslant 1$ and a function $f \in L^1(X)$ that vanishes outside of X_m, f also vanishes outside of X_n and thus satisfies

$$\int_{X_m} \langle g_m, f\rangle \, d\mu = \Lambda(f) = \int_{X_n} \langle g_n, f\rangle \, d\mu = \int_{X_m} \langle g_n, f\rangle \, d\mu,$$

implying $\int_{X_m} \langle g_m - g_n, f \rangle \, d\mu = 0$ for all $f \in L^1(X_m)$. It follows that $g_m - g_n|_{X_m} \in L^\infty(X_m)$ defines the trivial element of $(L^1(X_m))^*$ and therefore vanishes almost everywhere. This shows that each g_n can in fact be regarded as the restriction to X_n of a single function $g : X \to V$, and since $\|g_n\|_{L^\infty} \leqslant \|\Lambda\|_{(L^1)^*}$ for every n, the set on which $|g| > \|\Lambda\|_{(L^1)^*}$ is the union of countably many sets of measure zero, implying $g \in L^\infty(X)$ with $\|g\|_{L^\infty} \leqslant \|\Lambda\|_{(L^1)^*}$.

We claim finally that $\Lambda(f) = \int_X \langle g, f \rangle \, d\mu$ holds for every $f \in L^1(X)$. To see this, for each $n \in \mathbb{N}$ define $h_n \in L^1(X)$ as the product of f with the characteristic function of X_n, so $\|f - h_n\|_{L^1} = \int_{X \setminus X_n} |f| \, d\mu \to 0$ as $n \to \infty$. Using the continuity of Λ and Hölder's inequality, we now conclude

$$\Lambda(f) = \lim_{n \to \infty} \Lambda(h_n) = \lim_{n \to \infty} \int_X \langle g, h_n \rangle \, d\mu = \int_X \langle g, f \rangle \, d\mu.$$

□

3. Separability of L^p

Recall that a topological space is called **separable** if it contains a countable dense subset. The simplest examples that come to mind are finite-dimensional vector spaces, e.g. $\mathbb{Q}^n$ is a countable dense subset of $\mathbb{R}^n$. In this section, we would like to prove that $L^p(X)$ is also separable when $1 \leqslant p < \infty$. This requires some measure-theoretic assumptions on X, so in order to avoid overcomplicating the problem, we shall restrict ourselves to the case where X is a subset Ω of $\mathbb{R}^n$. (See [Sal16, §4.3] for a treatment of more general situations.)

Theorem 3.1. *For any $p \in [1, \infty)$ and any Lebesgue-measurable set $\Omega \subset \mathbb{R}^n$ endowed with the Lebesgue measure m, the space $L^p(\Omega)$ is separable.*

We shall prove this by constructing an explicit countable set of functions $Q(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$ that is dense in $L^p(\mathbb{R}^n)$. Given any $f \in L^p(\Omega)$ for $\Omega \subset \mathbb{R}^n$, one can then extend f to a function $\widehat{f} \in L^p(\mathbb{R}^n)$ that vanishes outside of Ω, find a sequence $\widehat{f}_k \in Q(\mathbb{R}^n)$ converging to $\widehat{f}$ in L^p, and observe that the restrictions $f_k := \widehat{f}_k|_\Omega$ therefore converge in L^p to f, proving that the countable set $Q(\Omega) := \{ f|_\Omega \mid f \in Q(\mathbb{R}^n) \}$ is dense in $L^p(\Omega)$.

The set $Q(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$ is easy to describe. In the following, we denote the characteristic function of a subset $A \subset \mathbb{R}^n$ by

$$\chi_A : \mathbb{R}^n \to \mathbb{R}, \qquad \chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Let us first fix a countable dense subset V_0 in the vector space V where our functions take their values; this is clearly possible since $\dim V < \infty$. (If we were allowing V to be an infinite-dimensional Banach space, then we would now add the assumption that V is separable.) We refer to a set $Q \subset \mathbb{R}^n$ as a **dyadic cube** if Q is of the form

$$Q = \left[\frac{m_1}{2^N}, \frac{m_1 + 1}{2^N} \right] \times \ldots \times \left[\frac{m_n}{2^N}, \frac{m_n + 1}{2^N} \right] \subset \mathbb{R}^n$$

for some $m_1, \ldots, m_n, N \in \mathbb{Z}$ with $N \geqslant 0$. Observe that the set of all dyadic cubes is countable, and so therefore is the set of characteristic functions $\chi_Q : \mathbb{R}^n \to \mathbb{R}$ of dyadic cubes. It follows that for every $k \in \mathbb{N}$, the set of k-tuples of dyadic cubes is countable, and thus so is the set of all finite tuples of dyadic cubes. Finally, for each individual tuple $(Q_1, \ldots, Q_k)$ of dyadic cubes, there is a countable set of functions $f : \mathbb{R}^n \to V$ of the form

$$f = \sum_{j=1}^{k} \chi_{Q_j} v_j, \qquad v_1, \ldots, v_k \in V_0.$$

We define $Q(\mathbb{R}^n)$ to be the set of all functions of this type, i.e. all finite linear combinations (with coefficients in the countable set V_0) of characteristic functions of dyadic cubes. All of these functions are bounded and have compact support, so they belong to $L^p(\mathbb{R}^n)$ for every $p \in [1, \infty]$. Our goal is to prove:

Proposition 3.2. *For every $p \in [1, \infty)$, the countable set $Q(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$.*

We will use the following fundamental fact from the theory of Lebesgue integration. Recall that a function is called **simple** (or sometimes a **step function**) if it takes only finitely many values. A simple function on a measure space (X, μ) is measurable if and only if it is a finite linear combination of characteristic functions of measurable sets, and it is then integrable if and only if all of those sets have finite measure, which is equivalent to saying that the function's support has finite measure. The integrable simple functions form a linear subspace of $L^p(X)$ for every $p \in [1, \infty]$, and shall denote it by

$$S(X) \subset L^p(X).$$

Lemma 3.3. *For every measure space (X, μ) and $1 \leqslant p < \infty$, $S(X)$ is dense in $L^p(X)$.*

Proof. Denote the measure on X by μ. Depending on your definition of integration, the $p = 1$ case may be understood as either a theorem or a tautology; e.g. [Lan93] defines $L^1(X)$ to be a quotient (modulo equality almost everywhere) of the L^1-closure of $S(X)$. Let us take the more common definition as in [Sal16], where $\int_{\mathbb{R}^n} f\, d\mu \in [0, \infty]$ for a measurable function $f : X \to [0, \infty]$ is the supremum of $\int_{\mathbb{R}^n} s\, d\mu$ for all measurable simple functions with $0 \leqslant s \leqslant f$, and for $f : X \to \mathbb{R}$, $\int_X f\, d\mu := \int_X f^+ \, d\mu - \int_X f^- \, d\mu$ with $f^\pm : X \to [0, \infty)$ such that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Then given $f : X \to \mathbb{R}$ of class L^p, there exist increasing sequences of measurable simple functions $0 \leqslant f_1^\pm \leqslant f_2^\pm \leqslant \ldots \leqslant f^\pm$ such that $f_n^\pm \to f^\pm$ pointwise as $n \to \infty$. Then since $(x + y)^p \geqslant x^p + y^p$ for all $x, y \geqslant 0$ and $p \geqslant 1$,[8]

$$\int_X |f^+|^p \, d\mu + \int_X |f^-|^p \, d\mu = \int_X (|f^+|^p + |f^-|^p) \, d\mu \leqslant \int_X |f^+ + f^-|^p \, d\mu = \int_X |f|^p \, d\mu < \infty,$$

so $\int_X |f_n^\pm|^p \, d\mu \leqslant \int_X |f|^p \, d\mu < \infty$. This implies that every $|f_n^\pm|^p$ (and therefore also every $f_n^\pm$) is a finite linear combination of characteristic functions of sets with finite measure, so $f_n^\pm \in S(X)$, and thus $f_n := f_n^+ - f_n^- \in S(X)$. Now $|f - f_n|^p \to 0$ pointwise, and using the convexity of the function $x \mapsto x^p$,

$$|f - f_n|^p = |(f^+ - f_n^+) - (f^- - f_n^-)|^p \leqslant 2^{p-1}|f^+ - f_n^+|^p + 2^{p-1}|f^- - f_n^-|^p \leqslant 2^p \left(|f^+|^p + |f^-|^p\right),$$

where the function on the right hand side is integrable, so the dominated convergence theorem implies $\int_X |f - f_n|^p \, d\mu \to 0$. The result for real-valued functions now easily extends to functions valued in the finite-dimensional vector space V by choosing a real basis as in §0.1. $\square$

Exercise 3.4. Show that $S(X)$ is dense in $L^\infty(X)$ if and only if $\mu(X) < \infty$.

With Lemma 3.3 in hand, our goal is now to show that $Q(\mathbb{R}^n)$ is dense in $S(\mathbb{R}^n)$.

Lemma 3.5. *Every open subset $A \subset \mathbb{R}^n$ is a union of a sequence of dyadic cubes $Q_1, Q_2, Q_3, \ldots$ whose interiors are all pairwise disjoint.*

Proof. Let $\mathcal{O}$ denote the set of all dyadic cubes that are contained in A. Since dyadic cubes can be arbitrarily small, $A = \bigcup_{Q \in \mathcal{O}} Q$, and the set $\mathcal{O}$ is countable since there are only countably many dyadic cubes in total. Write $\mathcal{O} = \{\widehat{Q}_1, \widehat{Q}_2, \ldots\}$; this is not the desired sequence since it contains pairs $\widehat{Q}_j, \widehat{Q}_k$ whose interiors intersect, but observe that for any such pair, the part of $\widehat{Q}_j$ disjoint from $\widehat{Q}_k$ can be covered by finitely many smaller dyadic cubes whose interiors are disjoint from each other and from $\widehat{Q}_k$. We can therefore construct a new sequence $Q_1, Q_2, \ldots$ by setting $Q_1 := \widehat{Q}_1$ and then replacing each $\widehat{Q}_k$ for $k \geqslant 2$ with a finite collection of dyadic cubes with interiors that are disjoint from each other and from $\bigcup_{j=1}^{k-1} \widehat{Q}_j$. $\square$

Lemma 3.6. *For every open subset $A \subset \mathbb{R}^n$ with $m(A) < \infty$ and every $v \in V$, $\epsilon > 0$ and $p \in [1, \infty)$, $Q(\mathbb{R}^n)$ contains a function f with $\|\chi_A v - f\|_{L^p} < \epsilon$.*

Proof. Pick $v_0 \in V_0$ with $|v - v_0| < \epsilon^p / m(A)$ and let $Q_1, Q_2, Q_3, \ldots$ denote the sequence of dyadic cubes provided by Lemma 3.5 to cover A. Since $\sum_{k=1}^\infty m(Q_k) = m(A) < \infty$, we have $\lim_{k\to\infty} \sum_{j=k}^\infty m(Q_j) = 0$, so the functions $f_k := \left(\sum_{j=1}^k \chi_{Q_j}\right) v_0$ satisfy

$$\begin{aligned} \|\chi_A v - f_k\|_{L^p}^p &= \sum_{j=1}^k |v - v_0|^p m(Q_j) + \sum_{j=k+1}^\infty |v|^p m(Q_j) \\ &\leqslant |v - v_0| \cdot m(A) + |v|^p \sum_{j=k+1}^\infty m(Q_j) \to |v - v_0| \cdot m(A) \quad \text{as} \quad k \to \infty, \end{aligned}$$

thus $\|\chi_A v - f_k\|_{L^p} < \epsilon$ for k sufficiently large. $\square$

[8] This inequality is an easy algebraic exercise when $p \in \mathbb{N}$, but when p is not an integer, one can argue as follows. Assume $y > 0$ since otherwise the result is obvious. Dividing by y^p, it is then equivalent to prove $(1+x)^p \geqslant 1 + x^p$ for all $x \geqslant 0$ and $p \geqslant 1$. Differentiating with respect to x, it is easy to show that $(1+x)^p - 1 - x^p$ is an increasing function on $\{x \geqslant 0\}$ if $p \geqslant 1$, and since it vanishes at $x = 0$, it is therefore nonnegatve.

We next appeal to the fact that the Lebesgue measure m is **outer regular**, meaning that for every Lebesgue-measurable set $A \subset \mathbb{R}^n$,

$$m(A) = \inf \left\{ m(A') \;\middle|\; A \subset A' \subset \mathbb{R}^n,\ A' \text{ open} \right\}.$$

It follows that whenever $m(A) < \infty$, there exists a nested sequence of open sets $A_1 \supset A_2 \supset A_3 \supset \ldots \supset A' := \bigcap_{k \in \mathbb{N}} A_k \supset A$ such that $m(A) = m(A')$. The set A' is not generally open, but it is a Borel set, a so-called G_δ. In this situation, $|\chi_A - \chi_{A_n}|^p$ converges almost everywhere to 0, and since it is clearly also bounded by a fixed integrable function for every n, the dominated convergence theorem implies $\chi_{A_n} \to \chi_A$ in L^p. Since Lemma 3.6 provides arbitrarily good approximations $f_n \in Q(\mathbb{R}^n)$ for each $\chi_{A_n} v \in L^p(\mathbb{R}^n)$, we've proved:

Lemma 3.7. *Lemma 3.6 remains true with A replaced by an arbitrary Lebesgue-measurable set in $\mathbb{R}^n$ with finite measure.* □

Proof of Proposition 3.2 (and thus Theorem 3.1). By Lemma 3.3, it suffices to prove that $Q(\mathbb{R}^n)$ is dense in $S(\mathbb{R}^n)$ in the L^p-norm. Elements of $S(\mathbb{R}^n)$ are of the form $\sum_{j=1}^k \chi_{A_j} v_j$, where each $A_j \subset \mathbb{R}^n$ is Lebesgue measurable with finite measure and $v_j \in V$. By Lemma 3.7, each $\chi_{A_j} v_j$ can be approximated arbitrarily well in the L^p-norm by functions in $Q(\mathbb{R}^n)$, so we are done. □

It is not hard to see that for almost any interesting measure space (X, μ), $L^\infty(X)$ is not separable:

Exercise 3.8.

(a) Show that if E is a Banach space containing an uncountable discrete subset, then E is not separable.
(b) Suppose (X, μ) is a measure space containing infinitely many disjoint subsets with positive measure. Show that $L^\infty(X)$ contains an uncountable subset $S \subset L^\infty(X)$, consisting of functions that take only the values 0 and 1, such that $\|f - g\|_{L^\infty} = 1$ for any two distinct $f, g \in S$.
Hint: If you've forgotten or never seen the proof via Cantor's diagonal argument that $\mathbb{R}$ is uncountable, looking it up may help.

Exercise 3.9. Here is another nonseparable Banach space that sometimes arises naturally. Assume $\mathcal{H}$ is an infinite-dimensional separable Hilbert space, and let $\mathscr{L}(\mathcal{H})$ denote the Banach space of bounded linear operators $\mathcal{H} \to \mathcal{H}$. Use an orthonormal basis of $\mathcal{H}$ to find a continuous embedding of $L^\infty(X)$ into $\mathscr{L}(\mathcal{H})$ for a suitable measure space X, and deduce from this that $\mathscr{L}(\mathcal{H})$ cannot be separable.

4. Weak convergence

In finite dimensions, a sequence $x_k \in \mathbb{K}^n$ converges to $x_\infty \in \mathbb{K}^n$ if and only if the n sequences formed by the coordinates of these vectors all converge to the corresponding coordinates of x_∞. Writing $e_1, \ldots, e_n \in \mathbb{K}^n$ for the standard orthonormal basis, the latter condition can be expressed equivalently as

$$\lim_{k\to\infty} \langle e_j, x_k\rangle = \langle e_j, x_\infty\rangle \quad \text{for all} \quad j = 1, \ldots, n.$$

There is an obvious way to generalize this condition for a sequence x_k in an infinite-dimensional Hilbert space $\mathcal{H}$, though the resulting notion of convergence turns out to depend on a choice of orthonormal basis (see Exercise 4.2 below). A stronger condition that is clearly independent of any choice of basis is

$$\lim_{k\to\infty} \langle v, x_k\rangle = \langle v, x_\infty\rangle \quad \text{for all} \quad v \in \mathcal{H}.$$

In light of the Riesz representation theorem, this can be expressed equivalently as:

$$\lim_{k\to\infty} \Lambda(x_k) = \Lambda(x_\infty) \quad \text{for all} \quad \Lambda \in \mathcal{H}^*.$$

In this form, the condition also makes sense in arbitrary normed vector spaces, leading to the following important definition.

Definition 4.1. A sequence x_n in a normed vector space E is said to **converge weakly** to $x \in E$, written

$$x_n \rightharpoonup x,$$

if $\Lambda(x_n) \to \Lambda(x)$ for all $\Lambda \in E^*$.

With this definition in mind, the usual notion of convergence in a normed vector space (written "$x_n \to x$") is sometimes also called **strong convergence**. If $\dim E < \infty$, then it is easy to check that there is no difference between weak and strong convergence. In infinite-dimensional spaces, strong convergence clearly implies weak convergence due to the continuity of the functionals $\Lambda \in E^*$, but the following exercise shows that the converse is false.

Exercise 4.2. Suppose $\mathcal{H}$ is a Hilbert space containing an infinite orthonormal set $\{e_n \in \mathcal{H}\}_{n=1}^\infty$. Prove:

(a) The sequence e_n converges weakly to 0 but has no strongly convergent subsequence.
(b) For any bounded sequence $\lambda_n \in \mathbb{K}$, the sequence $x_n := \lambda_n e_n \in \mathcal{H}$ converges weakly to 0.
(c) For any unbounded sequence $\lambda_n \in \mathbb{K}$, $x_n := \lambda_n e_n \in \mathcal{H}$ satisfies $\lim_{n\to\infty}\langle e_j, x_n\rangle = 0$ for every $j \in \mathbb{N}$, but is nonetheless not weakly convergent.
Hint: Given a subsequence λ_{n_j} with $|\lambda_{n_j}| \geqslant j$ for $j = 1, 2, 3, \ldots$, find a convergent series of the form $v := \sum_{j=1}^\infty a_j e_{n_j} \in \mathcal{H}$ for suitable scalars $a_j \in \mathbb{K}$ such that $\langle v, x_{n_j}\rangle \not\to 0$ as $j \to \infty$.

Whenever we discuss a notion of convergence, there should be a topology in the background. Every normed vector space E comes with a natural topology, usually called the **norm topology** (sometimes also the **strong topology**), for which a set is open if and only if it is a union of open balls. The **weak topology** on E is generally different: it is the locally convex topology defined via the uncountably infinite family of seminorms

$$\{\|\cdot\|_\Lambda : E \to [0, \infty)\}_{\Lambda \in E^*}, \qquad \text{where} \qquad \|x\|_\Lambda := |\Lambda(x)|.$$

Notice that these are not norms since $\Lambda(x) = 0$ does not imply $x = 0$, but they are seminorms due to the linearity of Λ. The weak topology on E is thus the topology generated by all subsets of the form $\{x \in E \mid |\Lambda(x) - \Lambda(x_0)| < \epsilon\}$ for $x_0 \in E$, $\epsilon > 0$ and $\Lambda \in E^*$, and a sequence $x_n \in E$ converges to $x \in E$ in the weak topology if and only if it converges in all the seminorms, which means precisely that $\Lambda(x_n) \to \Lambda(x)$ for all $\Lambda \in E^*$, i.e. $x_n \rightharpoonup x$. A subset $\mathcal{U} \subset E$ that belongs to the weak topology is sometimes called **weakly open**. We will see below (see Remark 4.5) that all weakly open sets are also open in the usual sense, but the converse is generally false.

Remark 4.3. On a locally convex space E with topology generated by a family of seminorms $\{\|\cdot\|_\alpha : E \to [0,\infty)\}_{\alpha\in I}$, it is conventional to require that no nonzero $x \in E$ can satisfy $\|x\|_\alpha = 0$ for every $\alpha \in I$. This guarantees that a convergent sequence in E can only have one limit, and is equivalent to the condition that the topology defined by the seminorms on E is Hausdorff. The weak topology does satisfy this condition on every normed vector space, but this fact is not always obvious: it depends on the knowledge that for every nonzero $x \in E$ there exists a dual vector $\Lambda \in E^*$ with $\Lambda(x) \neq 0$. In all of the explicit examples that we deal with, it will be clear that this is true, e.g. Lemma 2.1 guarantees it for the L^p-spaces. For arbitrary Banach spaces, it follows from the Hahn-Banach theorem (see Lemma 2.6).

Exercise 4.4. This exercise gives an alternative characterization of the weak topology on a normed vector space E as the smallest topology for which the map $E \to \mathbb{K}$ defined by every dual vector $\Lambda \in E^*$ is continuous. In other words, the weak topology contains exactly the sets that must be considered open in order for these maps to be called continuous, but no more.

(a) Show that for every $\Lambda \in E^*$, the map $\Lambda : E \to \mathbb{K}$ is continuous in the weak topology.

Continuity of the maps $\Lambda : E \to \mathbb{K}$ means that for every $\Lambda \in E^*$ and every open set $\mathcal{U} \subset \mathbb{K}$, the set $\Lambda^{-1}(\mathcal{U}) \subset E$ needs to be open. Let $\mathcal{T}$ denote the smallest topology on E that contains all sets of this form, which means that a set is in $\mathcal{T}$ if and only if it is a union of finite intersections of sets of the form $\Lambda^{-1}(\mathcal{U})$ for arbitrary dual vectors $\Lambda \in E^*$ and open sets $\mathcal{U} \subset \mathbb{K}$. Part (a) shows that the weak topology contains $\mathcal{T}$. We now aim to show that these two topologies are the same.

(b) Show that for every $y \in E$, the translation map $\tau_y : E \to E : x \mapsto x + y$ is continuous with respect to the topology $\mathcal{T}$.
(c) Show that for every $\Lambda \in E^*$, $x_0 \in E$ and $\epsilon > 0$, the set $\{x \in E \mid |\Lambda(x - x_0)| < \epsilon\}$ is in $\mathcal{T}$, and conclude that $\mathcal{T}$ contains the weak topology.

Remark 4.5. Since every bounded linear functional $\Lambda : E \to \mathbb{K}$ is continuous in the norm topology on E, it follows from Exercise 4.4 that the norm topology contains the weak topology, i.e. every weakly open set is also open with respect to the norm. In general, however, the norm topology is strictly larger, e.g. if E is an infinite-dimensional Hilbert space, then Exercise 4.2 exhibits a sequence $x_n \in E$ that converges to 0 in the weak topology but not in the norm topology—the reason being that the norm topology has too many open neighborhoods of 0 for x_n to lie in all of them for n large. Relatedly, the fact that $\|x_n\| = 1$ for all n but $x_n \rightharpoonup 0$ in that exercise demonstrates that the norm $\|\cdot\| : E \to [0,\infty)$ is not a continuous function in the weak topology, though it is of course continuous in the norm topology.

Exercise 4.6. In the setting of Exercise 4.2, show that every neighborhood of $0 \in \mathcal{H}$ in the weak topology contains infinitely many of the vectors $x_n := \sqrt{n}e_n$ for $n \in \mathbb{N}$. In particular, the closure of the set $\{e_1, \sqrt{2}e_2, \sqrt{3}e_3, \dots\} \subset \mathcal{H}$ contains 0.
Remark: In a topological space, a set is closed if and only if its complement is open, and the closure of a set is by definition the intersection of all closed sets containing that set. Exercise 4.2 shows that the sequence $\sqrt{n}e_n$ has no subsequence weakly convergent to 0, so the present exercise demonstrates that the notion of the "closure" of a discrete set in the weak topology does not match your intuition from the theory of metric spaces—this shows in fact that the weak topology on $\mathcal{H}$ is not metrizable.

Combining Definition 4.1 with the Riesz representation theorem leads naturally to the following notion:

Definition 4.7. For a measure space (X,μ) and $1 \leqslant p < \infty$ such that either X is σ-finite or $p > 1$, we say that a sequence $f_n \in L^p(X)$ is **weakly L^p-convergent** to a function $f \in L^p(X)$ and write $f_n \overset{L^p}{\rightharpoonup} f$ if for every $g \in L^q(X)$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$\int_X \langle g, f_n \rangle \, d\mu \to \int_X \langle g, f \rangle \, d\mu.$$

For $p = \infty$, the notion of convergence in Definition 4.7 still makes sense but cannot be called "weak convergence" since the dual space of $L^\infty(X)$ is generally larger than $L^1(X)$. But we could instead view $L^\infty(X)$ as the dual space of $L^1(X)$ and fit this notion into the following context. For a normed vector space E, there is a natural topology on its dual space E^* that is generally even weaker[9] than the weak topology. The **weak* topology** on E^* is, namely, the locally convex topology defined via the family of seminorms

$$\{\|\cdot\|_x : E^* \to [0,\infty)\}_{x\in E}, \qquad \text{where} \qquad \|\Lambda\|_x := |\Lambda(x)|.$$

In light of the natural inclusion $E \hookrightarrow E^{**}$, this family of seminorms is a subset of the family that defines the weak topology, though the two families are exactly the same whenever E is a reflexive Banach space, that is:

Proposition 4.8. *If E is a reflexive Banach space, then the weak and weak* topologies on E^* are identical.* □

We observe that since the space $L^p(X)$ for $1 < p < \infty$ is reflexive and can be identified with the dual space of $L^q(X)$ for $\frac{1}{p} + \frac{1}{q} = 1$, $L^p(X)$ has a natural weak* topology which is the same as its weak topology. On the other hand, the analogue of Definition 4.7 for $p = \infty$ describes convergence in the weak* topology on $L^\infty(X)$, which is strictly weaker than the weak topology, due to the fact that the dual of $L^\infty(X)$ is strictly larger than $L^1(X)$.

In analogy with Exercise 4.4, one can show that the weak* topology is the smallest topology such that for every $x \in E$, the function $E^* \to \mathbb{K} : \Lambda \mapsto \Lambda(x)$ is continuous. A sequence $\Lambda_n \in E^*$ is **weak* convergent** if and only if for every $x \in E$, $\Lambda_n(x) \to \Lambda(x)$, i.e. the functionals $\Lambda_n : E \to \mathbb{K}$ converge *pointwise* to $\Lambda : E \to \mathbb{K}$. Notice that for every nonzero $\Lambda \in E^*$, there necessarily exists a vector $x \in E$ for which $\|\Lambda\|_x \neq 0$, thus limits of weak* convergent sequences are unique and the weak* topology is Hausdorff (cf. Remark 4.3). This provides an easy proof (without requiring the Hahn-Banach theorem) that the weak topology on E^* is also Hausdorff, since every weak* open subset of E^* is also weakly open; or in terms of convergence, every weakly convergent sequence also converges in the weak* topology.

Remark 4.9. The definitions above do not require E to be complete, but there is a subtlety to be aware of when considering normed vector spaces that are not Banach spaces. If E is a Banach space and $F \subset E$ is a proper dense subspace, then $F^* = E^*$ since every bounded linear functional on F extends uniquely to one on E. The norms on F^* and E^* are also the same, so as Banach spaces they are identical, but their weak* topologies may nonetheless be different. In practice, we will only consider examples in which E is complete, in which case the reader may feel free to ignore this remark.

The next result demonstrates that the weak* topology is often, indeed, *much* weaker than the norm topology on E^*. Having fewer open sets means that sequences can more easily converge, so they are more likely to have convergent subsequences.

Theorem 4.10 (Banach-Alaoglu theorem, separable case)**.** *Assume E is a separable normed vector space. Then every bounded sequence in E^* has a weak* convergent subsequence.*

Proof. Fix a sequence $\Lambda_n \in E^*$ satisfying $\|\Lambda_n\| \leqslant C$ for some constant $C > 0$.

Claim 1: If $F \subset E$ is a countable subset, then after replacing Λ_n with a suitable subsequence, we can assume $\Lambda_n(x)$ converges for every $x \in F$. We prove this via the Cantor diagonal argument. Let $F = \{x_1, x_2, x_3, \ldots\}$, and observe that for each $k, n \in \mathbb{N}$, $|\Lambda_n(x_k)| \leqslant C\|x_k\|$, thus for every fixed $k \in \mathbb{N}$ the sequence $\{\Lambda_n(x_k)\}_{n=1}^\infty$ is bounded in $\mathbb{K}$. Let $\Lambda_n^{(1)}$ denote a subsequence of Λ_n such

[9]When comparing two topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ on the same set, one says that $\mathcal{T}_1$ is **weaker** than $\mathcal{T}_2$ if $\mathcal{T}_1 \subset \mathcal{T}_2$. In this context, "weaker" is a synonym for "smaller," and the word **coarser** is also sometimes used with the same meaning, while in the other direction, one says that $\mathcal{T}_2$ is **stronger** / **finer** / **larger** than $\mathcal{T}_1$. Weakening a topology makes it easier for sequences to converge, i.e. every $\mathcal{T}_2$-convergent sequence is also $\mathcal{T}_1$-convergent, but there may also be $\mathcal{T}_1$-convergent sequences that are not $\mathcal{T}_2$-convergent. Similarly, weakening the topology makes it easier for maps from other spaces into X to be continuous, but harder for functions defined on X to be continuous.

that the sequence $\Lambda_n^{(1)}(x_1)$ converges in $\mathbb{K}$. Then choose $\Lambda_n^{(2)}$ to be a subsequence of $\Lambda_n^{(1)}$ such that the sequence $\Lambda_n^{(2)}(x_2)$ also converges in $\mathbb{K}$. Continuing in this way, we obtain a sequence of sequences such that the diagonal subsequence $\Lambda_n^{(n)}$ has the desired property.

Claim 2: If $F \subset E$ is a dense subset such that $\Lambda_n(x)$ converges for every $x \in F$, then $\Lambda_n(x)$ also converges for every $x \in E$. Indeed, for any given $x \in E$, one can choose $x' \in F$ arbitrarily close to x and then estimate

$$\begin{aligned} |\Lambda_m(x) - \Lambda_n(x)| &\leqslant |\Lambda_m(x) - \Lambda_m(x')| + |\Lambda_m(x') - \Lambda_n(x')| + |\Lambda_n(x') - \Lambda_n(x)| \\ &\leqslant 2C\|x - x'\| + |\Lambda_m(x') - \Lambda_n(x')|. \end{aligned}$$

Since $\Lambda_n(x')$ is a Cauchy sequence in $\mathbb{K}$, this shows that $\Lambda_n(x)$ is also a Cauchy sequence.

Finally, since E is separable, we are free to assume the two subsets denoted by $F \subset E$ in claims 1 and 2 are the same set, so both claims together allow us to replace Λ_n with a subsequence such that $\Lambda_n(x)$ is convergent for every $x \in E$. Define $\Lambda : E \to \mathbb{K}$ by

$$\Lambda(x) := \lim_{n\to\infty} \Lambda_n(x).$$

It is easy to check that Λ is linear and satisfies $|\Lambda(x)| \leqslant C\|x\|$, thus $\Lambda \in E^*$ and Λ_n is weak* convergent to Λ. □

Since $L^p(\Omega)$ is separable and reflexive for $1 < p < \infty$ and $\Omega \subset \mathbb{R}^n$, this implies:

Corollary 4.11. *Assume $\Omega \subset \mathbb{R}^n$ is a Lebesgue-measurable subset and $1 < p < \infty$. Then every L^p-bounded sequence $f_k \in L^p(\Omega)$ has a weakly L^p-convergent subsequence.* □

Exercise 4.12. Find a sequence $f_n \in L^p(\mathbb{R})$ for $1 < p < \infty$ that converges weakly to 0 but satisfies $\|f_n\|_{L^p} = 1$ for all n, and deduce that f_n has no L^p-convergent subsequence.

Remark 4.13. $L^\infty(\Omega)$ is also the dual space of a separable Banach space, namely $L^1(\Omega)$, so Theorem 4.10 implies that L^∞-bounded sequences have weak* convergent subsequences. This case was not included in Corollary 4.11 since the weak and weak* topologies on $L^\infty(\Omega)$ are not the same.

Example 4.14. There are two troubles with the case $p = 1$ in Corollary 4.11, one more serious than the other. The less serious problem is that $L^1(\Omega)$ is not the dual space of $L^\infty(\Omega)$, though since it is *contained* in the dual of $L^\infty(\Omega)$, one could still deduce from Theorem 4.10 a result about weakly L^1-convergent subsequences if $L^\infty(\Omega)$ were separable. The lack of separability is the more serious problem, and the following example shows that it cannot be overcome. For $n \in \mathbb{N}$, define $f_n \in L^1(\mathbb{R})$ to be the characteristic function of the interval $[n-1, n]$, so clearly $\|f_n\|_{L^1} = 1$ for every n. Consider an arbitrary subsequence f_{n_k} for some $1 \leqslant n_1 < n_2 < n_3 < \ldots$, and define a function $g \in L^\infty(\mathbb{R})$ such that $g = (-1)^k$ on $[n_k - 1, n_k]$ for each $k \in \mathbb{N}$ and $g = 0$ everywhere else. Then the sequence $\int_{-\infty}^{\infty} g(x) f_{n_k}(x)\, dx = (-1)^k$ does not converge, thus f_{n_k} cannot be weakly convergent. The problem here is in essence that $L^\infty(\mathbb{R})$ is just too large a space, and as a consequence, weak L^1-convergence is harder to achieve than in the case $p > 1$.

The Banach-Alaoglu theorem implies that even though the unit sphere in $L^p(X)$ for $1 < p < \infty$ is not compact, it is *weakly* compact: arbitrary sequences with unit norm need not have accumulation points with respect to the L^p-norm, but in the weak topology they do. You may be wondering *which* points can arise as accumulation points in this scenario, e.g. must they also lie in the unit sphere? Let us show that they are at least bounded:

Proposition 4.15. *In any normed vector space $(E, \|\cdot\|)$, if $x_n \rightharpoonup x$, then $\|x\| \leqslant \liminf_{n\to\infty} \|x_n\|$.*

Proof. Using Lemma 2.6, choose $\Lambda \in E^*$ with $\|\Lambda\| = 1$ and $\Lambda(x) = \|x\|$. Then since $\Lambda(x_n) \to \Lambda(x)$ and $|\Lambda(x_n)| \leqslant \|x_n\|$,

$$\|x\| = \Lambda(x) = \liminf_{n\to\infty} \Lambda(x_n) \leqslant \liminf_{n\to\infty} \|x_n\|.$$

□

Recall from §2.3 that $L^p(X)$ is uniformly convex for $1 < p < \infty$. The next result therefore gives a useful criterion for strong L^p-convergence in terms of weak convergence. But in light of the Banach-Alaoglu theorem, it also says something that your geometric intuition may find shocking: the unit sphere is not closed in the weak topology. In particular, every sequence in the unit sphere that fails to have a strongly convergent subsequence has one that converges weakly to something in the *interior* of the unit sphere. It is known in fact that for any infinite-dimensional normed vector space E, the weak* closure of the unit sphere in E^* is the entire closed unit ball (cf. [BS18, Corollary 3.28]).

Theorem 4.16. *If $(E, \|\cdot\|)$ is a uniformly convex Banach space and $x_n \in E$ is a sequence with $x_n \rightharpoonup x$ and $\|x_n\| \to \|x\|$, then $x_n \to x$.*

Proof. We can assume $x \neq 0$ since the statement is otherwise trivial. Since the norms converge, we can also replace x_n and x with $x_n/\|x_n\|$ and $x/\|x\|$ respectively in order to assume $\|x_n\| = \|x\| = 1$ for all n without loss of generality. The weak convergence $x_n \rightharpoonup x$ implies $x_n + x \rightharpoonup 2x$, so combining Proposition 4.15 with the triangle inequality now gives

$$2 = \|2x\| \leqslant \liminf_{n\to\infty} \|x_n + x\| \leqslant \limsup_{n\to\infty} \|x_n + x\| \leqslant \limsup_{n\to\infty} (\|x_n\| + \|x\|) = 2,$$

and hence $\|x_n + x\| \to 2$, or equivalently, $\left\|\frac{x_n+x}{2}\right\| \to 1$. The conclusion $\|x_n - x\| \to 0$ then follows from uniform convexity. □

Just out of interest, let us state the more general version of the Banach-Alaoglu theorem, which does not require E to be separable. Its meaning is a bit harder to interpret, since the weak* topology is not generally first countable, so compactness need not imply sequential compactness.[10] We will neither prove nor make use of this version of the theorem, but proofs may be found e.g. in [RS80, §IV.5] or [BS18, §3.2]; it is a consequence of *Tychonoff's theorem* on the compactness of arbitrary products of compact topological spaces, which is equivalent to the axiom of choice (see [Wen18, §6]).

Theorem 4.17 (Banach-Alaoglu theorem, general case)**.** *For any normed vector space E, the closed unit ball in E^* is compact in the weak* topology.* □

[10] A topological space X is called **first countable** if for every $x \in X$, there is a countable sequence $\mathcal{U}_n \subset X$ of neighborhoods of x such that every neighborhood of x contains $\mathcal{U}_n$ for some $n \in \mathbb{N}$. First countability is a sufficient condition for the compactness of a subset to imply that all of its sequences have convergent subsequences (see e.g. [Wen18, §5]). It is easy to show that all metrizable topologies have this property, but scenarios like that of Exercise 4.6 reveal that the weak and weak* topologies generally do not.

5. Mollification

For this section, we consider functions defined on Lebesgue-measurable sets $\Omega \subset \mathbb{R}^n$ and define all integrals with respect to the Lebesgue measure m. For a Lebesgue-integrable function $f : \Omega \to V$, we write the integral as

$$\int_\Omega f\,dm =: \int_\Omega f(x)\,dx =: \int_\Omega f(x_1, \dots, x_n)\,dx_1 \dots dx_n.$$

We saw in §3 that the space $Q(\mathbb{R}^n)$ of functions that take constant values on finitely many dyadic cubes is dense in $L^p(\mathbb{R}^n)$ for every $p \in [1, \infty)$. It is not hard to convince oneself that every function in $Q(\mathbb{R}^n)$ can in turn be approximated arbitrarily well in the L^p-norm (again for $p < \infty$) by a compactly supported *continuous* function, thus proving that the space of continuous functions with compact support is dense in $L^p(\Omega)$.[11] We would now like to prove something more ambitious, and far more useful in applications:

Theorem 5.1. *For every $p \in [1, \infty)$, $C^\infty(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ is a dense subspace of $L^p(\mathbb{R}^n)$.*

Two important generalizations of Theorem 5.1 follow almost immediately. First: one can replace $\mathbb{R}^n$ by an arbitrary open subset $\Omega \subset \mathbb{R}^n$ and show that $C^\infty(\Omega) \cap L^p(\Omega)$ is dense in $L^p(\Omega)$. For the proof, one extends any given function $f \in L^p(\Omega)$ to $\widetilde{f} \in L^p(\mathbb{R}^n)$ via

$$\widetilde{f} := \begin{cases} f & \text{on } \Omega, \\ 0 & \text{on } \mathbb{R}^n \backslash \Omega, \end{cases}$$

and then approximates f with $f_\epsilon|_\Omega$ for smooth functions $f_\epsilon \in C^\infty(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ that approximate $\widetilde{f}$ in $L^p(\mathbb{R}^n)$. Further: $C^\infty(\Omega) \cap L^p(\Omega)$ in this statement can be replaced with

$$C_0^\infty(\Omega) := \left\{ f \in C^\infty(\Omega) \;\middle|\; f \text{ has compact support in } \Omega \right\}.$$

To see this, one first chooses for any given $f \in L^p(\Omega)$ and $\epsilon > 0$ an approximation $f_\epsilon \in C^\infty(\Omega) \cap L^p(\Omega)$ with $\|f - f_\epsilon\|_{L^p} < \frac{\epsilon}{2}$, and then replaces f_ϵ with βf_ϵ for a smooth compactly supported function $\beta : \Omega \to [0, 1]$ that satisfies $\beta|_{\mathcal{U}} \equiv 1$ for a sufficiently large open subset $\mathcal{U} \subset \Omega$. Taking a sequence of such cutoff functions β_N and subsets $\mathcal{U}_N$ such that $\bigcup_{N \in \mathbb{N}} \mathcal{U}_N = \Omega$, one can arrange that

$$\|f_\epsilon - \beta_N f_\epsilon\|_{L^p} < \frac{\epsilon}{2} \quad \text{und therefore} \quad \|f - \beta_N f_\epsilon\|_{L^p} < \epsilon$$

for $N \gg 0$ sufficiently large. For more details on this generalization, see e.g. [LL01, §2.19]; we summarize the result as follows:

Corollary 5.2. *For every $p \in [1, \infty)$ and every open subset $\Omega \subset \mathbb{R}^n$, $C_0^\infty(\Omega)$ is dense in $L^p(\Omega)$.* □

Exercise 5.3. Show that the space of bounded continuous functions is not dense in $L^\infty(\mathbb{R})$.

We prove Theorem 5.1 in the next several subsections using the *convolution*, a construction that is worth getting to know well, as it has a multitude of applications beyond this one theorem.

5.1. Continuity under translation.

For $v \in \mathbb{R}^n$ and a function $f : \mathbb{R}^n \to V$, the **translation operator** τ_v produces a new function $\tau_v f : \mathbb{R}^n \to V$ defined by

$$(\tau_v f)(x) := f(x + v).$$

Clearly τ_v defines a bounded and norm-preserving linear map $L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$. Continuity of f is equivalent to the condition that for every convergent sequence $v_k \to v_\infty$ in $\mathbb{R}^n$, the functions $\tau_{v_k} f$ converge pointwise to $\tau_{v_\infty} f$. This is not true in general for functions $f \in L^p(\mathbb{R}^n)$ since they are not generally continuous, but it will be useful to know that it becomes true if pointwise convergence is replaced by L^p-convergence:

[11] For a discussion of the density of $C(X)$ in $L^p(X)$ on more general measure spaces X, see [Sal16, §4.3].

Theorem 5.4. *If $1 \leqslant p < \infty$ and $f \in L^p(\mathbb{R}^n)$, then the map $\mathbb{R}^n \to L^p(\mathbb{R}^n) : v \mapsto \tau_v f$ is continuous.*

Proof. Since every τ_v defines a bounded linear operator $L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$ with $\|\tau_v\|_{\mathscr{L}(L^p)} = 1$, we have $\|\tau_{w+v} f - \tau_w f\|_{L^p} = \|\tau_w(\tau_v f - f)\|_{L^p} \leqslant \|\tau_v f - f\|_{L^p}$. It will thus suffice to prove that $\|\tau_v f - f\|_{L^p} \to 0$ as $v \to 0$ for all f belonging to some dense subset of $L^p(\mathbb{R}^n)$. Let $\widehat{Q}(\mathbb{R}^n)$ denote the space of all finite linear combinations $\sum_j \chi_{Q_j} f_j : \mathbb{R}^n \to V$, where $f_j \in V$ and each $Q_j \subset \mathbb{R}^n$ is a cube, i.e. any set of the form $[a_1, a_1 + d] \times \ldots \times [a_n, a_n + d]$ for $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and $d > 0$. Then $\widehat{Q}(\mathbb{R}^n)$ contains the set $Q(\mathbb{R}^n)$ spanned by characteristic functions of dyadic cubes, and having proved in Proposition 3.2 that the latter is dense in $L^p(\mathbb{R}^n)$, it follows that $\widehat{Q}(\mathbb{R}^n)$ is also dense. For an individual cube $Q = [a_1 + d, \ldots, a_n + d]$, we have

$$\|\tau_v \chi_Q - \chi_Q\|_{L^p}^p = \int_{\mathbb{R}^n} |\tau_v \chi_Q - \chi_Q|^p \, dm = m\big((v + Q)\backslash Q\big) + m\big(Q\backslash(v + Q)\big) \to 0 \quad \text{as} \quad v \to 0,$$

thus for any $f = \sum_j \chi_{Q_j} f_j \in \widehat{Q}(\mathbb{R}^n)$, Minkowski's inequality gives

$$\|\tau_v f - f\|_{L^p} \leqslant \sum_j \left\|\tau_v \chi_{Q_j} - \chi_{Q_j}\right\|_{L^p} \cdot |f_j| \to 0 \quad \text{as} \quad v \to 0.$$

□

5.2. **Convolution and regularity.** The **convolution** of two scalar-valued functions $f, g : \mathbb{R}^n \to \mathbb{K}$ is a scalar-valued function $f * g$ defined by

$$(f * g)(x) := \int_{\mathbb{R}^n} f(x - y) g(y) \, dy. \tag{5.1}$$

More generally, one can also allow one of f or g to take values in the vector space V, so that $f * g$ also takes values in V; we will generally assume this in the following without further commentary. The domain of $f * g$ is the set of all points $x \in \mathbb{R}^n$ for which the integrand on the right hand side of (5.1) is a Lebesgue-integrable function of y. It may happen that $(f * g)(x)$ is defined for some but not all $x \in \mathbb{R}^n$. In practice, we will only consider situations in which $(f * g)(x)$ is defined for *almost every* x; the function $f * g$ is then defined almost everywhere on $\mathbb{R}^n$. Since $f * g$ is defined via an integral, it does not change if either f or g is changed on a set of measure zero; it can therefore make sense to speak of the convolution $f * g$ of two elements $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$, and in such discussions we will typically not distinguish between actual functions and equivalence classes of functions defined almost everywhere.

Remark 5.5. In many situations, it can also make sense to define $f * g$ on a suitable subset of $\mathbb{R}^n$ for two functions f and g that are not defined everywhere on $\mathbb{R}^n$. One case that often arises is when f is defined on some open subset $\Omega \subset \mathbb{R}^n$ and g is defined on $\mathbb{R}^n$ but has compact support in the r-ball $B_r \subset \mathbb{R}^n$ about the origin for some small $r > 0$. If x belongs to the set

$$\Omega_r := \left\{ x \in \Omega \;\middle|\; \operatorname{dist}(x, \mathbb{R}^n \backslash \Omega) \geqslant r \right\},$$

then either $x - y \in \Omega$ or $g(y) = 0$ holds for every $y \in \mathbb{R}^n$, thus one can make sense of the right hand side of (5.1) by interpreting the integrand to be 0 whenever $g(y) = 0$. The convolution $f * g$ is thus defined on all points of Ω_r for which this integrand (suitably interpreted) is integrable.

Exercise 5.6. Use a change of variables to prove $f * g = g * f$.

An important property of the convolution is that $f * g$ is an general at least as "nice" as the nicest function among f and g.[12] In particular, if either f or g is of class C^1, then Exercise 5.6 allows us to relabel the functions so that f is in C^1 without loss of generality, and we can then try to prove the formula

$$\partial_k (f * g)(x) = \frac{\partial}{\partial x_k} \int_{\mathbb{R}^n} f(x - y) g(y) \, dy = \int_{\mathbb{R}^n} \partial_k f(x - y) g(y) \, dy = (\partial_k f * g)(x).$$

[12]The technical term for this notion of "niceness" is *regularity*, e.g. proving regularity of a function typically means proving that it is differentiable or smooth etc.

This will be valid whenever f and g satisfy suitable conditions to apply Theorem 0.4 and justify differentiating under the integral sign—in practice it is often easy to verify these conditions, and importantly, they do not require g to be differentiable, nor even continuous. For example:

Theorem 5.7. *For any $f \in C_0^\infty(\mathbb{R}^n)$ and $g \in L^1_{\text{loc}}(\mathbb{R}^n)$, the function $f * g$ is smooth on $\mathbb{R}^n$, and for every multi-index α,*

$$\partial^\alpha (f * g) = (\partial^\alpha f) * g.$$

Proof. By assumption f is smooth and vanishes outside of a compact subset $K \subset \mathbb{R}^n$, which implies that f is bounded. For every $x \in \mathbb{R}^n$, the integrand $y \mapsto f(x-y)g(y)$ can then only be nontrivial on the compact subset $K_x := \{x - k \in \mathbb{R}^n \mid k \in K\}$, and g is integrable on this domain, implying that the whole integrand is integrable on $\mathbb{R}^n$ and $(f*g)(x)$ is therefore defined for every $x \in \mathbb{R}^n$.

The function $x \mapsto (f * g)(x)$ is now defined as a parameter-dependent integral, where in the integrand only $f(x-y)$ depends on the parameter x. The result thus follows from Theorem 0.4, since:

- The integrand is Lebesgue integrable for every $x \in \mathbb{R}^n$;
- The integrability is also "locally uniform" in the sense that to every $x_0 \in \mathbb{R}^n$, one can associate a neighborhood $\mathcal{U} \subset \mathbb{R}^n$ of x_0 and an integrable function that bounds the integrand from above for every $x \in \mathcal{U}$.
- The function $x \mapsto f(x-y)g(y)$ is smooth for every $y \in \mathbb{R}^n$ and has partial derivative with respect to x_j given by $x \mapsto \partial_j f(x-y)g(y)$, which is again a continuous function of x.

Theorem 0.4 now implies $\partial_j(f * g) = (\partial_j f) * g$, and the generalization to arbitrary multi-indices follows by induction. ☐

5.3. **Young's inequality.** The following result is an elegant application of Fubini's theorem and Hölder's inequality.[13]

Theorem 5.8. *For arbitrary functions $f \in L^1(\mathbb{R}^n)$ und $g \in L^p(\mathbb{R}^n)$ with $1 \leqslant p \leqslant \infty$, $f * g$ is defined almost everywhere on $\mathbb{R}^n$, belongs to $L^p(\mathbb{R}^n)$ and satisfies*

$$\|f * g\|_{L^p} \leqslant \|f\|_{L^1} \cdot \|g\|_{L^p}.$$

Proof. The case $p = \infty$ is an easy exercise, so consider the case $1 \leqslant p < \infty$. Let $q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$; then

$$|f(x-y)g(y)| = |f(x-y)|^{1/p}|g(y)| \cdot |f(x-y)|^{1/q},$$

and Hölder's inequality implies for every $x \in \mathbb{R}^n$,

$$\begin{aligned}
\varphi(x) &:= \int_{\mathbb{R}^n} |f(x-y)g(y)|\, dy \\
&\leqslant \left(\int_{\mathbb{R}^n} |f(x-y)| \cdot |g(y)|^p \, dy \right)^{1/p} \cdot \left(\int_{\mathbb{R}^n} |f(x-y)|\, dy \right)^{1/q} \\
&\leqslant \|f\|_{L^1}^{1/q} \left(\int_{\mathbb{R}^n} |f(x-y)| \cdot |g(y)|^p \, dy \right)^{1/p}.
\end{aligned}$$

Now apply Fubini's theorem for nonnegative measurable functions to

$$\mathbb{R}^n \times \mathbb{R}^n \to [0, \infty] : (x, y) \mapsto |f(x-y)| \cdot |g(y)|^p;$$

[13] For various more general forms of Young's inequality, see [Sal16, Theorem 7.33] or [LL01, 4.2].

it follows that φ^p is a measurable function and

$$
\begin{aligned}
\|\varphi\|_{L^p}^p &= \int_{\mathbb{R}^n} [\varphi(x)]^p \, dx = \int_{\mathbb{R}^n} \|f\|_{L^1}^{p/q} \left(\int_{\mathbb{R}^n} |f(x-y)| \cdot |g(y)|^p \, dy \right) dx \\
&= \|f\|_{L^1}^{p/q} \int_{\mathbb{R}^n \times \mathbb{R}^n} |f(x-y)| \cdot |g(y)|^p \, dx \, dy \\
&= \|f\|_{L^1}^{p/q} \int_{\mathbb{R}^n} |g(y)|^p \left(\int_{\mathbb{R}^n} |f(x-y)| \, dx \right) dy = \|f\|_{L^1}^{p/q+1} \cdot \|g\|_{L^p}^p \\
&= \|f\|_{L^1}^p \cdot \|g\|_{L^p}^p < \infty.
\end{aligned} \tag{5.2}
$$

The function φ^p must therefore satisfy $\varphi^p < \infty$ almost everywhere, implying that $\varphi < \infty$ also holds almost everywhere, from which it follows that the convoluation $f * g$ is defined almost everywhere.

As a further application of Fubini's theorem, one can show that $f * g$ is also a measurable function; in fact, the convolution of two Lebesgue-measurable functions is always Borel measurable. We'll skip the proof of this, though see [Sal16, Theorem 7.32(iii)]. Since $|f * g| \leqslant \varphi$, the estimate $\|f * g\|_{L^p} \leqslant \|f\|_{L^1} \cdot \|g\|_{L^p}$ now follows. ☐

Exercise 5.9. Prove as a corollary of Theorem 5.8 that the convolution defines a continuous bilinear operator

$$L^1(\mathbb{R}^n) \times L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n) : (f, g) \mapsto f * g.$$

5.4. **Approximate identities.** We can now prove Theorem 5.1, and in the process explain a useful general trick called **mollification**, by which non-smooth functions can be approximated by smooth ones. One of the motivating ideas in the background is that of the "Dirac δ-function,", a fictional function $\delta : \mathbb{R}^n \to \mathbb{R}$ that one imagines being defined by $\delta(x) = 0$ for $x \neq 0$ and $\delta(0) = \infty$ so that

$$\int_{\mathbb{R}^n} \varphi(x)\delta(x) \, dx = \varphi(0)$$

for all φ in some reasonable class of functions on $\mathbb{R}^n$. While δ cannot be defined as an actual function, it can easily be *approximated* by smooth functions—such an approximation is sometimes called a **mollifier**.

Definition 5.10. An **approximate identity** on $\mathbb{R}^n$ is a sequence of smooth functions $\rho_j : \mathbb{R}^n \to [0, \infty)$ such that for every smooth compactly supported function φ on $\mathbb{R}^n$,

$$\int_{\mathbb{R}^n} \varphi(x)\rho_j(x) \, dx \to \varphi(0) \quad \text{as} \quad j \to \infty.$$

The functions ρ_j in Definition 5.10 are not required to have compact support, and it will be important when we prove the Fourier inversion formula in §8.5 to be able to choose specific examples that are not compactly supported but have other nice properties. For applications involving the convolution, however, it is useful to impose the following stricter condition.

Definition 5.11. A sequence of functions ρ_j on $\mathbb{R}^n$ will be said to have **shrinking support** if for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that the support of ρ_j is contained in the ϵ-ball about $0 \in \mathbb{R}^n$ for every $j \geqslant N$.

Lemma 5.12. *A sequence of smooth functions $\rho_j : \mathbb{R}^n \to [0, \infty)$ with shrinking support is an approximate identity if and only if $\int_{\mathbb{R}^n} \rho_j \, dm \to 1$ as $j \to \infty$, and in this case, the condition in Definition 5.10 is also satisfied for all (not necessarily smooth or compactly supported) measurable functions φ on $\mathbb{R}^n$ that are continuous at the origin.*

Proof. Assume $\mathrm{supp}(\rho_j)$ is contained in the ball $B_{r_j} \subset \mathbb{R}^n$ of radius $r_j > 0$ for some sequence $r_j \to 0$. If ρ_j is an approximate identity, then we can choose $N \in \mathbb{N}$ and a smooth compactly supported function $\varphi : \mathbb{R}^n \to [0, 1]$ that equals 1 on B_{r_j} for all $j \geqslant N$, and write

$$\int_{\mathbb{R}^n} \rho_j \, dm = \int_{B_{r_j}} \rho_j \, dm = \int_{B_{r_j}} \varphi\rho_j \, dm = \int_{\mathbb{R}^n} \varphi\rho_j \, dm \longrightarrow \varphi(0) = 1 \quad \text{as} \quad j \to \infty.$$

Conversely, if $\int_{\mathbb{R}^n} \rho_j \, dm \to 1$, then for any function φ on $\mathbb{R}^n$ that is continuous at 0,

$$\left|\varphi(0) - \int_{\mathbb{R}^n} \varphi \, \rho_j \, dm\right| = \left|\varphi(0)\left(1 - \int_{\mathbb{R}^n} \rho_j \, dm\right) + \int_{\mathbb{R}^n} [\varphi(0) - \varphi(x)] \, \rho_j(x) \, dx\right|$$
$$\leqslant |\varphi(0)| \cdot \left|1 - \int_{\mathbb{R}^n} \rho_j \, dm\right| + \sup_{x \in B_{r_j}} |\varphi(0) - \varphi(x)| \int_{\mathbb{R}^n} \rho_j \, dm \to 0.$$

□

Example 5.13. Choose a smooth function $\rho : \mathbb{R}^n \to [0, \infty)$ with compact support in the unit ball B_1 such that $\int_{\mathbb{R}^n} \rho \, dm = 1$. For $j \in \mathbb{N}$, the functions $\rho_j : \mathbb{R}^n \to [0, \infty)$ defined by $\rho_j(x) := j^n \rho(jx)$ then satisfy $\int_{\mathbb{R}^n} \rho_j \, dm = 1$ and have compact support in $B_{1/j}$ for all j, so this sequence forms an approximate identity with shrinking support.

Theorem 5.14. *Fix an approximate identity ρ_j with shrinking support, and given $f \in L^p(\mathbb{R}^n)$ with $1 \leqslant p < \infty$, let $f_j := \rho_j * f = f * \rho_j$ for $j \in \mathbb{N}$, that is,*

$$f_j(x) := \int_{\mathbb{R}^n} f(x - y) \rho_j(y) \, dy. \tag{5.3}$$

Then:

1. *f_j is a smooth function on $\mathbb{R}^n$ for every $j \in \mathbb{N}$.*
2. *$\|f_j\|_{L^p} \leqslant C \|f\|_{L^p}$ for every $j \in \mathbb{N}$ and a constant $C > 0$, which may be assumed arbitrarily close to 1 for sufficiently large j.*
3. *f_j converges in $L^p(\mathbb{R}^n)$ to f as $j \to \infty$.*

Remark 5.15. The formula (5.3) can be interpreted as defining $f_j(x)$ to be a *weighted average* of the values of f in a neighborhood of x, where the size of the neighborhood becomes arbitrarily small as j becomes large. The latter follows from the assumption that ρ_j has shrinking support.

Remark 5.16. The motivation for the term “approximate identity” is that if the δ-function existed as an actual function, it would satisfy $\delta * f = f * \delta = f$ for all reasonable functions f, making it an identity element in the algebra defined via the convolution product. We will see in §10 that this notion can be made rigorous by interpreting δ as a so-called *generalized function*, or *distribution*.

Proof of Theorem 5.14. The first two statements in the theorem follow from Theorems 5.7 and 5.8 since, by Lemma 5.12, $\|\rho_j\|_{L^1} = \int_{\mathbb{R}^n} \rho_j \, dm \to 1$. Let us write

$$\operatorname{supp}(\rho_j) \subset B_{r_j} \qquad \text{and} \qquad \left|\int_{\mathbb{R}^n} \rho_j \, dm - 1\right| < \epsilon_j$$

for a pair of sequences $r_j, \epsilon_j > 0$ that converge to zero. For the third statement in the theorem, we first give a proof under the additional assumption that f is almost everywhere bounded and has compact support, i.e. assume there exists a constant $R > 0$ such that

$$\|f\|_{L^\infty} \leqslant R \quad \text{and} \quad f|_{\mathbb{R}^n \setminus B_R} \equiv 0. \tag{5.4}$$

Since $\|\rho_j\|_{L^1}$ is bounded, Young's inequality (Theorem 5.8) now implies that f_j satisfies a uniform L^∞-bound for all j, and since $\operatorname{supp}(\rho_j) \subset B_{r_j}$ with $r_j \to 0$, we can also assume for large j that f_j has compact support in B_{R+1}. It follows that f and f_j are in $L^1(\mathbb{R}^n)$, and we claim: $f_j \to f$ in $L^1(\mathbb{R}^n)$. To prove this, we use (5.3) and estimate

$$\begin{aligned} |f_j(x) - f(x)| &= \left|\int_{\mathbb{R}^n} [f(x - y) - f(x)] \, \rho_j(y) \, dy + f(x)\left(\int_{\mathbb{R}^n} \rho_j \, dm - 1\right)\right| \\ &\leqslant \int_{\mathbb{R}^n} |f(x - y) - f(x)| \, \rho_j(y) \, dy + \epsilon_j |f(x)|, \end{aligned} \tag{5.5}$$

so by Fubini's theorem,

$$\begin{aligned}\|f_j - f\|_{L^1} &\leqslant \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |f(x-y) - f(x)| \cdot \rho_j(y)\, dy \right) dx + \epsilon_j \|f\|_{L^1} \\ &= \int_{\mathbb{R}^n} \rho_j(y) \left(\int_{\mathbb{R}^n} |f(x-y) - f(x)|\, dx \right) dy + \epsilon_j \|f\|_{L^1} \\ &= \int_{B_{r_j}} \rho_j(y) \|\tau_{-y} f - f\|_{L^1}\, dy + \epsilon_j \|f\|_{L^1} \leqslant \sup_{y \in B_{r_j}} \|\tau_{-y} f - f\|_{L^1} \cdot \|\rho_j\|_{L^1} + \epsilon_j \|f\|_{L^1}.\end{aligned}$$

This goes to 0 as $j \to \infty$ since $\epsilon_j, r_j \to 0$ and (by Theorem 5.4), $y \mapsto \tau_y f$ is a continuous map $\mathbb{R}^n \to L^1(\mathbb{R}^n)$.

Having established $f_j \to f$ in L^1, we also know that f_j has a subsequence for which $|f_j - f|^p$ converges pointwise almost everywhere to 0, and $|f_j - f|^p$ is also uniformly bounded by a constant multiple of the characteristic function of B_{R+1}, which is integrable. The dominated convergence theorem then implies

$$\|f_j - f\|_{L^p}^p = \int_{\mathbb{R}^n} |f_j - f|^p\, dm \to \int_{\mathbb{R}^n} 0\, dm = 0.$$

This conclusion applies at first to a subsequence, but if f_j were not convergent to f in $L^p(\mathbb{R}^n)$, then we could now find a subsequence that stays a positive distance away from f in the L^p-norm, and the L^1-convergence would then give a contradiction via the argument above, thus we have actually proved the convergence $f_j \xrightarrow{L^p} f$.

Without the additional conditions (5.4), one can instead argue as follows: for a given function $f \in L^p(\mathbb{R}^n)$ and a constant $R > 0$, define

$$f^R(x) := \begin{cases} f(x) & \text{if } x \in B_R \text{ and } |f(x)| \leqslant R, \\ 0 & \text{otherwise.} \end{cases}$$

It is not hard to show that $\|f - f^R\|_{L^p}$ can be made arbitrarily small by choosing $R > 0$ sufficiently large. Then f^R satisfies the conditions (5.4) and can therefore be approximated arbitrarily well in the L^p-norm by $f_j^R := \rho_j * f^R$. By Young's inequality,

$$\|f_j - f_j^R\|_{L^p} = \|\rho_j * (f - f^R)\|_{L^p} \leqslant \|\rho_j\|_{L^1} \cdot \|f - f^R\|_{L^p}$$

can then also be made arbitrarily small, thus $\|f - f_j\|_{L^p}$ becomes arbitrarily small for j sufficiently large. □

While we are on this subject, we can prove a similar result on approximation of C^m-functions that will be useful when we talk about distributions in §10. The statement requires a slight expansion of the notion of C^m_{loc}-convergence defined in §0.3. Observe that if

$$\Omega_1 \subset \Omega_2 \subset \ldots \subset \bigcup_{j \in \mathbb{N}} \Omega_j = \Omega \subset \mathbb{R}^n$$

is a nested sequence of open subsets in $\mathbb{R}^n$, then every compact set $K \subset \Omega$ belongs to Ω_j for $j \in \mathbb{N}$ sufficiently large. A sequence of C^m-functions $f_j : \Omega_j \to V$ is said to be **convergent in** $C^m_{\text{loc}}(\Omega)$ to a function $f : \Omega \to V$ if for every compact subset $K \subset \Omega$ and $N \in \mathbb{N}$ such that $K \subset \Omega_N$, the sequence of functions $f_N, f_{N+1}, f_{N+2}, \ldots$ restricted to K is C^m-convergent to $f|_K$. The ony difference between this and the definition in §0.3 is that the limit function f may be defined on a strictly larger domain than any function in the sequence.

Theorem 5.17. *Suppose $\Omega \subset \mathbb{R}^n$ is an open subset, $f \in C^m(\Omega)$ for some integer $m \geqslant 0$, and $\rho_j : \mathbb{R}^n \to [0, \infty)$ for $j \in \mathbb{N}$ is an approximate identity with shrinking support. Then there exists a nested sequence of open subsets $\Omega_1 \subset \Omega_2 \subset \ldots \subset \bigcup_{j \in \mathbb{N}} \Omega_j = \Omega$ such that for each $j \in \mathbb{N}$, $f_j := \rho_j * f$ is defined (in the sense of Remark 5.5) and smooth on Ω_j, and the sequence f_j converges to f in $C^m_{\text{loc}}(\Omega)$.*

Proof. Assume $\mathrm{supp}(\rho_j) \subset B_{r_j}$ with $r_j \to 0$, and define

$$\Omega_j := \left\{ x \in \Omega \;\middle|\; \mathrm{dist}(x, \mathbb{R}^n \backslash \Omega) > 2r_j \right\}.$$

Then $f_j(x) = \int_{\mathbb{R}^n} \rho_j(x-y) f(y)\, dy$ can be defined for all $x \in \Omega_j$ since $y \in \Omega$ whenever $x - y \in \mathrm{supp}(\rho_j)$. Smoothness follows by differentiating under the integral sign as in Theorem 5.7 to prove $\partial^\alpha f_j(x) = (\partial^\alpha \rho_j * f)(x)$ for all multi-indices α and $x \in \Omega_j$; here Theorem 0.4 is applicable because ρ_j is bounded and f is integrable on the region $B_{r_j}(x)$ where $\rho_j(x - \cdot)$ can be nonzero. To prove $f_j \to f$ in C^m_{loc}, suppose $K \subset \Omega$ is compact, and pick $N \in \mathbb{N}$ large enough so that $K \subset \Omega_N$ and the slightly larger compact set

$$K' := \left\{ x \in \mathbb{R}^n \;\middle|\; \mathrm{dist}(x, K) \leqslant r_j \right\}$$

is also contained in Ω. Then for $x \in K$ and $j \geqslant N$, (5.5) gives

$$|f_j(x) - f(x)| \leqslant \sup_{y \in B_{r_j}} |f(x-y) - f(x)| \cdot \|\rho_j\|_{L^1} + \epsilon_j \|f\|_{C^0(K)}.$$

Since x and $x - y$ in this expression both belong to K' and f is uniformly continuous on K', this implies uniform convergence $f_j \to f$ on K. To prove the same for derivatives up to order m, we observe that for any multi-index α with $|\alpha| \leqslant m$, $x \in K'$ and j sufficiently large,

$$\partial^\alpha f(x) = \frac{\partial^{|\alpha|}}{\partial x^\alpha} \int_{B_{r_j}} f(x-y)\rho_j(y)\, dy = \int_{B_{r_j}} \partial^\alpha f(x-y) \rho_j(y)\, dy = (\partial^\alpha f * \rho_j)(x),$$

where Theorem 0.4 justifies differentiation under the integral sign since $\partial^\alpha f$ is well defined and bounded on $B_{r_j}(x)$ while ρ_j is integrable. The same argument that was used for f_j then implies uniform convergence $\partial^\alpha f_j \to \partial^\alpha f$ on K. $\square$

6. Absolute continuity

6.1. **The fundamental theorem of calculus.** Let us consider the following question.

Question 6.1. *What is the largest class of functions f on a compact interval $[a, b] \subset \mathbb{R}$ such that the formula $f(x) = f(a) + \int_a^x f'(t)\, dt$ holds?*

Here we regard $\int_a^b f(t)\, dt$ as alternative notation for the Lebesgue integral $\int_{[a,b]} f\, dm$ if $a \leqslant b$, or $-\int_{[b,a]} f\, dm$ if $b \leqslant a$. The formula is easy to prove under the assumption that f is continuously differentiable, but we already know it is valid somewhat more generally than this, e.g. it clearly also holds if f is continuous and only piecewise C^1, and it is not hard to think up examples in which f is non-differentiable on a countably infinite subset but the formula still holds. In order for the right hand side to make sense at all, f only needs to be differentiable almost everywhere on $[a, b]$, and its (almost everywhere well-defined) derivative needs to be in $L^1([a, b])$. Is that enough? No:

Example 6.2. The **Cantor function** is a continuous, surjective and monotone increasing function $f : [0, 1] \to [0, 1]$ whose derivative is well defined and vanishes on a subset of full measure, namely the complement of the Cantor ternary set $C \subset [0, 1]$. In particular, f is defined to be constant on each of the intervals that are removed in order to define C:

$$f|_{(1/3,2/3)} := \frac{1}{2},$$
$$f|_{(1/9,2/9)} := \frac{1}{4}, \qquad f|_{(7/9,8/9)} := \frac{3}{4},$$
$$f|_{(1/27,2/27)} := \frac{1}{8}, \quad f|_{(7/27,8/27)} := \frac{3}{8}, \quad f|_{(19/27,20/27)} := \frac{5}{8}, \quad f|_{(25/27,26/27)} := \frac{7}{8}$$

and so forth (see Figure 4). The easiest way to define f at all other points is as the uniform limit of a sequence of piecewise affine, continuous, increasing and surjective functions $f_n : [0, 1] \to [0, 1]$. Such a sequence is uniquely determined by the following conditions (Figure 5):

- $f_0(x) := x$;
- For each $n \in \mathbb{N}$, f_n takes the same constant values as f on each of the 2^{n-1} intervals of length $1/3^n$ that are removed in the definition of C, and has constant slope on all other subintervals of $[0, 1]$.

It is easy to check from this definition that $|f_n - f_{n-1}| \leqslant c/2^n$ for some constant $c > 0$ and all $n \in \mathbb{N}$, thus the sequence f_n is uniformly Cauchy and therefore converges to a continuous function f, which is automatically monotone and surjective.[14]

Since the Cantor function has values on the entire interval $[0, 1]$ in spite of its derivative vanishing almost everywhere, it clearly lacks whatever property is needed for the fundamental theorem of calculus to hold. Let us reformulate the question slightly: suppose $f \in L^1([a, b])$, and consider the function F defined on $[a, b]$ by

$$F(x) := \int_a^x f(t)\, dt.$$

One of the main results of this section (Corollary 6.12 below) will show that F must be differentiable almost everywhere and its derivative is f. The Cantor function also has the first property, but since it is evidently not the integral of its derivative, we deduce that the Cantor function cannot be written as an integral of *any* Lebesgue-integrable function on $[0, 1]$. So, how do we tell the difference, i.e. what properties does the function F have that the Cantor function does not? Both are continuous, but it turns out that F satisfies a stronger condition than continuity.

[14] A more precise formula for f can be deduced from the fact that it is continuous and constant on a sequence of intervals whose union is dense. It is easiest to express in terms of base-3 and base-2 expansions: since all points $x \in C$ have unique base-3 expansions $0.a_1a_2a_3\ldots$ with $a_n \in \{0, 2\}$ for all $n = 1, 2, 3, \ldots$, one can write $f(x) \in [0, 1]$ so that its base-2 expansion is $0.b_1b_2b_3\ldots$ with $b_n := a_n/2$ for all n. In other words, $f\left(\sum_{n=1}^{\infty} \frac{2a_n}{3^n}\right) = \sum_{n=1}^{\infty} \frac{a_n}{2^n}$, assuming $a_n \in \{0, 1\}$ for all $n \in \mathbb{N}$.

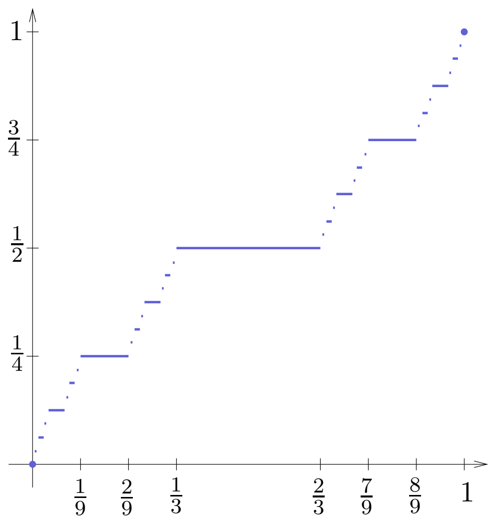

FIGURE 4. An imperfect picture of the Cantor function. Despite the appearance of jump discontinuities in the approximate graph drawn here, it is continuous.

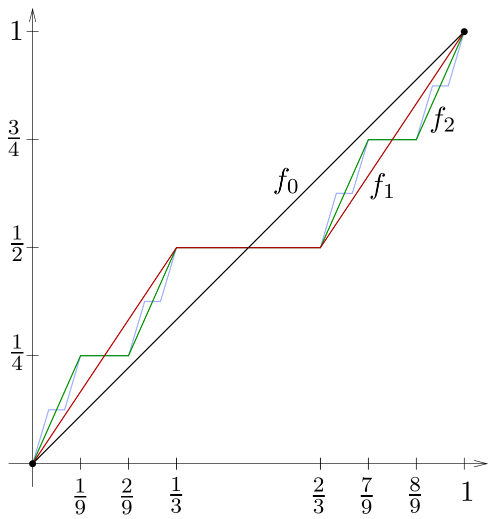

FIGURE 5. A sequence of piecewise affine functions converging uniformly to the Cantor function.

Lemma 6.3. *For any measure space (X, μ) and any $f \in L^1(X)$, given $\epsilon > 0$, there exists $\delta > 0$ such that for all measurable subsets $A \subset X$,*

$$\mu(A) < \delta \quad \Rightarrow \quad \int_A |f| \, d\mu < \epsilon.$$

Proof. If the result is not true, then there exists a number $\epsilon > 0$ and a sequence of measurable sets $A_n \subset X$ such that

$$\mu(A_n) < \frac{1}{2^n} \quad \text{but} \quad \int_{A_n} |f| \, d\mu \geqslant \epsilon.$$

Define $B_n := \bigcup_{k=n}^{\infty} A_k$, so we have

$$B_1 \supset B_2 \supset B_3 \supset \ldots B := \bigcap_{n \in \mathbb{N}} B_n,$$

and $\mu(B_n) \leqslant \sum_{k=n}^{\infty} \mu(A_k) < \sum_{k=n}^{\infty} \frac{1}{2^k} = \frac{1}{2^{n-1}}$, thus $\mu(B) = \lim_{n\to\infty} \mu(B_n) = 0$. This implies $\lim_{n\to\infty} \int_{B_n} |f| \, d\mu = \int_B |f| \, d\mu = 0$, which is a contradiction since $B_n \supset A_n$ for every n and thus $\int_{B_n} |f| \, d\mu \geqslant \int_{A_n} |f| \, d\mu \geqslant \epsilon > 0$. ∎

Returning to the function $F(x) = \int_a^x f(t)\,dt$ with $f \in L^1([a,b])$, consider the consequences of Lemma 6.3 for subsets $A \subset [a,b]$ defined as finite unions of intervals $A = \bigcup_{j=1}^N [a_j, b_j]$ with $a \leqslant a_1 \leqslant b_1 \leqslant \ldots \leqslant a_N \leqslant b_N \leqslant b$. The lemma provides for every $\epsilon > 0$ a $\delta > 0$ such that whenever $m(A) = \sum_{j=1}^N (b_j - a_j) < \delta$, it follows that

$$\sum_{j=1}^N |F(b_j) - F(a_j)| = \sum_{j=1}^N \left| \int_{a_j}^{b_j} f(t)\,dt \right| \leqslant \sum_{j=1}^N \int_{[a_j,b_j]} |f|\,dm = \int_A |f|\,dm < \epsilon.$$

In other words, F satisfies the following condition:

Definition 6.4. A function F on an interval $I \subset \mathbb{R}$ is called **absolutely continuous** if for every $\epsilon > 0$ there exists $\delta > 0$ such that for all finite sequences $a_1 \leqslant b_1 \leqslant \ldots \leqslant a_N \leqslant b_N$ of points in I,

$$\sum_{j=1}^N (b_j - a_j) < \delta \quad \Rightarrow \quad \sum_{j=1}^N |F(b_j) - F(a_j)| < \epsilon.$$

This definition would be the same as uniform continuity for functions on $I \subset \mathbb{R}$ if one only allowed $N = 1$, but the extension to all finite unions of intervals makes it a strictly stronger condition than uniform continuity. The Cantor function, for example, is uniformly continuous (as are all continuous functions on compact intervals), but the next exercise shows that it is not absolutely continuous:

Exercise 6.5. Show that if $F : [a,b] \to \mathbb{R}$ is absolutely continuous, then it maps every set of measure zero in $[a,b]$ to a set of measure zero in $\mathbb{R}$.

Exercise 6.6. Show that every Lipschitz continuous function on a compact interval $[a,b]$ is also absolutely continuous.

Here is the answer to Question 6.1:

Theorem 6.7 (Fundamental theorem of calculus for the Lebesgue integral)**.** *For a nontrivial compact interval $[a,b] \subset \mathbb{R}$ and functions f on $[a,b]$, the following conditions are equivalent:*

(1) f is absolutely continuous;
(2) f is differentiable almost everywhere, its derivative f' is in $L^1([a,b])$, and $f(x) = f(a) + \int_a^x f'(t)\,dt$ for all $x \in [a,b]$.

We have already proved the easy direction of this theorem, as a consequence of Lemma 6.3. We will show in Corollary 6.12 that for any $g \in L^1([a,b])$, the absolutely continuous function given by $F(x) = c + \int_a^x g(t)\,dt$ for a constant $c = F(a)$ is almost everywhere differentiable and its derivative is g. This statement is a consequence of the *Lebesgue differentiation theorem*, introduced in the next subsection. What then remains to be proved is that *every* absolutely continuous function f on $[a,b]$ can be written in the form $f(x) = f(a) + \int_a^x g(t)\,dt$ for some $g \in L^1([a,b])$. We will prove this in §6.5 as a consequence of a simple version of the *Radon-Nikodým theorem*, proved in §6.4.

Combining Exercise 6.6 with Theorem 6.7 produces a slightly surprising consequence:

Corollary 6.8. *Every Lipschitz continuous function on a compact interval $[a,b] \subset \mathbb{R}$ with $b > a$ is differentiable almost everywhere.* □

Corollary 6.8 also holds for functions on open domains in $\mathbb{R}^n$, and is known in that level of generality as *Rademacher's theorem*. For a concise proof built on top of the one-dimensional case, see [Hei05].

6.2. **The Lebesgue differentiation theorem.** Here is another natural question, which we will need to answer before we learn how to differentiate integrals of L^1-functions.

Question 6.9. *For locally integrable functions f on $\mathbb{R}^n$, what relation is there between $f(x)$ and the "average" value of f on arbitrarily small balls about x?*

Let us denote by

$$B_r(x) \subset \mathbb{R}^n$$

the open ball of radius $r > 0$ about a point $x \in \mathbb{R}^n$.

Definition 6.10. For a function $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, $x \in \mathbb{R}^n$ is called a **Lebesgue point** of f if the average value of $|f - f(x)|$ on $B_r(x)$ converges to zero as $r \to 0$, i.e.

$$\lim_{r \to 0^+} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f(y) - f(x)| \, dy = 0.$$

Whenever x is a Lebesgue point of f, one has

$$\lim_{r \to 0^+} \frac{1}{m(B_r(x))} \int_{B_r(x)} f(y) \, dy = f(x) \tag{6.1}$$

since

$$\begin{aligned} \left| \frac{1}{m(B_r(x))} \int_{B_r(x)} f \, dm - f(x) \right| &= \left| \frac{1}{m(B_r(x))} \int_{B_r(x)} (f - f(x)) \, dm \right| \\ &\leqslant \frac{1}{m(B_r(x))} \int_{B_r(x)} |f - f(x)| \, dm. \end{aligned}$$

Clearly x is a Lebesgue point whenever f is continuous at x, but Legesgue-integrable functions can easily be discontinuous everywhere. Moreover, changing f on a set of measure zero changes the right hand side of (6.1) at some points but not the left hand side, so the most one could hope for in general is for (6.1) to be true for almost every x. That turns out to be true, and thus gives the best possible answer to Question 6.9:

Theorem 6.11 (Lebesgue differentiation theorem)**.** *For any $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, almost every point of $\mathbb{R}^n$ is a Lebesgue point of f.*

To see why this is called a *differentiation* theorem, consider the case $n = 1$. If $f \in L^1([a, b])$, extend f to a function in $L^1(\mathbb{R})$ that vanishes outside $[a, b]$, and consider the function $F(x) := \int_a^x f(t) \, dt$. If $x \in (a, b)$ is a Lebesgue point of f, then for all $h > 0$ sufficiently small, we have

$$\begin{aligned} \left| \frac{F(x+h) - F(x)}{h} - f(x) \right| &= \left| \frac{1}{h} \int_x^h f(t) \, dt - f(x) \right| \leqslant \frac{1}{h} \int_x^h |f(t) - f(x)| \, dt \\ &\leqslant 2 \frac{1}{m((x-h, x+h))} \int_{x-h}^{x+h} |f(t) - f(x)| \, dt, \end{aligned}$$

and the latter becomes arbitrarily small when $h > 0$ is small. A similar statement is proved in the same manner for $h < 0$ and shows that at every Lebesgue point x, $F'(x) = f(x)$.

Corollary 6.12. *For every $f \in L^1([a, b])$, the function $F(x) := \int_a^x f(t) \, dt$ is differentiable almost everywhere on (a, b) and satisfies $F' = f$.* □

The proof of Theorem 6.11 requires a result called the *Hardy-Littlewood maximal inequality*, which we will discuss in the next subsection. In order to see what is needed, let us set up the general framework of the proof first.

We begin with two easy observations:

(1) If f is a continuous function on $\mathbb{R}^n$, then every point in $\mathbb{R}^n$ is a Lebesgue point.
(2) If almost every point is a Lebesgue point for all $f \in L^1(\mathbb{R}^n)$, then the same holds for all $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$.

The second statement follows from the purely local nature of the Lebesgue point condition, i.e. it depends on f only in arbitrarily small neighborhoods of x. Then if we cut off the values of $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ outside the ball $B_k(0) \subset \mathbb{R}^n$ to produce a function in $L^1(\mathbb{R}^n)$ whose set of non-Lebesgue points in $B_k(0)$ we can prove has measure zero, it follows that the set of non-Lebesgue points of f will be the union of these sets for all $k \in \mathbb{N}$, and thus also has measure zero.

With this understood, let us associate to any $f \in L^1(\mathbb{R}^n)$ and $r \geqslant 0$ the functions $f^r : \mathbb{R}^n \to [0, \infty]$ defined by

$$f^r(x) := \frac{1}{m(B_r(x))} \int_{B_r(x)} |f - f(x)| \, dm \quad \text{for } r > 0, \qquad f^0(x) := \limsup_{r \to 0} f^r(x).$$

The goal is to prove that $f^0 = 0$ almost everywhere. For each $N \in \mathbb{N}$, let

$$A_N := \left\{ x \in \mathbb{R}^n \;\middle|\; f^0(x) > 1/N \right\}.$$

We will deduce the desired result from:

Lemma 6.13. *For every $N \in \mathbb{N}$, A_N is contained in a Lebesgue-measurable set of measure less than $\frac{1}{N}$.*

Indeed, if this lemma holds, then since $A_1 \subset A_2 \subset A_3 \subset \ldots$, it follows that every A_N is a set of measure zero. Their union therefore also has measure zero, and that is precisely the set on which $f^0 > 0$.

In order to estimate the measure of A_N, we appeal to the density of continuous functions in $L^1(\mathbb{R}^n)$ and choose a sequence $f_1, f_2, f_3, \ldots$ of continuous functions on $\mathbb{R}^n$ such that $f_k \to f$ in L^1. We can then pick k large and use f_k to estimate $f^r(x)$ for $r > 0$ small:

$$\begin{aligned} f^r(x) &= \frac{1}{m(B_r(x))} \int_{B_r(x)} |f - f(x)| \, dm \\ &\leqslant \frac{1}{m(B_r(x))} \int_{B_r(x)} (|f - f_k| + |f_k - f_k(x)| + |f_k(x) - f(x)|) \, dm \\ &= \frac{1}{m(B_r(x))} \int_{B_r(x)} |f - f_k| \, dm + f_k^r(x) + |f_k(x) - f(x)|. \end{aligned} \tag{6.2}$$

We cannot assume $f_k \to f$ uniformly, so in this last expression, the third term might not become arbitrarily small for all x as $k \to 0$, but it is easy to show that it does so outside of a set of small measure. Indeed, we can associate to any given measurable function g on a measure space (X, μ) the sets $A_t := \left\{ x \in X \;\middle|\; |g(x)| > t \right\}$ for $t > 0$, and then estimate $\|g\|_{L^1} \geqslant \int_{A_t} |g| \, d\mu \geqslant \mu(A_t)t$. The result is known as **Chebyshev's inequality**:

$$\mu\big(\left\{ x \in X \;\middle|\; |g(x)| > t \right\} \big) \leqslant \frac{\|g\|_{L^1}}{t}. \tag{6.3}$$

Applying this to $f - f_k \in L^1(\mathbb{R}^n)$, we can arrange by choosing k sufficiently large to make $|f_k(x) - f(x)|$ arbitrarily small for all x outside of a set that has arbitrarily small measure. Having chosen k in this way, the second term in the last line of (6.2) also becomes arbitrarily small as $r \to 0$. However, estimating the first term requires some non-obvious input: we would like to claim that since $|f - f_k|$ has a small L^1-norm, its average value over $B_r(x)$ also satisfies some small bound as $r \to 0$. If we were first fixing $r > 0$ and then letting $k \to \infty$, it would be obvious that this term vanishes in the limit, but unfortunately the order of quantifiers is the other way around: we have already fixed k and need to estimate the average for all small $r > 0$ in terms of $\|f - f_k\|_{L^1}$. This is what the Hardy-Littlewood maximal inequality is for.

6.3. **Maximal functions and weak L^1.** We now introduce a missing ingredient in the proof of Lemma 6.13.

Definition 6.14. For $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, the **maximal function** $Mf : \mathbb{R}^n \to [0, \infty]$ is defined by

$$Mf(x) := \sup_{r > 0} \frac{1}{m(B_r(x))} \int_{B_r(x)} |f| \, dm.$$

Lemma 6.15. *For every $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, the maximal function $Mf : \mathbb{R}^n \to [0, \infty]$ is Borel measurable.*

Proof. It suffices to show that $(Mf)^{-1}((t,\infty))$ is a Borel set for every $t \in \mathbb{R}$; we will show in fact that it is open, i.e. Mf is lower semicontinuous. The condition $Mf(x) > t$ implies that for some $r > 0$, $\frac{1}{m(B_r(x))}\int_{B_r(x)} |f|\,dm > t$. This remains true after replacing $m(B_r(x))$ with $m(B_{r'}(x))$ for some slightly larger $r' > r$, and since $B_r(x) \subset B_{r'}(x')$ for all $x' \in \mathbb{R}^n$ sufficiently close to x, we then have

$$t < \frac{1}{m(B_{r'}(x))}\int_{B_r(x)} |f|\,dm \leqslant \frac{1}{m(B_{r'}(x))}\int_{B_{r'}(x)} |f|\,dm \leqslant Mf(x').$$

□

It is clearly not possible to achieve a general pointwise bound on Mf for $f \in L^1(\mathbb{R}^n)$, e.g. if f is defined as $1/\sqrt{|x|}$ on $[-1,1]$ and 0 on the rest of $\mathbb{R}$, then $f \in L^1(\mathbb{R})$ but its average values on $[-r,r]$ diverge to ∞ as $r \to \infty$, giving $Mf(0) = \infty$. A realistic hope, however, would be to bound the measure of sets on which Mf exceeds any given value. If $Mf \in L^1(\mathbb{R}^n)$, then such a bound follows from the Chebyshev inequality (6.3). In general, it would be too much to hope for Mf to be globally integrable, but there also exist functions that are not in $L^1(\mathbb{R}^n)$ and nonetheless satisfy a bound of the form (6.3), with the L^1-norm replaced by some other constant. A simple example is $f(x) := 1/x$, which is not in $L^1(\mathbb{R})$ but satisfies $m\big(\{x \in X \mid |f(x)| > t\}\big) = 2/t$, thus it belongs to the following class of functions.

Definition 6.16. A measurable function f on (X,μ) is called **weakly integrable** if there exists a constant $C > 0$ such that

$$\mu\big(\{x \in X \mid |f(x)| > t\}\big) \leqslant \frac{C}{t} \quad \text{for all} \quad t > 0.$$

We will denote the space of such functions by $L^1_{\text{weak}}(X)$.

One can define a "norm" on $L^1_{\text{weak}}(X)$ by

$$\|f\|_{L^1_{\text{weak}}} := \sup_{t>0} t\mu\big(\{x \in X \mid |f(x)| > t\}\big).$$

Just one caveat: $\|\cdot\|_{L^1_{\text{weak}}}$ satisfies $\|cf\|_{L^1_{\text{weak}}} = |c|\cdot\|f\|_{L^1_{\text{weak}}}$ for $c \in \mathbb{K}$ and $\|f\|_{L^1_{\text{weak}}} = 0$ if and only if f vanishes almost everywhere, but it does not satisfy the triangle inequality. Instead it satisfies (see [Sal16, Lemma 6.2])

$$\|f+g\|_{L^1_{\text{weak}}} \leqslant \frac{\|f\|_{L^1_{\text{weak}}}}{\lambda} + \frac{\|g\|_{L^1_{\text{weak}}}}{1-\lambda} \quad \text{for} \quad 0 < \lambda < 1,$$

and

$$\sqrt{\|f+g\|_{L^1_{\text{weak}}}} \leqslant \sqrt{\|f\|_{L^1_{\text{weak}}}} + \sqrt{\|g\|_{L^1_{\text{weak}}}}.$$

As a consequence, $L^1_{\text{weak}}(X)$ is not a normed vector space, but one can regard it as a topological vector space with respect to the metric $\operatorname{dist}(f,g) := \sqrt{\|f-g\|_{L^1_{\text{weak}}}}$. The inequality (6.3) can now be interpreted as saying that there is a natural continuous inclusion $L^1(X) \hookrightarrow L^1_{\text{weak}}(X)$.

Theorem 6.17 (Hardy-Littlewood)**.** *There exists a constant $C > 0$ depending only on the dimension n such that the estimate $\|Mf\|_{L^1_{\text{weak}}} \leqslant C\|f\|_{L^1}$ holds for all $f \in L^1(\mathbb{R}^n)$.*

The proof requires a simple version of a result known as the *Vitali covering lemma.*

Lemma 6.18 (Vitali)**.** *For every finite collection of open balls $B_{r_1}(x_1),\dots,B_{r_N}(x_N) \subset \mathbb{R}^n$, there exists a subset $I \subset \{1,\dots,N\}$ such that $B_{r_i}(x_i) \cap B_{r_j}(x_j) = \varnothing$ for every $i,j \in I$ with $i \ne j$, and*

$$\bigcup_{i=1}^N B_{r_i}(x_i) \subset \bigcup_{j\in I} B_{3r_j}(x_j).$$

In particular, $m\left(\bigcup_{i=1}^N B_{r_i}(x_i)\right) \leqslant 3^n \sum_{j\in I} m(B_{r_j}(x_j)).$

Proof. Abbreviate $B_i := B_{r_i}(x_i)$, and reorder the balls so that, without loss of generality, $r_1 \geqslant \ldots \geqslant r_N$. Define $I = \{i_1, \ldots, i_\ell\} \subset \{1, \ldots, N\}$ such that $i_1 := 1$ and, for each $j \geqslant 1$, i_{j+1} is the smallest number greater than i_j such that $B_{i_{j+1}}$ is disjoint from $B_{i_1} \cup \ldots \cup B_{i_j}$. This process terminates after finitely many steps, and if $k \notin I$, it means that B_k intersects B_i for some $i \in I$ with $i < k$. Since $r_k \leqslant r_i$, it follows that $B_k \subset B_{3r_i}(x_i)$. □

Proof of Theorem 6.17. We shall prove that the stated inequality holds with $C = 3^n$: in other words, for every $f \in L^1(\mathbb{R}^n)$ and $t > 0$,

$$m(A_t) \leqslant \frac{3^n \cdot \|f\|_{L^1}}{t}, \qquad \text{where} \qquad A_t := \left\{x \in \mathbb{R}^n \;\middle|\; Mf(x) > t\right\}. \tag{6.4}$$

By the inner regularity of the Lebesgue measure (see [Sal16, Theorem 2.13]), it will suffice to prove that every compact set $K \subset \mathbb{R}^n$ with $Mf|_K > t$ satisfies $m(K) \leqslant \frac{3^n \cdot \|f\|_{L^1}}{t}$. For each $x \in K$, the condition $Mf(x) > t$ means there exists a ball $B(x) \subset \mathbb{R}^n$ about x such that

$$\frac{1}{m(B(x))} \int_{B(x)} |f|\, dm > t. \tag{6.5}$$

Using the compactness of K, choose a finite subcollection $B_1, \ldots, B_N$ of such balls so that $K \subset \bigcup_{j=1}^N B_j$. After reordering the balls, we can then apply Lemma 6.18 to assume that $B_1, \ldots, B_\ell$ (for some $\ell \leqslant N$) are all disjoint and, using (6.5),

$$m(K) \leqslant m\left(\bigcup_{j=1}^N B_j\right) \leqslant 3^n \sum_{j=1}^{\ell} m(B_j) < 3^n \sum_{j=1}^{\ell} \frac{1}{t} \int_{B_j} |f|\, dm \leqslant \frac{3^n}{t} \|f\|_{L^1}.$$

□

We now have enough tools to complete the proof of the Lebesgue differentiation theorem.

Proof of Lemma 6.13 (and thus Theorem 6.11). The estimate (6.2) implies

$$f^r(x) \leqslant M(f - f_k)(x) + f_k^r(x) + |f_k(x) - f(x)|. \tag{6.6}$$

Given $N \in \mathbb{N}$, choose k large enough so that

$$\|f - f_k\|_{L^1} < \frac{1}{3^n \cdot 4N^2}.$$

Then Chebyshev's inequality (6.3) implies

$$m\left(\left\{x \in \mathbb{R}^n \;\middle|\; |f_k(x) - f(x)| > 1/2N\right\}\right) \leqslant 2N \|f - f_k\|_{L^1} < \frac{1}{3^n \cdot 2N} < \frac{1}{2N},$$

and by Theorem 6.17 (in particular (6.4)),

$$m\left(\left\{x \in \mathbb{R}^n \;\middle|\; M(f - f_k)(x) > 1/2N\right\}\right) \leqslant 2N \cdot 3^n \|f - f_k\|_{L^1} < \frac{1}{2N},$$

thus for all $x \in \mathbb{R}^n$ outside a set of measure at most $1/N$, (6.6) becomes

$$f^r(x) \leqslant \frac{1}{2N} + f_k^r(x) + \frac{1}{2N} = f_k^r(x) + \frac{1}{N}.$$

Letting r go to 0, we conclude $f^0(x) \leqslant \frac{1}{N}$ since $f_k^0(x) = 0$ by the continuity of f_k. □

6.4. **The Radon-Nikodým theorem.** Recall that if (X, μ) is a measure space and $f : X \to [0, \infty]$ a measurable function, then one can define another measure λ on the same σ-algebra by

$$\lambda(A) := \int_A f\, d\mu. \tag{6.7}$$

In order to show that absolutely continuous functions can always be written as integrals, we will first answer the following question, which turns out to be easier:

Question 6.19. *Given two measures μ and λ defined on the same measurable space X, does there exist a measurable function $f : X \to [0, \infty]$ such that (6.7) holds for all measurable sets A?*

The function $f : X \to [0,\infty]$ in this relation, if it exists, is sometimes called the **Radon-Nikodým derivative** of λ with respect to μ, and written as

$$\frac{d\lambda}{d\mu} := f.$$

It is not hard to think up necessary conditions for the existence of such a function. For example, there clearly is no such function if μ is the counting measure on $\mathbb{R}$ and λ is the Lebesgue measure, as $\int_A f\,d\mu = \sum_{x\in A} f(x)$ then can only be finite when $A \subset \mathbb{R}$ is a countable set, whose Lebesgue measure is therefore zero. It is also impossible if one takes the Lebesgue measure on $\mathbb{R}^n$ as μ and the **Dirac measure**

$$\delta(A) := \begin{cases} 1 & \text{if } 0 \in A, \\ 0 & \text{otherwise} \end{cases}$$

as λ; this is just the statement that the "Dirac δ-function" so popular among physicists does not actually exist. One thing both of these counterexamples have in common is that one can find measurable sets $A \subset X$ for which $\mu(A) = 0$ but $\lambda(A) \neq 0$. This possibility clearly needs to be excluded since $\int_A f\,d\mu = 0$ for every function when $\mu(A) = 0$.

Definition 6.20. Given a measure space (X,μ), a measure λ defined on the same σ-algebra is called **absolutely continuous** with respect to μ (written "$\lambda \ll \mu$") if the implication

$$\mu(A) = 0 \quad \Rightarrow \quad \lambda(A) = 0$$

holds for all measurable sets $A \subset X$.

The following exercise is not logically necessary for our exposition, but it demonstrates that there are nontrivial connections between Definition 6.20 and the notion of absolutely continuous functions.

Exercise 6.21 (cf. Lemma 6.3)**.** Show that if $\lambda \ll \mu$ and $\lambda(X) < \infty$, then for every $\epsilon > 0$ there exists a $\delta > 0$ such that for measurable sets $A \subset X$,

$$\mu(A) < \delta \quad \Rightarrow \quad \lambda(A) < \epsilon.$$

Theorem 6.22 (Radon-Nikodým)**.** *If μ and λ are two σ-finite measures on the same measurable space X, then the following conditions are equivalent:*

(1) $\lambda \ll \mu$;
(2) There exists a measurable function $f : X \to [0,\infty]$ satisfying $\lambda(A) = \int_A f\,d\mu$ for all measurable sets $A \subset X$.

The implication (2) $\Rightarrow$ (1) is immediate. We will prove (1) $\Rightarrow$ (2) using the natural isomorphism between L^∞ and the dual space of L^1. To see how this arises, note that if λ is given by (6.7), then for every real-valued λ-integrable function $g \in L^1(X,\lambda)$,

$$\int_X g\,d\lambda = \int_X gf\,d\mu. \tag{6.8}$$

The non-obvious trick is to view $\Lambda(g) := \int_X g\,d\lambda \in \mathbb{R}$ as defining a bounded linear functional $\Lambda : L^1(X,\lambda+\mu) \to \mathbb{R}$, which makes sense since $(\lambda+\mu)(A) := \lambda(A) + \mu(A)$ defines yet another measure on the same σ-algebra as λ and μ, and we have

$$|\Lambda(g)| = \left|\int_X g\,d\lambda\right| \leqslant \int_X |g|\,d\lambda \leqslant \int_X |g|\,d(\lambda+\mu) = \|g\|_{L^1(\lambda+\mu)}. \tag{6.9}$$

Since $\lambda + \mu$ is σ-finite, it follows then from the Riesz representation theorem that there exists a real-valued function $h \in L^\infty(X,\lambda+\mu)$ with $\|h\|_{L^\infty(\lambda+\mu)} \leqslant 1$ such that

$$\int_X g\,d\lambda = \int_X hg\,d(\lambda+\mu) \qquad \text{for all} \qquad g \in L^1(X,\lambda+\mu). \tag{6.10}$$

This is enough information to derive a formula for f in terms of h: indeed, combining (6.8) and (6.10) gives

$$\int_X gf\,d\mu = \int_X hg\,d(\lambda+\mu) = \int_X hg\,d\lambda + \int_X hg\,d\mu = \int_X (hgf + hg)\,d\mu = \int_X gh(f+1)\,d\mu$$

for all $g \in L^1(X,\lambda+\mu) \subset L^1(X,\lambda)$. This suggests the relation $f = h(f+1)$, or equivalently

$$f = \frac{h}{1-h}.$$

There are a few subtle issues to check before we can call this a proof—you may notice for instance that we have not yet used the condition $\lambda \ll \mu$.

Proof of Theorem 6.22. Following the trick described above, we note that since λ and μ are both σ-finite, $\lambda+\mu$ is also a σ-finite measure, so that the Riesz representation theorem (Theorem 2.4) gives a natural isomorphism between $L^\infty(X,\lambda+\mu)$ and the dual space of $L^1(X,\lambda+\mu)$. The bounded linear functional $\Lambda : L^1(X,\lambda+\mu) \to \mathbb{R}$ defined by $\Lambda(g) := \int_X f\,d\lambda$ therefore gives rise to a unique (up to equality almost everywhere) real-valued function $h \in L^\infty(X,\lambda+\mu)$ satisfying (6.10).

We claim that h satisfies $0 \leqslant h < 1$ almost everywhere with respect to the measure μ. Indeed, for $n \in \mathbb{N}$, let $A_n := \{x \in X \mid h \leqslant -1/n\}$, and suppose $A_n' \subset A_n$ is any subset for which $\lambda(A_n') + \mu(A_n') < \infty$. The function $g := \chi_{A_n'}$ is then in $L^1(X,\lambda+\mu)$, so plugging it into (6.10) gives

$$0 \leqslant \lambda(A_n') = \int_X g\,d\lambda = \int_X hg\,d(\lambda+\mu) = \int_{A_n'} h\,d(\lambda+\mu) \leqslant -\frac{1}{n}\left[\lambda(A_n') + \mu(A_n')\right] \leqslant 0,$$

implying $\mu(A_n') = \lambda(A_n') = 0$. Since λ and μ are both σ-finite, A_n is a union of countably many subsets on which λ and μ are both finite, so having shown that μ and λ vanish on all of these subsets, it follows that $\mu(A_n) = \lambda(A_n) = 0$. The set on which $h < 0$ is now the countable union of the sets A_n for $n \in \mathbb{N}$, and therefore also has measure zero with respect to both μ and λ. The other bound follows similarly by setting $A := \{x \in X \mid h(x) \geqslant 1\}$: for any subset $A' \subset A$ with $\lambda(A')$ and $\mu(A')$ both finite, we can plug $g := \chi_{A'}$ into (6.10) and find

$$\lambda(A') = \int_X g\,d\lambda = \int_X hg\,d(\lambda+\mu) = \int_{A'} h\,d\lambda + \int_{A'} h\,d\mu \geqslant \lambda(A') + \mu(A'),$$

implying $\mu(A') = 0$. (Notice that this time, we do not immediately also obtain $\lambda(A') = 0$; the latter follows since $\lambda \ll \mu$, but it need not be true without the absolute continuity assumption.) Appealing once more to the σ-finiteness of λ and μ, this implies $\mu(A) = 0$.

The function $f := \frac{h}{1-h}$ therefore satisfies $0 \leqslant f < \infty$ almost everywhere with respect to μ, so we can define a measure μ_f by

$$\mu_f(A) := \int_X f\,d\mu.$$

We claim $\mu_f = \lambda$. To see this, let us rewrite the relation (6.10) in the form

$$\int_X (1-h)g\,d\lambda = \int_X hg\,d\mu \quad \text{for all} \quad g \in L^1(X,\lambda+\mu).$$

Now if $A \subset X$ is any measurable subset for which the function $g := \frac{1}{1-h}\chi_A$ is in $L^1(X,\lambda+\mu)$, we obtain

$$\lambda(A) = \int_X \chi_A\,d\lambda = \int_X (1-h)g\,d\lambda = \int_X hg\,d\mu = \int_X \frac{h}{1-h}\chi_A\,d\mu = \int_A f\,d\mu = \mu_f(A).$$

To extend this to an arbitrary measurable subset $A \subset X$, we can again appeal to σ-finiteness and write

$$X = \bigcup_{n\in\mathbb{N}} X_n = \bigcup_{n\in\mathbb{N}} Y_n$$

for two sequences of subsets $X_1 \subset X_2 \subset \ldots \subset X$ and $Y_1 \subset Y_2 \subset \ldots X$ with $\mu(X_n) < \infty$ and $\lambda(Y_n) < \infty$. For $n \in \mathbb{N}$, let

$$A_n := \{x \in A \mid 1 - h(x) \geqslant 1/n\} \cap X_n \cap Y_n \subset X.$$

Then A_n has finite $(\mu + \lambda)$-measure and $\frac{1}{1-h} \leqslant n$ on A_n, thus $\frac{1}{1-h}\chi_{A_n} \in L^1(X, \lambda + \mu)$, so the calculation above proves $\lambda(A_n) = \mu_f(A_n)$. To finish, observe that since $h < 1$ almost everywhere with respect to μ, absolute continuity $\lambda \ll \mu$ implies that this is also true with respect to λ, and thus

$$\lambda\Big(A \backslash \bigcup_{n \in \mathbb{N}} A_n\Big) = \mu_f\Big(A \backslash \bigcup_{n \in \mathbb{N}} A_n\Big) = 0.$$

This justifies the following limit computation:

$$\lambda(A) = \lim_{n \to \infty} \lambda(A_n) = \lim_{n \to \infty} \mu_f(A_n) = \mu_f(A).$$

□

Remark 6.23. Without the condition $\lambda \ll \mu$, the function h constructed in the proof above may satisfy $h = 1$ on a set with positive λ-measure, in which case $\lim_{n\to\infty} \lambda(A_n) \leqslant \lambda(A)$ in the last step, producing an inequality

$$\int_A f \, d\mu \leqslant \lambda(A)$$

which may in general be strict. The argument still proves that equality holds for every measurable set $A \subset X$ such that the function $\frac{1}{1-h}\chi(A) \in L^1(X, \lambda + \mu)$, but in pathological examples, there may be no interesting sets with this property.

Exercise 6.24. Find the function $f : \mathbb{R}^n \to [0, \infty]$ that is constructed in the proof of Theorem 6.22 for the case where $\mu := m$ is the Lebesgue measure and $\lambda := \delta$ the Dirac measure. On which sets $A \subset \mathbb{R}^n$ is equality achieved in $\int_A f \, dm \leqslant \delta(A)$?

Remark 6.25. There are more general versions of the Radon-Nikdým theorem for so-called *signed measures* and *complex measures*, in which f in the formula $\lambda(A) = \int_A f \, d\mu$ may be a real or complex-valued μ-integrable function. See for example [Rud87, Chapter 6] or [Sal16, §5.4].

6.5. **Absolutely continuous functions are integrals.** In light of Corollary 6.12, the hard direction of the fundamental theorem of calculus for the Lebesgue integral now follows from:

Lemma 6.26. *Every absolutely continuous function F on $[a, b] \subset \mathbb{R}$ is given by*

$$F(x) = F(a) + \int_a^x f(t) \, dt \tag{6.11}$$

for some $f \in L^1([a, b])$.

The lemma is valid for functions $F : [a, b] \to V$ with values in an arbitrary finite-dimensional vector space, but we will focus on the case $V = \mathbb{R}$, which immediately implies the general case after choosing a real basis of V. For real-valued functions, we will deduce it from the Radon-Nikodým theorem. Some alternative approaches and interesting related facts are outlined in §6.5.3.

6.5.1. *The case of strictly increasing functions.* Let us assume $F : [a, b] \to \mathbb{R}$ is absolutely continuous and strictly increasing, i.e. it satisfies

$$F(y) > F(x) \quad \text{whenever} \quad y > x.$$

The following lemma produces another connection between the notions of absolute continuity for functions and measures.

Lemma 6.27. *F maps every set of measure zero in $[a, b]$ to a set of measure zero in $\mathbb{R}$.*

Proof. If $A \subset [a,b]$ has measure zero, then for any given $\delta > 0$, A is contained in the union of a sequence of disjoint intervals (a_i, b_i) such that $\sum_{i=1}^{\infty}(b_i - a_i) < \delta$. Absolute continuity guarantees that if $\epsilon > 0$ is given, $\delta > 0$ in the previous sentence can be chosen so that for every $k \in \mathbb{N}$, $\sum_{i=1}^{k} |f(b_i) - f(a_i)| < \epsilon$, and consequently,

$$\sum_{i=1}^{\infty} |f(b_i) - f(a_i)| \leqslant \epsilon.$$

The image $F(A)$ is therefore contained in a countable union of open intervals $(F(a_i), F(b_i))$ whose total measure is at most ϵ. □

Exercise 6.28. Find a set of measure zero whose image under the Cantor function of Example 6.2 has positive measure. (Note that in Lemma 6.27, we did not actually need to assume that F is *strictly* increasing.)

Proof of Lemma 6.26 for F strictly increasing. We claim that the formula

$$\lambda(A) := m(F(A))$$

defines a measure on $[a,b]$ with $\lambda \ll m$.

We need to check first that the image under F of every Lebesgue measurable set $A \subset [a,b]$ is Lebesgue measurable. By the inner regularity of the Lebesgue measure (see [Sal16, Theorem 2.13]), we can choose a sequence of compact subsets $K_1 \subset K_2 \subset K_3 \subset \ldots \subset A$ such that A is the disjoint union of a set A_0 of measure zero with $K_\infty := \bigcup_{n \in \mathbb{N}} K_n$. Since F is continuous, $F(K_n) \subset \mathbb{R}$ is also compact for each n, which makes $F(K_\infty)$ a countable union of compact sets and thus a Borel set. By Lemma 6.27, $F(A_0)$ is another set of measure zero, thus $F(A) = F(A_0) \cup F(K_\infty)$ is Lebesgue measurable.

Since $F : [a,b] \to \mathbb{R}$ is strictly increasing, it is also injective, so disjoint measurable subsets $A_1, A_2, A_3, \ldots \subset [a,b]$ have disjoint images, and it follows that λ as defined above is σ-additive. It is clearly also finite since F has bounded image, so this proves that λ is a measure, and Lemma 6.27 implies that it is absolutely continuous with respect to the Lebesgue measure.

The Radon-Nikodým theorem now provides a measurable function $f : [a,b] \to [0,\infty]$ such that for all measurable subsets $A \subset [a,b]$,

$$m(F(A)) = \int_A f\, dm.$$

In particular for $A := [a,x]$, this gives $F(x) - F(a) = \int_a^x f(t)\, dt$, and the global integrability of f follows from this by setting $x = b$. □

6.5.2. *The general case.* If $F : [a,b] \to \mathbb{R}$ is increasing but not strictly, then there is a cheap trick to reduce Lemma 6.26 to the strictly increasing case: we consider the function

$$G(x) := x + F(x),$$

which *is* strictly increasing (and also absolutely continuous), even if F is constant on some subinterval. The strictly increasing case therefore provides a function $g \in L^1([a,b])$ such that $G(x) = G(a) + \int_a^x g(t)\, dt$, and it follows that

$$F(x) = F(a) + \int_a^x [g(t) - 1]\, dt.$$

The proof for increasing functions is thus complete.

The conclusion of the proof now follows from an important general observation about all functions of *bounded variation*, which includes the absolutely continuous functions. Given any function f on $[a,b]$, we define the **total variation** of f by

$$TV(f) := \sup\left\{ \sum_{i=1}^{N} |f(x_i) - f(x_{i-1})| \;\middle|\; N \geqslant 1 \text{ and } a = x_0 < x_1 < \ldots < x_N = b \right\} \in [0,\infty],$$

and say that f is of **bounded variation** if $TV(f) < \infty$. Notice that if f is of bounded variation, then its restriction to every compact subinterval of $[a,b]$ is also of bounded variation.

Lemma 6.29. *If f is absolutely continuous on $[a,b]$, then it is of bounded variation, and the function $Vf : [a,b] \to [0,\infty)$ defined by $Vf(x) := TV(f|_{[a,x]})$ is also absolutely continuous.*

Proof. We note first that if $a \leqslant x < c < y \leqslant b$, then $|f(y) - f(x)| \leqslant |f(y) - f(c)| + |f(c) - f(x)|$, thus if $a \leqslant x_0 < x_1 < \ldots < x_N = b$ is any partition of the interval $[a,b]$ that does not include $c \in (a,b)$, adding c to the partition can only increase the value of the sum in the definition of $TV(f)$. It follows that we lose no generality if we modify the definition of $TV(f)$ so that the supremum ranges only over partitions that include c, which gives rise to the relation

$$TV(f) = TV(f|_{[a,c]}) + TV(f|_{[c,b]}).$$

In particular, this implies

$$Vf(y) = Vf(x) + TV(f|_{[x,y]}) \quad \text{for every } y > x \text{ in } [a,b], \tag{6.12}$$

so the function $Vf : [a,b] \to [0,\infty]$ is increasing. Now choose $\epsilon > 0$ and $\delta > 0$ as in the definition of absolute continuity, and choose a partition $a = t_0 < t_1 < \ldots < t_N = b$ such that $t_i - t_{i-1} < \delta$ for every i. Any partition of $[t_{i-1}, t_i]$ is then a finite collection of closures of disjoint open intervals with total length less than δ, implying $TV(f|_{[t_{i-1},t_i]}) < \epsilon$ and thus

$$TV(f) = \sum_{i=1}^{N} TV(f|_{[t_{i-1},t_i]}) < N\epsilon < \infty.$$

Keeping the same ϵ and δ, if $a \leqslant a_1 < b_1 < \ldots < a_n < b_n \leqslant b$ satisfy $\sum_{i=1}^n (b_i - a_i) < \delta$, then (6.12) implies

$$\sum_{i=1}^{n} |Vf(b_i) - Vf(a_i)| = \sum_{i=1}^{n} TV(f|_{[a_i,b_i]}).$$

The latter is the supremum of sums $\sum_j |f(t_j) - f(t_{j-1})|$ over finite collections of intervals $[t_{j-1}, t_j]$ with disjoint interiors whose lengths add up to $\sum_{i=1}^n (b_i - a_i) < \delta$, hence the sum is less than ϵ. ☐

Lemma 6.30. *For any function $f : [a,b] \to \mathbb{R}$ of bounded variation, the functions Vf, $Vf + f$ and $Vf - f$ on $[a,b]$ are increasing.*

Proof. That Vf is increasing follows already from (6.12), and in fact for $y > x$,

$$Vf(y) - Vf(x) = TV(f|_{[x,y]}) \geqslant |f(y) - f(x)|$$

since $x < y$ is a particular example of a partition of $[x,y]$. It follows that $Vf(y) - Vf(x) \geqslant f(y) - f(x)$ and $Vf(y) - Vf(x) \geqslant f(x) - f(y)$, thus

$$Vf(y) - f(y) \geqslant Vf(x) - f(x) \quad \text{and} \quad Vf(y) + f(y) \geqslant Vf(x) + f(x).$$

☐

Conclusion of the proof of Lemma 6.26. An arbitrary absolutely continuous function $F : [a,b] \to \mathbb{R}$ can be decomposed as

$$F = \frac{1}{2}(VF + F) - \frac{1}{2}(VF - F),$$

where by Lemmas 6.29 and 6.30, $VF + F$ and $VF - F$ are each absolutely continuous and increasing. We have already proved therefore that both can be represented as integrals of L^1-functions on $[a,b]$, and the same thus follows for F. ☐

6.5.3. *Alternative approaches.* There are other ways of proving Lemma 6.26 without using the Radon-Nikodým theorem. Since we expect an absolutely continuous function F to be the integral of its derivative, one method for finding the function f in (6.11) is to prove directly that F is differentiable almost everywhere. This can be deduced from the following somewhat surprising classical result of Lebesgue:

Theorem 6.31 (Lebesgue)**.** *Every monotone function $f : [a,b] \to \mathbb{R}$ is differentiable almost everywhere.*

A proof of this theorem using a more elaborate version of the Vitali covering lemma (cf. Lemma 6.18) may be found in [Roy88, Chapter 5]; see also [RSN90] for a slightly different exposition. It takes only slightly more effort to see that for a monotone function $f : [a,b] \to \mathbb{R}$, f' belongs to $L^1([a,b])$. The argument goes as follows: consider the case where $f : [a,b] \to \mathbb{R}$ is increasing, and for convenience, extend f over $\mathbb{R}$ with constant values $f(x) = f(a)$ for $x \leqslant a$ and $f(x) = f(b)$ for $x \geqslant b$. The **difference quotients**

$$D_h f(x) := \frac{f(x+h) - f(x)}{h}$$

are then well-defined functions $D_h f : [a,b] \to \mathbb{R}$ for every $h \in \mathbb{R}\backslash\{0\}$, and they are clearly measurable functions since f is measurable. By Theorem 6.31, there exists a function $f' : [a,b] \to \mathbb{R}$ which can be defined as $f'(x) = \lim_{h\to 0} D_h f(x)$ wherever this limit exists and zero everywhere else; this means $D_h f \to f'$ almost everywhere as $h \to 0$, thus f' is measurable. The difference quotients are easily seen to satisfy a "discrete" variant of the fundamental theorem of calculus: instead of $\int_a^b f'(t)\,dt = f(b) - f(b)$, one has

$$\begin{aligned}\int_{[a,b]} D_h f\,dm &= \frac{1}{h}\int_a^b [f(x+h) - f(x)]\,dx = \frac{1}{h}\int_{a+h}^{b+h} f(x)\,dx - \frac{1}{h}\int_a^b f(x)\,dx \\ &= \frac{1}{h}\int_{[b,b+h]} f\,dm - \frac{1}{h}\int_{[a,a+h]} f\,dm.\end{aligned} \tag{6.13}$$

Taking $h > 0$, in the situation at hand we have defined f to be constant on $[b, b+h]$, and $-f$ is bounded above by $-f(a)$ on $[a, a+h]$, so this computation implies

$$\int_{[a,b]} D_h f\,dm \leqslant f(b) - f(a) \qquad \text{for } h > 0.$$

Now if we consider the sequence of functions $D_{1/n} f$ for $n \in \mathbb{N}$, they have nonnegative values since f is increasing, and they converge almost everywhere to f', thus Fatou's lemma (see [Sal16, Theorem 1.41]) gives

$$\int_{[a,b]} f'\,dm = \int_{[a,b]} \liminf_{n\to\infty} D_{1/n} h\,dm \leqslant \liminf_{n\to\infty} \int_{[a,b]} D_{1/n} h\,dm \leqslant f(b) - f(a).$$

Corollary 6.32. *For every monotone function $f : [a,b] \to \mathbb{R}$, f' is measurable and satisfies*

$$\|f'\|_{L^1} \leqslant |f(b) - f(a)|.$$

□

Remark 6.33. Corollary 6.32 is analogous to the inequality in Remark 6.23 for a measure λ on $[a,b]$ that need not satisfy $\lambda \ll m$. It comes with the caveat that without an assumption of absolute continuity, the inequality may fail to carry any interesting information, e.g. the Cantor function (Example 6.2) shows that the left hand side can simply vanish, even when the function f is far from being constant.

Theorem 6.31 and Corollary 6.32 have an immediate consequence for the class of functions $f : [a,b] \to \mathbb{R}$ that can be written as the difference $f_+ - f_-$ between two increasing functions $f_\pm : [a,b] \to \mathbb{R}$. This is precisely the class of functions with bounded variation that we saw in §6.5.2, which includes all absolutely continuous functions, thus:

Corollary 6.34. *Every absolutely continuous function $f : [a,b] \to \mathbb{R}$ is differentiable almost everywhere and its derivative belongs to $L^1([a,b])$.* ☐

With this result in place, we can compare any given absolutely continuous function F on $[a,b]$ with the function $F_1(x) := \int_a^x F'(t)\,dt$. The latter is also absolutely continuous and differentiable almost everywhere, with $F_1' = F'$ by Corollary 6.12, so looking at $F - F_1$ reduces the problem to proving:

Lemma 6.35. *Every absolutely continuous function on $[a,b]$ whose derivative vanishes almost everywhere is constant.*

For a fairly short proof of this, based again on the Vitali covering lemma, see [Roy88, §5.4, Lemma 13].

From a functional-analytic perspective, there is a more interesting alternative argument to be found in the most recent edition [RF10] of Royden's classic textbook, based originally on the article [FH15]. It makes use of the following notion, which should seem natural in light of Lemma 6.3:

Definition 6.36. A collection $\mathcal{F}$ of integrable functions on a measure space (X,μ) is called **uniformly integrable** (or **equi-intebrable**) if for every $\epsilon > 0$, there exists $\delta > 0$ such that for all $f \in \mathcal{F}$ and measurable subsets $A \subset X$,

$$\mu(A) < \delta \quad \Rightarrow \quad \int_A |f|\,d\mu < \epsilon.$$

You should think of this definition as a close analogue of *equicontinuity*: the point is that the correspondence between ϵ and δ is not allowed to depend on the choice of the function $f \in \mathcal{F}$.

Example 6.37. If $\mathcal{F}$ is any collection of measurable functions f satisfying $|f| \leqslant g$ for some fixed $g \in L^1(X)$, then Lemma 6.3 implies that $\mathcal{F}$ is uniformly integrable.

The relevance of uniform integrability to this discussion arises from the following observation:

Lemma 6.38 ([FH15] or [RF10, §6.4])**.** *A continuous function f on $[a,b]$ is absolutely continuous if and only if its family of difference quotients $\{D_h f\}_{0<h\leqslant 1}$ is uniformly integrable on $[a,b]$.*

With this understood, one can now appeal to a useful generalization of the dominated convergence theorem:

Theorem 6.39 (Vitali's convergence theorem; see [RF10, §4.6])**.** *For a Lebesgue-measurable subset $X \subset \mathbb{R}$ with finite measure, if $\{f_n\}_{n\in\mathbb{N}}$ is a uniformly integrable collection of functions on X such that $f_n \to f$ pointwise almost everywhere, then $f \in L^1(X)$ and $\int_X f_n\,dm \to \int_X f\,dm$.*

Remark 6.40. The convergence theorem is also true on general finite measure spaces if one adds the condition $f \in L^1(X)$ to the hypotheses; in our situation, that version would also suffice in light of Corollary 6.34. There is also a version for spaces with infinite measure; see [RF10, §18.3].

Knowing that an absolutely continuous function f is differentiable almost everywhere, one can now deduce $\int_{[a,b]} f'\,dt = f(b) - f(a)$ as follows: setting $h = 1/n$ in (6.13) gives a sequence of relations whose left hand sides converge as $n \to \infty$ to $\int_{[a,b]} f'\,dt$ due to Lemma 6.38 and Vitali's convergence theorem. At the same time, the right hand sides converge to $f(b) - f(a)$ since f is continuous, so we are done.

7. Fourier series

7.1. **Fully periodic functions.** In this section we consider functions f on $\mathbb{R}^n$ that are 1-periodic in every variable, meaning that the relation

$$f(x_1, \dots, x_{j+1}, x_j + 1, x_{j+1}, \dots, x_n) = f(x_1, \dots, x_n)$$

holds for every $j = 1, \dots, n$. We shall refer to functions with this property as **fully periodic** functions. Some obvious examples include the trigonometric functions

$$\begin{aligned} &\sin(2\pi k x_j), \quad \text{for} \quad k = 1, 2, 3, \dots \text{ and } j = 1, \dots, n, \\ &\cos(2\pi k x_j), \quad \text{for} \quad k = 0, 1, 2, \dots \text{ and } j = 1, \dots, n, \end{aligned} \tag{7.1}$$

plus all products and linear combinations of these functions. The idea of a Fourier series is to express arbitrary fully periodic functions as (possibly infinite) sums of products of precisely these functions. Algebraically, it is much easier to work with complex exponentials than trigonometric functions, thus we shall allow all our fully periodic functions to take values in a *complex* vector space and, instead of writing them in terms of the functions in (7.1), try to express them as linear combinations of products of the functions

$$e^{2\pi i k x_j}, \quad \text{for} \quad k \in \mathbb{Z} \text{ and } j = 1, \dots, n. \tag{7.2}$$

Notice that an arbitrary finite product of such functions takes the form

$$\varphi_k(x) := e^{2\pi i k \cdot x}, \quad \text{for} \quad k \in \mathbb{Z}^n, \tag{7.3}$$

where $k \cdot x$ denotes the standard Euclidean inner product of two vectors $k, x \in \mathbb{R}^n$. We use this notation to distinguish the inner product on $\mathbb{R}^n$ from the *complex* inner product $\langle\ ,\ \rangle$ on the finite-dimensional vector space V in which our functions will take their values. For this discussion, we explicitly set

$$\mathbb{K} := \mathbb{C},$$

and since it will often be relevant, we remind the reader that the standing convention for the complex inner product on V is

$$\langle iv, w \rangle = -i \langle v, w \rangle, \qquad \langle v, iw \rangle = i \langle v, w \rangle.$$

The main theorem on Fourier series states that every function in a sufficiently reasonable class of fully periodic functions $f : \mathbb{R}^n \to V$ can be expressed as a convergent sum of the form

$$f(x) = \sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot x} \widehat{f}_k \tag{7.4}$$

for a unique set of coefficients $\widehat{f}_k \in V$, called the **Fourier coefficients** of f. The right hand side of (7.4) is called the **Fourier series** of f. Since complex exponentials are linear combinations of trigonometric functions, it is always possible (though often tiresome) to rewrite a Fourier series as a sum of products of the trigonometric functions appearing in (7.1); in particular, the Fourier series of a *real*-valued fully periodic function $f : \mathbb{R}^n \to \mathbb{R}$ can always be re-expressed as a *real*-linear combination of products of real-valued trigonometric functions, so that complex numbers need not be mentioned. In applications, the complex numbers typically have no intrinsic meaning but make calculations much easier.

7.2. **Function spaces on the torus and the lattice.** A fully periodic function on $\mathbb{R}^n$ can equivalently be regarded as a funtion on an n-dimensional torus $\mathbb{T}^n$, which is by definition an n-fold Cartesian product of circles. The most convenient definition of $\mathbb{T}^n$ for our purposes is as follows. The **lattice** $\mathbb{Z}^n \subset \mathbb{R}^n$ is a subgroup of $\mathbb{R}^n$ with respect to the operation of vector addition, and since $\mathbb{R}^n$ is an abelian group, the subgroup is normal. We define $\mathbb{T}^n$ to be the quotient group

$$\mathbb{T}^n := \mathbb{R}^n / \mathbb{Z}^n,$$

so in other words, elements of $\mathbb{T}^n$ are equivalence classes of vectors in $\mathbb{R}^n$, such that two vectors $x, y \in \mathbb{R}^n$ are in the same equivalence class if and only if $x - y \in \mathbb{Z}^n$. In the case $n = 1$, the map

$$\mathbb{R}/\mathbb{Z} \ni [t] \mapsto (\cos(2\pi t), \sin(2\pi t)) \in \mathbb{R}^2$$

gives a natural bijection between $\mathbb{T}^1$ and the unit circle in $\mathbb{R}^2$, which is also often denoted by $S^1 \subset \mathbb{R}^2$ since it is a "1-dimensional sphere". Through this bijection, one can identify $\mathbb{T}^n$ with the n-fold product of copies of S^1.

We can make $\mathbb{T}^n$ into a metric space by defining

$$d([x],[y]) := \inf_{(x,y) \in [x]\times[y]} |x - y| ,$$

where $|\cdot|$ denotes the standard Euclidean norm on $\mathbb{R}^n$. You should take a moment to convince yourself that this expression really defines a metric on $\mathbb{T}^n$. Moreover, the natural projection map

$$\pi : \mathbb{R}^n \to \mathbb{T}^n : x \mapsto [x]$$

is continuous with respect to this metric, and since $\mathbb{T}^n$ is the image under π of the compact subset $[0,1]^n \subset \mathbb{R}^n$, it follows that $\mathbb{T}^n$ is compact.

The Lebesgue measure m on $\mathbb{R}^n$ also determines a natural measure on $\mathbb{T}^n$. Let $\mathcal{L}(\mathbb{R}^n)$ denote the σ-algebra of Lebesgue-measurable subsets of $\mathbb{R}^n$, and define $\mathcal{L}(\mathbb{T}^n) \subset 2^{\mathbb{T}^n}$ to consist of all sets $A \subset \mathbb{T}^n$ with the property that $\pi^{-1}(A) \in \mathcal{L}(\mathbb{R}^n)$. In other words, $\mathcal{L}(\mathbb{T}^n)$ is the largest σ-algebra on $\mathbb{T}^n$ for which the projection map $\pi : \mathbb{R}^n \to \mathbb{T}^n$ is measurable. For $A \in \mathcal{L}(\mathbb{T}^n)$, we then define

$$m(A) := m(\pi^{-1}(A) \cap [0,1)^n) \geqslant 0.$$

It is straightforward to check that $(\mathbb{T}^n, \mathcal{L}(\mathbb{T}^n), m)$ by this definition is a measure space, and moreover, since $[0,1)^n \subset \mathbb{R}^n$ has finite Lebesgue measure, $m(\mathbb{T}^n)$ is finite; indeed, $m(\mathbb{T}^n) = 1$.

Exercise 7.1. Show that every fully periodic function $f : \mathbb{R}^n \to V$ corresponds to a unique function $F : \mathbb{T}^n \to V$ such that

$$f(x) = F([x]) \quad \text{for all} \quad x \in \mathbb{R}^n,$$

and conversely, every function $F : \mathbb{T}^n \to V$ determines a fully periodic function $f : \mathbb{R}^n \to V$ via this same relation. Show moreover that f is continuous/measurable if and only if F is continuous/measurable, respectively, and for an integrable function $F : \mathbb{T}^n \to V$,

$$\int_{\mathbb{T}^n} F(x)\,dx := \int_{\mathbb{T}^n} F\,dm = \int_{[0,1)^n} F \circ \pi\,dm.$$

Since $\mathbb{T}^n$ is now both a compact metric space and a finite measure space, Exercise 7.1 has the following useful consequences. First, every continuous fully periodic function is equivalent to a continuous function on a compact metric space, and is therefore *bounded.* Second, a function $f : \mathbb{T}^n \to V$ can be integrable even if $f \circ \pi : \mathbb{R}^n \to V$ is not, as it is only the integral of $|f \circ \pi|$ over the cube $[0,1)^n$ that needs to be finite. In fact, periodicity guarantees that fully periodic functions $f : \mathbb{R}^n \to V$ can *never* be Lebesgue integrable on $\mathbb{R}^n$ unless they vanish almost everywhere, but this only happens because the function $f : \mathbb{R}^n \to V$ contains too much redundant information. Integrating f instead over the finite measure space $\mathbb{T}^n$ circumvents this problem.

In the following, we will keep Exercise 7.1 in mind and typically blur the distinction between arbitrary functions $\mathbb{T}^n \to V$ and fully periodic functions $\mathbb{R}^n \to V$. We will also drop the equivalence classes from the notation and denote elements of $\mathbb{T}^n$ simply as vectors $x \in \mathbb{R}^n$ when there is no ambiguity; when this notational shortcut is used, it means that *any* representative $x \in \mathbb{R}^n$ of the given element in $\mathbb{T}^n$ may be chosen, and no important results will depend on this choice.

Notice that if a fully periodic function is differentiable, then its partial derivatives are also periodic functions, thus we can sensibly speak of differentiable functions on $\mathbb{T}^n$ and define the hierarchy of function spaces

$$C^0(\mathbb{T}^n) \supset C^1(\mathbb{T}^n) \supset C^2(\mathbb{T}^n) \supset \ldots \supset C^\infty(\mathbb{T}^n),$$

where for each $k = 0, 1, 2, \ldots, \infty$ we define $C^k(\mathbb{T}^n)$ to be the vector space of fully periodic functions $\mathbb{R}^n \to V$ that are k-times continuously differentiable. For $k < \infty$, these spaces all admit natural Banach space structures, of which only the case $k = 0$ will be especially important for our purposes: the norm on $C^0(\mathbb{T}^n)$ is defined by

$$\|f\|_{C^0} := \max_{x \in \mathbb{R}^n} |f(x)|,$$

where the existence of the maximum is guaranteed by the fact that $\mathbb{T}^n$ is compact. Similarly, for each $p \in [1, \infty]$ the measure on $\mathbb{T}^n$ gives rise to a Banach space of V-valued functions (defined almost everywhere),

$$L^p(\mathbb{T}^n) := L^p(\mathbb{T}^n, m), \qquad \|f\|_{L^p} := \begin{cases} \left(\int_{\mathbb{T}^n} |f|^p \, dm\right)^{1/p} & \text{for } p < \infty, \\ \operatorname{ess\,sup} |f| & \text{for } p = \infty. \end{cases}$$

Note that since $\mathbb{T}^n$ is the image of the compact and finite-measure subset $[0, 1)^n$ under the projection $\pi : \mathbb{R}^n \to \mathbb{T}^n$, a function $f : \mathbb{T}^n \to V$ will belong to $L^p(\mathbb{T}^n)$ if and only if $f \circ \pi : \mathbb{R}^n \to V$ is *locally* of class L^p on $\mathbb{R}^n$, i.e. its restriction to every compact subset must be of class L^p, but $f \circ \pi$ itself will not usually belong to $L^p(\mathbb{R}^n)$. The space $L^2(\mathbb{T}^n)$ has a natural complex inner product defined by

$$\langle f, g \rangle_{L^2} := \int_{\mathbb{T}^n} \langle f(x), g(x) \rangle \, dx,$$

which makes $L^2(\mathbb{T}^n)$ into a Hilbert space.

Since the continuous functions on $\mathbb{T}^n$ are bounded and $\mathbb{T}^n$ has finite measure, $C^0(\mathbb{T}^n)$ is a subspace of $L^p(\mathbb{T}^n)$ for every $p \in [1, \infty]$; so, therefore, is $C^\infty(\mathbb{T}^n)$. In fact:

Proposition 7.2. *For every $p \in [1, \infty)$, $C^\infty(\mathbb{T}^n)$ is a dense linear subspace of $L^p(\mathbb{T}^n)$.*

Proof. We shall deduce this from the result in §5 that $C^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$. Given $f \in L^p(\mathbb{T}^n)$, define $F : \mathbb{R}^n \to V$ by

$$F = f \circ \pi \text{ on } [0, 1)^n, \qquad F = 0 \text{ elsewhere},$$

where $\pi : \mathbb{R}^n \to \mathbb{T}^n$ is the quotient projection. Then $F \in L^p(\mathbb{R}^n)$, so for every $\epsilon > 0$, there exists a smooth function $F_\epsilon \in C^\infty(\mathbb{R}^n)$ with

$$\|F - F_\epsilon\|_{L^p} < \epsilon.$$

Given any $\delta > 0$, we can also choose a smooth function $\beta_\delta : \mathbb{R}^n \to [0, 1]$ that has compact support in $(0, 1)^n$ and satisfies

$$\beta_\delta \equiv 1 \quad \text{on} \quad [\delta, 1 - \delta]^n.$$

The function $\beta_\delta F_\epsilon : \mathbb{R}^n \to V$ is then smooth and has compact support in $(0, 1)^n$, so it gives rise to a uniquely determined fully periodic smooth function $G_\epsilon^\delta : \mathbb{R}^n \to V$ such that $G_\epsilon^\delta = \beta_\delta F_\epsilon$ on $[0, 1)^n$. Let $g_\epsilon^\delta : \mathbb{T}^n \to V$ denote the corresponding function on the n-torus such that $G_\epsilon^\delta = g_\epsilon^\delta \circ \pi$.

We claim that $\|f - g_\epsilon^\delta\|_{L^p}$ can be made arbitrarily small if ϵ and δ are each chosen sufficiently small. Indeed, abbreviate $Q := [0, 1)^n$ and $Q_\delta := [\delta, 1 - \delta]^n$. Then

$$\begin{aligned} \|f - g_\epsilon^\delta\|_{L^p}^p &= \int_{\mathbb{T}^n} |f - g_\epsilon^\delta|^p \, dm = \int_Q |F - G_\epsilon^\delta|^p \, dm \\ &= \int_{Q_\delta} |F - G_\epsilon^\delta|^p \, dm + \int_{Q \setminus Q_\delta} |F - G_\epsilon^\delta|^p \, dm. \end{aligned}$$

Since $G_\epsilon^\delta = \beta_\delta F_\epsilon = F_\epsilon$ on Q_δ, the first term in the second line is

$$\int_{Q_\delta} |F - G_\epsilon^\delta|^p \, dm = \int_{Q_\delta} |F - F_\epsilon|^p \, dm \leqslant \int_{\mathbb{R}^n} |F - F_\epsilon|^p \, dm = \|F - F_\epsilon\|_{L^p}^p < \epsilon^p,$$

which is made arbitrarily small by choosing $\epsilon > 0$ small. To estimate the other term in the second line, we can use the fact that β_δ takes values in $[0, 1]$ and write

$$|F - G^\delta_\epsilon| = |F - \beta_\delta F_\epsilon| = |F - F_\epsilon + F_\epsilon(1 - \beta_\delta)| \leqslant |F - F_\epsilon| + |F_\epsilon|,$$

hence by Minkowski's inequality,

$$\begin{aligned}\left(\int_{Q\backslash Q_\delta} |F - G^\delta_\epsilon|^p \, dm\right)^{1/p} &\leqslant \left(\int_{Q\backslash Q_\delta} (|F - F_\epsilon| + |F_\epsilon|)^p \, dm\right)^{1/p} \\ &\leqslant \left(\int_{Q\backslash Q_\delta} |F - F_\epsilon|^p \, dm\right)^{1/p} + \left(\int_{Q\backslash Q_\delta} |F_\epsilon|^p \, dm\right)^{1/p}.\end{aligned}$$

Here, the first term in the second line is bounded above by $\|F - F_\epsilon\|_{L^p} < \epsilon$, while if $\epsilon > 0$ is fixed, the second term can be made arbitrarily small for sufficiently small $\delta > 0$ since $|F_\epsilon|^p$ is Lebesgue integrable and $\bigcap_{\delta>0}(Q\backslash Q_\delta)$ is a set of measure zero, so that $\lim_{\delta\to 0} \int_{Q\backslash Q_\delta} |F_\epsilon|^p \, dm = 0$. This proves the claim. □

Since the Fourier coefficients of a function $f : \mathbb{T}^n \to V$ are meant to be a collection of vectors $\widehat{f}_k$ associated to elements $k \in \mathbb{Z}^n$, it will be useful to regard the collection of all these coefficients as a function

$$\widehat{f} : \mathbb{Z}^n \to V.$$

There is no meaningful notion of continuity or differentiability for such functions, but we *can* speak of L^p-spaces on $\mathbb{Z}^n$ with respect to the **counting measure**, i.e. let $\nu : 2^{\mathbb{Z}^n} \to [0, \infty]$ denote the measure such that $\nu(A)$ for each $A \subset \mathbb{Z}^n$ is the number of elements in A. The L^p-spaces with respect to this measure are conventionally denoted by

$$\ell^p(\mathbb{Z}^n) := L^p(\mathbb{Z}^n, \nu), \qquad 1 \leqslant p \leqslant \infty,$$

and since nonempty subsets in $\mathbb{Z}^n$ always have positive measure, the elements in these spaces are actual functions, not just equivalence classes of functions. The counting measure identifies integrals with infinite series and integrability with absolute summability, so for each $p \in [1, \infty)$, the ℓ^p-norm of a function $f : \mathbb{Z}^n \to V$ is

$$\|f\|_{\ell^p} := \left(\sum_{k\in\mathbb{Z}^n} |f(k)|^p\right)^{1/p},$$

while

$$\|f\|_{\ell^\infty} := \sup_{k\in\mathbb{Z}^n} |f(k)|.$$

There is one more space of functions $f : \mathbb{Z}^n \to V$ that we will need to consider, called the space of **rapidly decreasing** coefficients and denoted by $\mathscr{S}(\mathbb{Z}^n)$. A function $f : \mathbb{Z}^n \to V$ is defined to be in $\mathscr{S}(\mathbb{Z}^n)$ if and only if for every n-variable polynomial function $P : \mathbb{R}^n \to \mathbb{R}$, the function

$$\mathbb{Z}^n \to V : k \mapsto P(k)f(k)$$

is bounded. Equivalently, this means that for every $m \in \mathbb{N}$, the function $k \mapsto |k|^m f(k)$ is bounded on $\mathbb{Z}^n$. Since m in this expression can be chosen arbitrarily large, it is clear that functions $f(k)$ in $\mathscr{S}(\mathbb{Z}^n)$ always decay to 0 as $|k| \to \infty$. In fact:

Exercise 7.3. Show that $\mathscr{S}(\mathbb{Z}^n)$ is a dense linear subspace of $\ell^p(\mathbb{Z}^n)$ for every $p \in [1, \infty)$. *Hint: All functions $\mathbb{Z}^n \to V$ with bounded support are in $\mathscr{S}(\mathbb{Z}^n)$.*

7.3. **The transformations $\mathscr{F}$ and $\mathscr{F}^*$.** Suppose for the moment that $(V, \langle\ ,\ \rangle)$ is $\mathbb{C}$ with its standard Hermitian inner product. The functions $\varphi_k(x) := e^{2\pi i k \cdot x}$ defined in (7.3) for $k \in \mathbb{Z}^n$ can then be regarded as elements of $C^\infty(\mathbb{T}^n)$ since they are fully periodic and smooth. Since they are bounded and $\mathbb{T}^n$ has finite measure, they can also be regarded as elements of $L^2(\mathbb{T}^n)$, and as it turns out, they form an orthonormal set:

Exercise 7.4. Show that $\|\varphi_k\|_{L^2} = 1$ for every $k \in \mathbb{Z}^n$ and $\langle \varphi_k, \varphi_{k'} \rangle_{L^2} = 0$ for every $k \neq k' \in \mathbb{Z}^n$.

If we now assume $\hat{f}_k \in \mathbb{C}$ are coefficients such that the sum $\sum_{k \in \mathbb{Z}^n} \hat{f}_k \varphi_k$ converges in the L^2-norm to $f \in L^2(\mathbb{T}^n)$, then Exercise 7.4 makes it easy to compute the Fourier coefficients in terms of f: we have

$$\hat{f}_k = \sum_{p \in \mathbb{Z}^n} \langle \varphi_k, \hat{f}_p \varphi_p \rangle_{L^2} = \langle \varphi_k, f \rangle_{L^2} = \int_{\mathbb{T}^n} e^{-2\pi i k \cdot x} f(x)\, dx.$$

This computation generalizes in a straightforward way to functions valued in a general finite-dimensional complex inner product space $(V, \langle\ ,\ \rangle)$ if we engage in a slight abuse of notation: let us define

$$\langle \varphi, \psi v \rangle_{L^2} := \langle \varphi, \psi \rangle_{L^2} v \in V \qquad \text{for } \varphi, \psi : \mathbb{T}^n \to \mathbb{C},$$

which by linearity gives rise to a natural pairing $\langle \varphi, f \rangle \in V$ for any pair of L^2-functions $\varphi : \mathbb{T}^n \to \mathbb{C}$ and $f : \mathbb{T}^n \to V$. The computation of $\hat{f}_k$ above then becomes valid for vector-valued functions. We shall take this formula as a definition of a transformation $\mathscr{F}$, which sends functions $f : \mathbb{T}^n \to V$ to functions

$$\mathscr{F} f := \hat{f} : \mathbb{Z}^n \to V : k \mapsto \hat{f}_k$$

whenever the integral on the right hand side of the following expression is well defined for all k:

$$(\mathscr{F} f)_k := \hat{f}_k := \int_{\mathbb{T}^n} e^{-2\pi i k \cdot x} f(x)\, dx. \tag{7.5}$$

It is clear, for instance, that if $f \in L^1(\mathbb{T}^n)$, then all of the coefficients $\hat{f}_k$ are well defined and they satisfy a uniform bound

$$|\hat{f}_k| \leqslant \|f\|_{L^1},$$

hence $\mathscr{F}$ defines a bounded linear operator

$$\mathscr{F} : L^1(\mathbb{T}^n) \to \ell^\infty(\mathbb{Z}^n). \tag{7.6}$$

There is a similar transformation $\mathscr{F}^*$ that associates to a function $g : \mathbb{Z}^n \to V : k \mapsto g_k$ a function

$$\mathscr{F}^* g := \check{g} : \mathbb{T}^n \to V$$

defined by

$$(\mathscr{F}^* g)(x) := \check{g}(x) := \sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot x} g_k. \tag{7.7}$$

As with the definition of $\mathscr{F}$ in (7.5), this definition comes with the caveat that at first glance, it only makes sense if the sum converges for every x. So for instance, it makes sense whenever $g \in \ell^1(\mathbb{Z}^n)$, as the sum then converges absolutely and uniformly; since the partial sums are all finite sums of continuous functions, it follows in this case that $\check{g} : \mathbb{T}^n \to V$ is a continuous function and satisfies $|\check{g}(x)| \leqslant \sum_{k \in \mathbb{Z}^n} |g_k| = \|g\|_{\ell^1}$ for all $x \in \mathbb{T}^n$, thus $\mathscr{F}^*$ defines a bounded linear operator

$$\mathscr{F}^* : \ell^1(\mathbb{Z}^n) \to C^0(\mathbb{T}^n). \tag{7.8}$$

We have already seen that under certain circumstances, the operators $\mathscr{F}$ and $\mathscr{F}^*$ are inverse to each other, e.g. the computation following Exercise 7.4 above shows that if $g : \mathbb{Z}^n \to \mathbb{C}$ is a function such that $\mathcal{F}^* g \in L^2(\mathbb{T}^n)$ and the series in the definition of $\mathcal{F}^* g$ converges in the L^2-norm, then $\mathcal{F}\mathcal{F}^* g = g$. The next two theorems are the main results we need to prove about Fourier series.

Theorem 7.5. *The transformations $\mathscr{F}$ and $\mathscr{F}^*$ defined in* (7.5) *and* (7.7) *respectively have the following properties:*

1. *$\mathscr{F}$ maps $C^\infty(\mathbb{T}^n)$ bijectively onto $\mathscr{S}(\mathbb{Z}^n)$.*
2. *$\mathscr{F}^*$ maps $\mathscr{S}(\mathbb{Z}^n)$ bijectively onto $C^\infty(\mathbb{T}^n)$.*
3. *The bijections $\mathscr{F} : C^\infty(\mathbb{T}^n) \to \mathscr{S}(\mathbb{Z}^n)$ and $\mathscr{F}^* : \mathscr{S}(\mathbb{Z}^n) \to C^\infty(\mathbb{T}^n)$ are inverse to each other.*
4. *For every $f \in C^\infty(\mathbb{T}^n)$, the series $\sum_{k\in\mathbb{Z}^n} e^{2\pi i k\cdot x}\widehat{f}_k$ converges absolutely and uniformly with all derivatives to f.*

Theorem 7.6 (Parseval's identity)**.** *For every $f, g \in C^\infty(\mathbb{T}^n)$,*

$$\langle \widehat{f}, \widehat{g}\rangle_{\ell^2} = \langle f, g\rangle_{L^2}.$$

Since $C^\infty(\mathbb{T}^n)$ is dense in $L^2(\mathbb{T}^n)$, Parseval's identity gives rise to a unique bounded linear extension of the operator $\mathscr{F} : C^\infty(\mathbb{T}^n) \to \mathscr{S}(\mathbb{Z}^n)$ to an operator

$$\mathscr{F} : L^2(\mathbb{T}^n) \to \ell^2(\mathbb{Z}^n). \tag{7.9}$$

In other words, for each $f \in L^2(\mathbb{T}^n)$, we can choose an approximating sequence $f_j \in C^\infty(\mathbb{T}^n)$ with $f_j \xrightarrow{L^2} f$ as $j \to \infty$, and define $\widehat{f} = \mathscr{F} f \in \ell^2(\mathbb{Z}^n)$ as the ℓ^2-limit of the ℓ^2-Cauchy sequence $\widehat{f}_j \in \mathscr{S}(\mathbb{Z}^n)$. This description makes $\mathscr{F} : L^2(\mathbb{T}^n) \to \ell^2(\mathbb{Z}^n)$ sound more abstract than it really is: in fact, since $\mathbb{T}^n$ has finite measure, $L^2(\mathbb{T}^n)$ is a subspace of $L^1(\mathbb{T}^n)$, so the operator in (7.9) is just the restriction of $\mathscr{F} : L^1(\mathbb{T}^n) \to \ell^\infty(\mathbb{Z}^n)$ to this subspace. In the other direction, the density of $\mathscr{S}(\mathbb{Z}^n)$ in $\ell^2(\mathbb{Z}^n)$ implies that $\mathscr{F}^* : \mathscr{S}(\mathbb{Z}^n) \to C^\infty(\mathbb{T}^n)$ extends uniquely to an operator

$$\mathscr{F}^* : \ell^2(\mathbb{Z}^n) \to L^2(\mathbb{T}^n),$$

defined similarly by choosing for any $g \in \ell^2(\mathbb{Z}^n)$ an approximating sequence $g_j \in \mathscr{S}(\mathbb{Z}^n)$ with $g_j \xrightarrow{\ell^2} g$ and writing $\mathscr{F}^* g = \check{g} \in L^2(\mathbb{T}^n)$ for the L^2-limit of the L^2-Cauchy sequence $\check{g}_j \in C^\infty(\mathbb{T}^n)$. Here there is an obvious choice of approximating sequence g_j available, namely

$$g_j(k) := \begin{cases} g(k) & \text{if } |k| \leqslant j, \\ 0 & \text{otherwise.} \end{cases}$$

This makes $\check{g}_j \in C^\infty(\mathbb{T}^n)$ a sequence of partial sums for the Fourier series $\sum_{k\in\mathbb{Z}^n} e^{2\pi i k\cdot x} g(k)$, so the conclusion is that this series converges to $\check{g}$ in the L^2-norm. It clearly cannot be expected to converge uniformly since $\check{g} \in L^2(\mathbb{T}^n)$ is not generally continuous, and there is also no guarantee of pointwise convergence, not even almost everywhere. The compositions $\mathscr{F}^*\mathscr{F} : C^\infty(\mathbb{T}^n) \to C^\infty(\mathbb{T}^n)$ and $\mathscr{F}\mathscr{F}^* : \mathscr{S}(\mathbb{Z}^n) \to \mathscr{S}(\mathbb{Z}^n)$, of course, each extend to $L^2(\mathbb{T}^n) \to L^2(\mathbb{T}^n)$ and $\ell^2(\mathbb{Z}^n) \to \ell^2(\mathbb{Z}^n)$ respectively as the identity map. We summarize this discussion as follows.

Corollary 7.7. *The transformations $\mathscr{F}$ and $\mathscr{F}^*$ defined in* (7.5) *and* (7.7) *give well-defined unitary maps*[15] *$L^2(\mathbb{T}^n) \to \ell^2(\mathbb{Z}^n)$ and $\ell^2(\mathbb{Z}^n) \to L^2(\mathbb{T}^n)$ respectively, where in the latter case, the series should be interpreted as an L^2-convergent (but not necessarily pointwise convergent) series of functions in $L^2(\mathbb{T}^n)$. Moreover, these two transformations are inverse to each other.* □

In light of Exercise 7.4, another way to say this is as follows:

Corollary 7.8. *For any orthonormal basis $v_1, \dots, v_m$ of $(V, \langle\ ,\ \rangle)$, the functions*

$$\left\{ e^{2\pi i k\cdot x} v_j \right\}_{k\in\mathbb{Z}^n,\ j=1,\dots,m}$$

form an orthonormal basis of the Hilbert space $L^2(\mathbb{T}^n)$. □

The remainder of §7 is concerned with the proofs of Theorems 7.5 and 7.6, and along the way, we will prove some relations between the Fourier operations and differentiation which are frequently useful in applications.

[15] A linear map $T : \mathcal{H} \to \mathcal{H}'$ between two Hilbert spaces is called **unitary** if it is an isometry, i.e. $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in \mathcal{H}$. Such maps also satisfy $\|Tx\| = \|x\|$ for all $x \in \mathcal{H}$, hence they are continuous.

7.4. **Fourier series and derivatives.** If one ignores the words "bijectively" and "onto," then the first statement in Theorem 7.5 becomes an easy consequence of the following exercise.

Exercise 7.9. Use integration by parts to show that for every $f \in C^1(\mathbb{T}^n)$, $k = (k_1, \dots, k_n) \in \mathbb{Z}^n$ and $j = 1, \dots, n$,

$$\widehat{\partial_j f}_k = 2\pi i k_j \widehat{f}_k.$$

Recall from §0.3 that a **multi-index** in n variables is an n-tuple $\alpha = (\alpha_1, \dots, \alpha_n)$ of nonnegative integers, and we denote its **order** by $|\alpha| := \alpha_1 + \ldots + \alpha_n$. This gives rise to the differential operator

$$\partial^\alpha := \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n}$$

of order $|\alpha|$ for functions on $\mathbb{R}^n$, as well as a complex-valued polynomial function of $z = (z_1, \dots, z_n) \in \mathbb{C}^n$ defined by

$$z^\alpha := z_1^{\alpha_1} \dots z_n^{\alpha_n}.$$

For $f \in C^\infty(\mathbb{T}^n)$, repeating the formula in Exercise 7.9 finitely many times now yields

$$\widehat{\partial^\alpha f}_k = (2\pi i k)^\alpha \widehat{f}_k \tag{7.10}$$

for any multi-index α.

Proof of Theorem 7.5, part 1. Assume $f \in C^\infty(\mathbb{T}^n)$, and choose any multi-index α. Since $\partial^\alpha f$ is bounded and $\mathbb{T}^n$ has finite measure, $\partial^\alpha f$ also belongs to $L^1(\mathbb{T}^n)$, implying in light of (7.6) that $\widehat{\partial^\alpha f} \in \ell^\infty(\mathbb{Z}^n)$. The relation (7.10) then implies that

$$k^\alpha \widehat{f}_k = \frac{k^\alpha}{(2\pi i)^{|\alpha|} k^\alpha} \widehat{\partial^\alpha f}_k = \frac{1}{(2\pi i)^{|\alpha|}} \widehat{\partial^\alpha f}_k$$

is bounded independently of $k \in \mathbb{Z}^n$. Since this is true for every multi-index α, it follows that $k \mapsto P(k)\widehat{f}_k$ is a bounded function $\mathbb{Z}^n \to V$ for every polynomial P, hence $\widehat{f} \in \mathscr{S}(\mathbb{Z}^n)$.

We've proved:

$$\mathscr{F}(C^\infty(\mathbb{T}^n)) \subset \mathscr{S}(\mathbb{Z}^n).$$

□

The next exercise is an easy application of the standard theorem on term-by-term differentiation of infinite series—the point is that whenever $g \in \ell^1(\mathbb{Z}^n)$, the partial sums of the series $\sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot x} g(k)$ converge uniformly with respect to $x \in \mathbb{T}^n$.

Exercise 7.10. Given a function $g : \mathbb{Z}^n \to V$ and $j \in \{1, \dots, n\}$, let $g_j : \mathbb{Z}^n \to V$ denote the function defined by $g_j(k) := k_j g(k)$ for $k = (k_1, \dots, k_n) \in \mathbb{Z}^n$. Show that if g and g_j both belong to $\ell^1(\mathbb{Z}^n)$, then $\check{g} : \mathbb{T}^n \to V$ is continuous and has a continuous partial derivative $\partial_j \check{g} : \mathbb{T}^n \to V$ given by

$$\partial_j \check{g}(x) = \widetilde{2\pi i g_j}(x).$$

Proof of Theorem 7.5, part 2. We consider the second statement in the theorem: suppose $g \in \mathscr{S}(\mathbb{Z}^n)$. Then the function $k \mapsto k^\alpha g(k)$ also belongs to $\mathscr{S}(\mathbb{Z}^n)$ for every multi-index α, and is therefore in $\ell^1(\mathbb{Z}^n)$. Iterating the result of Exercise 7.10 finitely many times then proves that for every multi-index α, $\partial^\alpha \check{g}$ exists and is continuous and is given by

$$\partial^\alpha \check{g} = (2\pi i)^{|\alpha|} \widetilde{g_\alpha}, \tag{7.11}$$

where $g_\alpha : \mathbb{Z}^n \to V$ is given by $g_\alpha(k) := k^\alpha g(k)$. In particular, $\check{g} : \mathbb{T}^n \to V$ is smooth.

We've proved:

$$\mathscr{F}^*(\mathscr{S}(\mathbb{Z}^n)) \subset C^\infty(\mathbb{T}^n).$$

□

The main remaining step in the proof of Theorem 7.5 is to show that

$$\mathscr{F}\mathscr{F}^*|_{\mathscr{S}(\mathbb{Z}^n)} = \mathrm{Id}_{\mathscr{S}(\mathbb{Z}^n)} \qquad \text{and} \qquad \mathscr{F}^*\mathscr{F}|_{C^\infty(\mathbb{T}^n)} = \mathrm{Id}_{C^\infty(\mathbb{T}^n)}.$$

We have already proved the first relation, as a consequence of the L^2-orthonormality of the functions φ_k. We shall prove in §7.6 that the relation $\mathscr{F}^*\widehat{f} = f$ holds for $f \in C^\infty(\mathbb{T}^n)$. As preparation for the latter, we first need a quick digression on the topic of approximate identities.

7.5. Approximate identities.

In §5.4, we considered sequences of smooth functions $\rho_j : \mathbb{R}^n \to [0,\infty)$ that approximate the so-called "Dirac δ-function". In the context of fully periodic functions, the analogous object to $\delta : \mathbb{R}^n \to [0,\infty)$ would be a nonnegative function δ on $\mathbb{T}^n$ that satisfies

$$\int_{\mathbb{T}^n} \varphi(x)\delta(x)\,dx = \varphi(0) \quad \text{for all} \quad \varphi \in C^\infty(\mathbb{T}^n).$$

If such a function existed, it would need to be identically zero on $\mathbb{T}^n\backslash\{0\}$ and have an infinite value at 0, so δ cannot be understood as a function in the classical sense, though one can make sense of it as either a measure or a distribution (i.e. a "generalized function", see §10). What is perhaps more important in many applications is that one can *approximate* it with actual smooth functions.

Definition 7.11. An **approximate identity** on $\mathbb{T}^n$ is a sequence $\rho_j : \mathbb{T}^n \to [0,\infty)$ of nonnegative smooth functions such that for every $\varphi \in C^\infty(\mathbb{T}^n)$,

$$\lim_{j\to\infty} \int_{\mathbb{T}^n} \rho_j(x)\varphi(x)\,dx = \varphi(0).$$

Remark 7.12. The convolution of two functions on $\mathbb{T}^n$ is defined analogously to functions on $\mathbb{R}^n$, by

$$(f * g)(x) := \int_{\mathbb{T}^n} f(x-y)g(y)\,dy,$$

where $x - y \in \mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ makes sense for $x, y \in \mathbb{T}^n$ since the lattice $\mathbb{Z}^n$ is a subgroup of $\mathbb{R}^n$ with respect to vector addition. One can again use a change of variables to show $f * g = g * f$ (cf. Exercise 5.6). The defining property of an approximate identity thus implies that for any $f \in C^\infty(\mathbb{T}^n)$, $(\rho_j * f)(x) = (f * \rho_j)(x) = \int_{\mathbb{T}^n} f(x-y)\rho_j(y)\,dy \to f(x)$, so

$$\rho_j * f \to f \text{ pointwise for } f \in C^\infty(\mathbb{T}^n). \tag{7.12}$$

The term "approximate identity" refers to the ring structure on $L^1(\mathbb{T}^n)$ defined via the convolution operator. If a δ-function "$\delta := \lim_{j\to\infty} \rho_j$" existed, then it would satisfy $\delta * f = f * \delta = f$ for every smooth function f, thus it would define an identity element in the convolution ring.

The next result describes one of several simple tricks for finding examples of approximate identities.

Proposition 7.13. *Suppose $\rho : \mathbb{T}^n \to [0,\infty)$ is a smooth function satisfying $\rho(0) = 1$ and $\rho(x) < 1$ for all $x \neq 0 \in \mathbb{T}^n$, and for each $j \in \mathbb{N}$, let $c_j := \int_{\mathbb{T}^n} [\rho(x)]^j\,dx > 0$. Then the sequence $\rho_j : \mathbb{T}^n \to [0,\infty)$ defined by*

$$\rho_j(x) := \frac{1}{c_j}[\rho(x)]^j$$

is an approximate identity.

Proof. Let $B_\delta(0) \subset \mathbb{T}^n$ denote the open ball of radius $\delta > 0$ about $0 \in \mathbb{T}^n$. We claim that for every $\delta > 0$,

$$\int_{\mathbb{T}^n\backslash B_\delta(0)} \rho_j(x)\,dx \to 0$$

as $j \to \infty$. Indeed, $\rho < 1$ on the compact set $\mathbb{T}^n \backslash B_\delta(0)$, thus $\rho \leqslant b$ on this set for some constant $b \in (0,1)$. Choose $a \in (b,1)$: then since $\rho(0) = 1$, there also exists a $\delta' \in (0,\delta)$ such that $\rho \geqslant a$ on $B_{\delta'}(0)$. This implies

$$c_j = \int_{\mathbb{T}^n} [\rho(x)]^j \, dx \geqslant \int_{B_{\delta'}(0)} [\rho(x)]^j \, dx \geqslant a^j m(B_{\delta'}(0)),$$

thus

$$\begin{aligned} \int_{\mathbb{T}^n \backslash B_\delta(0)} \rho_j(x)\, dx &= \frac{1}{c_j} \int_{\mathbb{T}^n \backslash B_\delta(0)} [\rho(x)]^j \, dx \leqslant \frac{b^j m(\mathbb{T}^n \backslash B_\delta(0)}{a^j m(B_{\delta'}(0))} \\ &= \left(\frac{b}{a}\right)^j \frac{m(\mathbb{T}^n \backslash B_\delta(0)}{m(B_{\delta'}(0))} \to 0. \end{aligned}$$

Now given $f \in C^\infty(\mathbb{T}^n)$ and any $\epsilon > 0$, choose $\delta > 0$ such that $|f(x) - f(0)| < \epsilon$ for all $x \in B_\delta(0)$. Since $\int_{\mathbb{T}^n} \rho_j(x)\, dx = 1$ for all j by construction, we then have

$$\begin{aligned} \left| \int_{\mathbb{T}^n} \rho_j(x) f(x)\, dx - f(0) \right| &= \left| \int_{\mathbb{T}^n} \rho_j(x) \left[f(x) - f(0) \right] \, dx \right| \leqslant \int_{\mathbb{T}^n} \rho_j(x) \left| f(x) - f(0) \right| \, dx \\ &= \int_{B_\delta(0)} \rho_j(x) \left| f(x) - f(0) \right| \, dx + \int_{\mathbb{T}^n \backslash B_\delta(0)} \rho_j(x) \left| f(x) - f(0) \right| \, dx \\ &\leqslant \epsilon \int_{B_\delta(0)} \rho_j(x)\, dx + 2 \max_{x \in \mathbb{T}^n} |f(x)| \int_{\mathbb{T}^n \backslash B_\delta(0)} \rho_j(x)\, dx \\ &\leqslant \epsilon + 2 \max_{x \in \mathbb{T}^n} |f(x)| \int_{\mathbb{T}^n \backslash B_\delta(0)} \rho_j(x)\, dx \to \epsilon \quad \text{as } j \to \infty. \end{aligned}$$

Since $\epsilon > 0$ can be chosen arbitrarily small, this proves $\int_{\mathbb{T}^n} \rho_j(x) f(x)\, dx \to f(0)$. ☐

If the δ-function existed, its Fourier coefficients would have to be $\delta_k = \int_{\mathbb{T}^n} e^{-2\pi i k \cdot x} \delta(x)\, dx = 1$ for all $k \in \mathbb{Z}^n$, giving rise to the formal expression

$$\delta(x) = \sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot x}. \tag{7.13}$$

Both sides of this formula are nonsense mathematically, but it is worth remembering anyway, as it encapsulates two rigorously provable statements about Fourier series of approximate identities:

Lemma 7.14. *For any approximate identity $\rho_j : \mathbb{T}^n \to [0,\infty)$, the Fourier coefficients $(\widehat{\rho}_j)_k \in \mathbb{C}$ satisfy a uniform bound $|(\widehat{\rho}_j)_k| \leqslant C$ for some constant $C > 0$ independent of $j \in \mathbb{N}$ and $k \in \mathbb{Z}^n$, and $\lim_{j\to\infty} (\widehat{\rho}_j)_k = 1$ for all k.*

Proof. The convergence $(\widehat{\rho}_j)_k \to 1$ as $j \to \infty$ follows immediately from the formula $(\widehat{\rho}_j)_k = \int_{\mathbb{T}^n} e^{2\pi i k \cdot x} \rho_j(x)\, dx$ and the defining property of an approximate identity. In particular for $k = 0 \in \mathbb{Z}^n$, we have $\lim_{j\to\infty} (\widehat{\rho}_j)_0 = 1$, so there exists a bound $(\widehat{\rho}_j)_0 = \int_{\mathbb{T}^n} \rho_j(x)\, dx \leqslant C$ independent of j. Then

$$|(\widehat{\rho}_j)_k| \leqslant \int_{\mathbb{T}^n} \left| e^{2\pi i k \cdot x} \rho_j(x) \right| \, dx \leqslant \int_{\mathbb{T}^n} \rho_j(x)\, dx \leqslant C$$

holds for every $j \in \mathbb{N}$ and $k \in \mathbb{Z}^n$. ☐

Lemma 7.15. *There exists an approximate identity $\rho_j : \mathbb{T}^n \to [0,\infty)$ that is equal to its own Fourier series for every j, i.e. it satisfies $\mathscr{F}^* \mathscr{F} \rho_j = \rho_j$.*

Proof. Define $\beta : \mathbb{T}^1 \to [0,\infty)$ by $\beta(t) := \frac{\cos(2\pi t)+1}{2}$ and $\rho : \mathbb{T}^n \to [0,\infty)$ by

$$\rho(x_1, \ldots, x_n) := \beta(x_1) \ldots \beta(x_n),$$

and let ρ_j denote the approximate identity described by Proposition 7.13 in terms of this particular choice of ρ. Since β is a complex linear combination of $e^{2\pi i t}$ and $e^{-2\pi i t}$, ρ is a finite linear combination of functions from the orthonormal set $\{\varphi_k\}_{k \in \mathbb{Z}^n}$, and the same is therefore true of all its powers $[\rho(x)]^j$ for $j \in \mathbb{N}$. This proves that ρ_j is equal to a finite Fourier series for every j. $\square$

7.6. **Completeness.** We are now ready to prove that $\mathscr{F}^* \widehat{f} = f$ for every $f \in C^\infty(\mathbb{T}^n)$. Let us first describe an informal "physicist's version" of the proof, in which we refuse to worry about annoying analytical issues like integrability, convergence, and whether the δ-function actually exists. The main tool needed for this is Fubini's theorem, which we apply for functions on $\mathbb{T}^n \times \mathbb{Z}^n$ with the product of our Lebesgue-type measure m on $\mathbb{T}^n$ with the counting measure ν on $\mathbb{Z}^n$. For $f \in C^\infty(\mathbb{T}^n)$, we compute:

$$
\begin{aligned}
(\mathscr{F}^* \widehat{f})(x) &= \sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot x} \widehat{f}_k \, dx = \int_{\mathbb{Z}^n} e^{2\pi i k \cdot x} \left(\int_{\mathbb{T}^n} e^{-2\pi i k \cdot y} f(y) \, dy \right) d\nu(k) \\
&= \int_{\mathbb{T}^n \times \mathbb{Z}^n} e^{2\pi i k \cdot (x-y)} f(y) \, d(m(y) \otimes \nu(k)) = \int_{\mathbb{T}^n} \left(\sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot (x-y)} f(y) \right) dy \\
&= \int_{\mathbb{T}^n} \left(\sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot (x-y)} \right) f(y) \, dy = \int_{\mathbb{T}^n} \delta(x-y) f(y) \, dy = \int_{\mathbb{T}^n} f(x-y) \delta(y) \, dy \\
&= f(x).
\end{aligned}
$$

Several steps in this derivation are formal manipulations that cannot be taken literally. The interchange of the integral and the summation is meant to be a result of applying Fubini's theorem to the function $(y,k) \mapsto e^{2\pi i k \cdot (x-y)} f(y)$ on $\mathbb{T}^n \times \mathbb{Z}^n$, though unfortunately, the latter is not $(m \otimes \nu)$-integrable. The δ-function then appears due to (7.13), and from there we apply a straightforward change of variables followed by the defining property of the δ-function.

The way to make all this mathematically precise is by introducing the Fourier coefficients of an approximate identity ρ_j in the second line. This will make the function on $\mathbb{T}^n \times \mathbb{Z}^n$ integrable and thus produce a mathematically correct formula, which converges to the desired formula as $j \to \infty$.

Proof of Theorem 7.5, part 3. Assume $f \in C^\infty(\mathbb{T}^n)$ is given. By Lemma 7.15, we can choose an approximate identity $\rho_j : \mathbb{T}^n \to [0, \infty)$ that equals its own Fourier series for every j, and by Lemma 7.14, its Fourier coefficients are uniformly bounded and converge to 1. Since ρ_j is smooth, the function $k \mapsto (\widehat{\rho}_j)_k$ on $\mathbb{Z}^n$ belongs to $\mathscr{S}(\mathbb{Z}^n) \subset \ell^1(\mathbb{Z}^n)$, so that the function $F : \mathbb{T}^n \times \mathbb{Z}^n \to V$ given by

$$
F(y,k) := e^{2\pi i k \cdot x} e^{-2\pi i k \cdot y} (\widehat{\rho}_j)_k f(y)
$$

satisfies

$$
|F(y,k)| \leqslant |(\widehat{\rho}_j)_k| \cdot |f(y)|,
$$

and is therefore $(m \otimes \nu)$-integrable as a consequence of Fubini's theorem for nonnegative measurable functions. We can then apply Fubini's theorem for integrable functions, giving

$$
\sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot x} (\widehat{\rho}_j)_k \left(\int_{\mathbb{T}^n} e^{-2\pi i k \cdot y} f(y) \, dy \right) = \int_{\mathbb{T}^n} \left(\sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot (x-y)} (\widehat{\rho}_j)_k \right) f(y) \, dy.
$$

The left hand side of this expression is $\sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot x} (\widehat{\rho}_j)_k \widehat{f}_k$, and since $\widehat{f} \in \mathscr{S}(\mathbb{Z}^n) \subset \ell^1(\mathbb{Z}^n)$, Lemma 7.14 implies via the dominated convergence theorem (applied on $\mathbb{Z}^n$ with the counting measure) that this converges to $(\mathscr{F}^* \widehat{f})(x)$ as $j \to \infty$. Since each ρ_j is equal to its Fourier series, the right hand side is

$$
\int_{\mathbb{T}^n} \rho_j(x-y) f(y) \, dy = (\rho_j * f)(x),
$$

which converges in turn to $f(x)$ by (7.12).

We've proved:

$$
\mathscr{F}^* \mathscr{F} f = f \quad \text{for all} \quad f \in C^\infty(\mathbb{T}^n).
$$

$\square$

Proof of Theorem 7.5, conclusion. The first three statements in the theorem have already been established, so it only remains to verify that for $g \in \mathscr{S}(\mathbb{Z}^n)$, the Fourier series $\sum_{k\in\mathbb{Z}^n} e^{2\pi i k\cdot x} g(k)$ converges uniformly and so do its derivatives of all orders. The uniform convergence is clear since $\mathscr{S}(\mathbb{Z}^n) \subset \ell^1(\mathbb{Z}^n)$. Applying an arbitrary differential operator ∂^α to the terms of the series changes the coefficients to $(-2\pi i k)^\alpha g(k)$, and this function of k is still in $\mathscr{S}(\mathbb{Z}^n)$, so the resulting series also converges uniformly. ☐

7.7. **Parseval's identity.** The proof of Theorem 7.6 is based mainly on the observation that $\mathscr{F}$ and $\mathscr{F}^*$ are adjoint operations.

Lemma 7.16. *For every $f \in C^\infty(\mathbb{T}^n)$ and $g \in \mathscr{S}(\mathbb{Z}^n)$,*

$$\langle g, \mathscr{F} f\rangle_{\ell^2} = \langle \mathscr{F}^* g, f\rangle_{L^2}.$$

Proof. We again use Fubini's theorem for a function on $\mathbb{T}^n \times \mathbb{Z}^n$ with the product measure $m\otimes\nu$:

$$\begin{aligned}
\langle g, \mathscr{F} f\rangle_{\ell^2} &= \sum_{k\in\mathbb{Z}^n} \left\langle g(k), \int_{\mathbb{T}^n} e^{-2\pi i k\cdot x} f(x)\, dx \right\rangle = \sum_{k\in\mathbb{Z}^n} \left(\int_{\mathbb{T}^n} \langle g(k), e^{-2\pi i k\cdot x} f(x)\rangle\, dx \right) \\
&= \int_{\mathbb{T}^n\times\mathbb{Z}^n} e^{-2\pi i k\cdot x} \langle g(k), f(x)\rangle\, d(m(x)\otimes\nu(k)) \\
&= \int_{\mathbb{T}^n} \left(\sum_{k\in\mathbb{Z}^n} \langle e^{2\pi i k\cdot x} g(k), f(x)\rangle \right) dx = \int_{\mathbb{T}^n} \left\langle \sum_{k\in\mathbb{Z}^n} e^{2\pi i k\cdot x} g(k), f(x) \right\rangle dx \\
&= \int_{\mathbb{T}^n} \langle \check{g}(x), f(x)\rangle\, dx = \langle \mathscr{F}^* g, f\rangle_{L^2}.
\end{aligned}$$

Here the use of Fubini's theorem is justified since f is smooth and g is rapidly decreasing, so $(x,k) \mapsto |e^{-2\pi i k\cdot x}\langle g(k), f(x)\rangle| \leqslant |g(k)| \cdot |f(x)|$ defines an integrable function on $\mathbb{T}^n \times \mathbb{Z}^n$. ☐

Proof of Theorem 7.6. For $f, g \in C^\infty(\mathbb{T}^n)$, we have $\widehat{f}, \widehat{g} \in \mathscr{S}(\mathbb{Z}^n)$, so Lemma 7.16 and the fact that $\mathscr{F}$ and $\mathscr{F}^*$ are inverses gives

$$\langle \widehat{f}, \widehat{g}\rangle_{\ell^2} = \langle \mathscr{F}^* f, \mathscr{F}^* g\rangle_{\ell^2} = \langle f, \mathscr{F}\mathscr{F}^* g\rangle_{L^2} = \langle f, g\rangle_{L^2}.$$

☐

8. The Fourier transform

8.1. **The Fourier transform on the Schwartz space.** In this section we again assume $(V, \langle\ ,\ \rangle)$ is a finite-dimensional complex inner product space, but we now consider functions $f : \mathbb{R}^n \to V$ that are not periodic. One cannot expect these to be expressible in terms of the fully periodic functions $\varphi_k(x) := e^{2\pi i k\cdot x}$ for $k \in \mathbb{Z}^n$. On the other hand, if the periodicity condition is dropped, then the oscillatory function φ_k is well defined on $\mathbb{R}^n$ for *every* $k \in \mathbb{R}^n$, and it is natural to wonder whether arbitrary functions on $\mathbb{R}^n$ can be regarded in some sense as linear combinations of oscillatory functions of this type. Since k can now take uncountable many distinct values, our notion of a "linear combination" needs to be expanded for this discussion: instead of trying to write $f(x)$ as a series, we would now like to write it as an integral

$$f(x) = \int_{\mathbb{R}^n} e^{2\pi i p\cdot x} \hat{f}(p)\, dp \tag{8.1}$$

for some function $\hat{f} : \mathbb{R}^n \to V$, called the **Fourier transform** of f. Our discussion of the Fourier transform in this section will closely parallel that of the Fourier series, but it is in some respects more elegant, as the theory of the Fourier transform exhibits a certain symmetry that is lacking in the periodic case. This is evident when one sees the formulas for the transformations $\mathscr{F}$ and $\mathscr{F}^*$, each of which converts a function $\mathbb{R}^n \to V$ into another function $\mathbb{R}^n \to V$: for any class of functions $f, g : \mathbb{R}^n \to V$ such that the following integrals converge, we define the **Fourier transform** of f and **Fourier inverse transform** of g respectively by[16]

$$(\mathscr{F} f)(p) := \hat{f}(p) := \int_{\mathbb{R}^n} e^{-2\pi i p\cdot x} f(x)\, dx, \tag{8.2}$$

and

$$(\mathscr{F}^* g)(x) := \check{g}(x) := \int_{\mathbb{R}^n} e^{2\pi i p\cdot x} g(p)\, dp. \tag{8.3}$$

Both are clearly well defined if $f, g \in L^1(\mathbb{R}^n)$, in which case $\hat{f}$ and $\check{g}$ are both bounded functions; in fact, one can easily show via Theorem 0.4 that in this case they are continuous, so that $\mathscr{F}$ and $\mathscr{F}^*$ each define bounded linear operators

$$\mathscr{F}, \mathscr{F}^* : L^1(\mathbb{R}^n) \to C^0_b(\mathbb{R}^n).$$

Recall from §0.3 that $C^0_b(\mathbb{R}^n)$ is the Banach space of *bounded* continuous functions on $\mathbb{R}^n$, with the usual sup-norm[17]

$$\|f\|_{C^0} := \sup_{x\in\mathbb{R}^n} |f(x)|.$$

Before we can discuss in what sense these two operators are inverse to each other, we must introduce suitable function spaces on which they will both be bijective. In the setting of Fourier series, this role was played by the spaces $C^\infty(\mathbb{T}^n)$ and $\mathscr{S}(\mathbb{Z}^n)$. In the present setting, we need a single space of functions on $\mathbb{R}^n$ that combines features of both of these.

Definition 8.1. The **Schwartz space** $\mathscr{S}(\mathbb{R}^n)$, also known as the space of **smooth and rapidly decreasing** functions, consists of all smooth functions $f : \mathbb{R}^n \to V$ with the property that for every pair of multi-indices α and β, the function $\mathbb{R}^n \to V$ given by $x^\alpha \partial^\beta f(x)$ is bounded.

---

[16] The literature contains several differing opinions on where the factor of 2π should appear in (8.2) and (8.3). Our convention is the same as in [LL01, DM72], but many books omit it from the exponent, at the cost of having to insert some power $1/2\pi$ (depending on the dimension) in front of one or both integrals. A professor of mine once told of a lecture on Fourier analysis in which the speaker had solved this problem right at the beginning by saying, "Let $2\pi = 1$."

[17] Unlike the norm for continuous functions on the compact space $\mathbb{T}^n$, the supremum in the definition of $\|f\|_{C^0}$ need not be achieved for continuous functions on $\mathbb{R}^n$, and $C^0_b(\mathbb{R}^n)$ does not contain *all* continuous functions on $\mathbb{R}^n$, since not all of them are bounded.

Exercise 8.2. Show that a smooth function f on $\mathbb{R}^n$ belongs to $\mathscr{S}(\mathbb{R}^n)$ if and only if for every multi-index α and every $k \in \mathbb{N}$, there exists a constant $C > 0$ dependent on α and k such that

$$|\partial^\alpha f(x)| \leqslant \frac{C}{1+|x|^k} \qquad \text{for all} \qquad x \in \mathbb{R}^n.$$

Exercise 8.3. Show that $\mathscr{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$ for every $p \in [1,\infty]$, and for every $f \in \mathscr{S}(\mathbb{R}^n)$ and every multi-index α, the functions $\partial^\alpha f$ and $x \mapsto x^\alpha f(x)$ also belong to $\mathscr{S}(\mathbb{R}^n)$.

The next two theorems are the main results we aim to prove in this section about the Fourier transform.

Theorem 8.4. *The transformations $\mathscr{F}$ and $\mathscr{F}^*$ each map $\mathscr{S}(\mathbb{R}^n)$ bijectively to itself, and they are inverse to each other.*

Theorem 8.5 (Plancherel's theorem). *For every $f, g \in \mathscr{S}(\mathbb{R}^n)$, $\langle f, g\rangle_{L^2} = \langle \hat{f}, \hat{g}\rangle_{L^2}$.*

In particular, the linear operators $\mathscr{F}, \mathscr{F}^* : \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$ are isometries (and are therefore continuous) with respect to the L^2-norm. The Schwartz space contains the space of smooth compactly supported functions, which is dense in $L^2(\mathbb{R}^n)$, thus $\mathscr{S}(\mathbb{R}^n)$ is also dense in $L^2(\mathbb{R}^n)$, so this result implies:

Corollary 8.6. *The operators $\mathscr{F}, \mathscr{F}^* : \mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$ admit unique extensions to unitary isomorphisms $L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ such that $\mathscr{F}^* = \mathscr{F}^{-1}$.* □

Proposition 8.7. *For $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, the definitions of $\mathscr{F}f$ and $\mathscr{F}^*f$ in* (8.2) *and* (8.3) *respectively agree (up to equality almost everywhere) with their definitions as described in Corollary 8.6 via Plancherel's theorem and the density of $\mathscr{S}(\mathbb{R}^n) \subset L^2(\mathbb{R}^n)$.*

Proof. To avoid confusion, let us denote by $\hat{f}_{L^1}$ and $\hat{f}_{L^2}$ the two possible definitions of $\hat{f}$ as defined via (8.2) or via the density of $\mathscr{S}(\mathbb{R}^n) \subset L^2(\mathbb{R}^n)$. We claim first that there exists a sequence of smooth compactly supported functions $f_j \in C_0^\infty(\mathbb{R}^n)$ that converge to f in both the L^1- and L^2-norms. Indeed, choosing an approximate identity $\rho_j : \mathbb{R}^n \to [0,\infty)$ with shrinking support as in §5, the smooth functions $\rho_j * f$ converge to f in both L^1 and L^2 according to Theorem 5.14, and one can then define f_j by multiplying these by suitable compactly supported cutoff functions as in the discussion preceding Corollary 5.2. With this sequence chosen, the functions $f_j \in C_0^\infty(\mathbb{R}^n)$ also belong to $\mathscr{S}(\mathbb{R}^n)$, so the L^2-convergence $f_j \to f$ implies that $\hat{f}_j$ converges in L^2 to $\hat{f}_{L^2}$, and it follows that $\hat{f}_j$ also has a subsequence converging pointwise almost everywhere to $\hat{f}_{L^2}$. But since $\mathscr{F} : L^1(\mathbb{R}^n) \to C_b^0(\mathbb{R}^n)$ as defined by (8.2) is a bounded linear map, the L^1-convergence $f_j \to f$ implies additionally that $\hat{f}_j$ converges uniformly to the continuous function $\hat{f}_{L^1}$. This can only be true if $\hat{f}_{L^1} = \hat{f}_{L^2}$ almost everywhere. A completely analogous argument works for $\mathscr{F}^*$. □

For a function $f \in L^2(\mathbb{R}^n)$ that is not in $L^1(\mathbb{R}^n)$, the formula for $\hat{f}$ in (8.2) does not strictly make sense, because the integral does not converge, but the continuity of $\mathscr{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ means that one can define $\hat{f}$ as the L^2-limit of the L^2-Cauchy sequence $\hat{f}_j \in C_b^0(\mathbb{R}^n)$ for any sequence $f_j \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ converging in L^2 to f. Exercise 8.9 below describes a reasonable trick for carrying this out in practice.

Remark 8.8. If $f \in L^2(\mathbb{R}^n)$ but $f \notin L^1(\mathbb{R}^n)$, then $\hat{f}$ and $\check{f}$ are not functions, strictly speaking, but rather equivalence classes of functions up to equality almost everywhere, so their values $\hat{f}(p)$ and $\check{f}(p)$ at individual points $p \in \mathbb{R}^n$ are not well defined. In contrast, $\hat{f}(p)$ and $\check{f}(p)$ are well defined for every $p \in \mathbb{R}^n$ via the integrals (8.2) or (8.3) if $f \in L^1(\mathbb{R}^n)$.

Exercise 8.9. Show that for $f, g \in L^2(\mathbb{R}^n)$, the following conditions are equivalent:

(1) $\hat{f} = g$ almost everywhere;

(2) There exists a sequence $R_j \to \infty$ such that $\lim_{j\to\infty} \int_{B_{R_j}} e^{-2\pi i p\cdot x} f(x)\, dx = g(p)$ for almost every $p \in \mathbb{R}^n$. (Here $B_R \subset \mathbb{R}^n$ denotes the ball of radius R about the origin.)

Hint: Multiply f by characteristic functions to define L^2-close approximations that are also in $L^1(\mathbb{R}^n)$.

8.2. **Fourier transforms and derivatives.** For $f \in L^1(\mathbb{R}^n)$, the function $e^{-2\pi p\cdot x} f(x)$ is continuous with respect to p and also, as a function of x, bounded for every $p \in \mathbb{R}^n$ by the fixed Lebesgue-integrable function $|f| : \mathbb{R}^n \to [0,\infty)$. Viewing $\widehat{f}(p)$ as a parameter-dependent integral and applying Theorem 0.4 (and similarly for $\check{f}(x)$) thus proves:

Proposition 8.10. *If $f \in L^1(\mathbb{R}^n)$, then $\mathscr{F}f$ and $\mathscr{F}^*f$ belong to $C^0_b(\mathbb{R}^n)$, and the resulting maps $\mathscr{F}, \mathscr{F}^* : L^1(\mathbb{R}^n) \to C^0_b(\mathbb{R}^n)$ are bounded linear operators.* □

Exercise 8.11. Use integration by parts and/or Theorem 0.4 to establish the following analogues of Exercises 7.9 and 7.10:

(1) Suppose $f \in L^1(\mathbb{R}^n)$, f has a continuous partial derivative $\partial_j f : \mathbb{R}^n \to V$ that also belongs to $L^1(\mathbb{R}^n)$ for some $j = 1,\dots,n$, and f "decays at infinity" in the sense that $\lim_{R\to\infty} \sup_{x\in\mathbb{R}^n\setminus B_R} |f(x)| = 0$, where $B_R \subset \mathbb{R}^n$ denotes the ball of radius R about $0 \in \mathbb{R}^n$. Then
$$\widehat{\partial_j f}(p) = 2\pi i p_j \widehat{f}(p) \quad \text{and} \quad \widecheck{\partial_j f}(x) = -2\pi i x_j \check{f}(x)$$
for each $p = (p_1,\dots,p_n) \in \mathbb{R}^n$ and $x = (x_1,\dots,x_n) \in \mathbb{R}^n$.

(2) Given $f : \mathbb{R}^n \to V$ and $j \in \{1,\dots,n\}$, let $f_j : \mathbb{R}^n \to V$ denote the function $f_j(x) := x_j f(x)$ for $x = (x_1,\dots,x_n) \in \mathbb{R}^n$. If f and f_j both belong to $L^1(\mathbb{R}^n)$, then $\check{f}, \widehat{f} : \mathbb{R}^n \to V$ are continuous and have continuous partial derivatives $\partial_j \check{f}, \partial_j \widehat{f} : \mathbb{R}^n \to V$ given by
$$\partial_j \check{f}(x) = \widecheck{2\pi i f_j}(x) \quad \text{and} \quad \partial_j \widehat{f}(p) = \widehat{-2\pi i f_j}(p).$$

If $f \in \mathscr{S}(\mathbb{R}^n)$, then the conditions in both parts of Exercise 8.11 are satisfied and the formulas may be iterated arbitrarily many times, proving that for every multi-index α,

$$\begin{aligned} \widehat{\partial^\alpha f}(p) &= (2\pi i p)^\alpha \widehat{f}(p), & \widecheck{\partial^\alpha f}(x) &= (-2\pi i x)^\alpha \check{f}(x), \\ \partial^\alpha \check{f}(x) &= (2\pi i)^{|\alpha|} \check{f_\alpha}(x), & \partial^\alpha \widehat{f}(p) &= (-2\pi i)^{|\alpha|} \widehat{f_\alpha}(p), \end{aligned} \tag{8.4}$$

where $f_\alpha(x) := x^\alpha f(x)$. Implicit in the last two formulas is that $\partial^\alpha \check{f}$ and $\partial^\alpha \widehat{f}$ exist for every α, i.e. in this case, $\check{f}$ and $\widehat{f}$ are also smooth.

Proof of Theorem 8.4, part 1. For $f \in \mathscr{S}(\mathbb{R}^n)$, we have already shown above that $\widehat{f}$ is smooth, and for each pair of multi-indices α, β $\partial^\alpha \widehat{f}$ satisfies
$$p^\beta \partial^\alpha \widehat{f}(p) = p^\beta (-2\pi i)^{|\alpha|} \widehat{f_\alpha}(p) = \frac{(-2\pi i)^{|\alpha|}}{(2\pi i)^{|\beta|}} (2\pi i p)^\beta \widehat{f_\alpha}(p) = \frac{(-2\pi i)^{|\alpha|}}{(2\pi i)^{|\beta|}} \widehat{\partial^\beta f_\alpha}(p).$$
By the definition of the Schwartz space, f_α and $\partial^\beta f_\alpha$ also belong to $\mathscr{S}(\mathbb{R}^n)$, so in particular, the latter is in $L^1(\mathbb{R}^n)$ and its Fourier transform is therefore bounded. This proves $\widehat{f} \in \mathscr{S}(\mathbb{R}^n)$. One shows in the same manner that $\check{f} \in \mathscr{S}(\mathbb{R}^n)$, so we have proved:
$$\mathscr{F}(\mathscr{S}(\mathbb{R}^n)) \subset \mathscr{S}(\mathbb{R}^n) \quad \text{and} \quad \mathscr{F}^*(\mathscr{S}(\mathbb{R}^n)) \subset \mathscr{S}(\mathbb{R}^n).$$
□

8.3. **The Gaussian.** One class of functions in $\mathscr{S}(\mathbb{R}^n)$ whose Fourier transforms can be computed explicitly are the *Gaussians*, i.e. functions of the form $Ae^{-c|x|^2}$ for constants $A, c > 0$. The computation carried out in this subsection is more than just an amusing exercise: the proof of the inversion formula in §8.5 will require an approximate identity with particular properties, and Gaussians furnish the most convenient construction of such an object.

Proposition 8.12. *For any constant $a > 0$, the function $f(x) := e^{-a^2|x|^2}$ on $\mathbb{R}^n$ satisfies*

$$\widehat{f}(x) = \check{f}(x) = \frac{\pi^{n/2}}{a^n} e^{-(\pi/a)^2|x|^2}.$$

Proof. By Fubini's theorem,

$$\begin{aligned}\widehat{f}(p) &= \int_{\mathbb{R}^n} e^{-a^2(x_1^2+\ldots+x_n^2)} e^{-2\pi i(p_1x_1+\ldots+p_nx_n)}\,dx = \int_{\mathbb{R}^n} e^{-a^2x_1^2}\ldots e^{-a^2x_n^2} e^{-2\pi ip_1x_1}\ldots e^{-2\pi ip_nx_n}\,dx \\ &= \left(\int_{-\infty}^{\infty} e^{-a^2x_1^2} e^{-2\pi ip_1x_1}\,dx_1\right)\ldots\left(\int_{-\infty}^{\infty} e^{-a^2x_n^2} e^{-2\pi ip_nx_n}\,dx_n\right),\end{aligned}$$

thus it will suffice to prove that the stated formula for $\widehat{f}$ is correct in the case $n = 1$. Consider $f(x) := e^{-a^2x^2}$ on $\mathbb{R}$. Instead of computing the integral

$$\widehat{f}(p) = \int_{-\infty}^{\infty} e^{-a^2x^2} e^{-2\pi ipx}\,dx$$

explicitly for every $p \in \mathbb{R}$, we shall identify the function $\widehat{f}$ as the unique solution to a certain initial value problem. For $p = 0$, we have

$$\widehat{f}(0) = \int_{-\infty}^{\infty} e^{-a^2x^2}\,dx = \frac{1}{a}\int_{-\infty}^{\infty} e^{-u^2}\,du = \frac{\sqrt{\pi}}{a},$$

which follows via the substitution $u = ax$ and the well-known formula $\int_{-\infty}^{\infty} e^{-u^2}\,du = \sqrt{\pi}$. Applying (8.4) and then integrating by parts, we also have

$$\begin{aligned}\widehat{f}'(p) &= -2\pi i\widehat{xf}(p) = -2\pi i\int_{-\infty}^{\infty} xe^{-a^2x^2}e^{-2\pi ipx}\,dx = \frac{i\pi}{a^2}\int_{-\infty}^{\infty}\frac{d}{dx}\left(e^{-a^2x^2}\right)\cdot e^{-2\pi ipx}\,dx \\ &= -\frac{i\pi}{a^2}\int_{-\infty}^{\infty} e^{-a^2x^2}\frac{d}{dx}\left(e^{-2\pi ipx}\right)\,dx = \frac{i\pi}{a^2}\cdot 2\pi ip\int_{-\infty}^{\infty} e^{-a^2x^2}e^{-2\pi ipx}\,dx = -\frac{2\pi^2}{a^2}p\widehat{f}(p),\end{aligned}$$

in other words, $\widehat{f} : \mathbb{R} \to \mathbb{C}$ satisfies the initial value problem

$$\begin{cases}\dfrac{d\widehat{f}}{dp} &= -2(\pi/a)^2p\widehat{f}, \\ \widehat{f}(0) &= \sqrt{\pi}/a.\end{cases}$$

The unique solution to this problem is $\widehat{f}(p) = \frac{\sqrt{\pi}}{a}e^{-(\pi/a)^2p^2}$.

Since f is a real-valued function, $\check{f}$ is the complex conjugate of $\widehat{f}$, which is $\widehat{f}$ itself. □

Corollary 8.13. *The Gaussian $f(x) = e^{-a^2|x|^2}$ with $a > 0$ satisfies $\mathscr{F}^*\mathscr{F}f = \mathscr{F}\mathscr{F}^*f = f$.* □

8.4. **Approximate identities revisited.** If the Dirac δ-function $\delta : \mathbb{R}^n \to [0, \infty]$ were an actual function in $\mathscr{S}(\mathbb{R}^n)$, its Fourier transform would be

$$\widehat{\delta}(p) = \int_{\mathbb{R}^n} e^{-2\pi ip\cdot x}\delta(x)\,dx = 1,$$

leading to the slightly nonsensical formula

$$\delta(x) = \int_{\mathbb{R}^n} e^{2\pi ip\cdot x}\,dp. \tag{8.5}$$

As with most things involving the δ-function, one can make mathematical sense of this formula in terms of approximate identities, and the proof of the Fourier inversion formula in the next subsection will require the existence of an approximate identity for which the inversion formula is known to hold. For our purposes in this context, "approximate identity" means the following:

Definition 8.14. A **tempered approximate identity** on $\mathbb{R}^n$ is a sequence $\rho_j : \mathbb{R}^n \to [0,\infty)$ of nonnegative functions in $\mathscr{S}(\mathbb{R}^n)$ such that for every $\varphi \in \mathscr{S}(\mathbb{R}^n)$,

$$\lim_{j\to\infty}\int_{\mathbb{R}^n} \rho_j(x)\varphi(x)\,dx = \varphi(0). \tag{8.6}$$

Note that the assumption $\rho_j \in \mathscr{S}(\mathbb{R}^n)$ implies that the convolution $\rho_j * f$ is a well-defined function $\mathbb{R}^n \to V$ for every $f \in \mathscr{S}(\mathbb{R}^n)$, and (8.6) then implies $(\rho_j * f)(x) = (f * \rho_j)(x) = \int_{\mathbb{R}^n} f(x-y)\rho_j(y)\,dy \to f(x)$, hence

$$\rho_j * f \to f \text{ pointwise for } f \in \mathscr{S}(\mathbb{R}^n). \tag{8.7}$$

Lemma 8.15. *Suppose $\rho : \mathbb{R}^n \to [0,\infty)$ is a smooth function satisfying $\int_{\mathbb{R}^n}\rho(x)\,dx = 1$. Then the sequence $\rho_j : \mathbb{R}^n \to [0,\infty)$ defined by*

$$\rho_j(x) := j^n\rho(jx)$$

satisfies (8.6) *for every bounded continuous function $\varphi : \mathbb{R}^n \to V$.*

Proof. We use the change of variables $y := jx$ to write

$$\int_{\mathbb{R}^n}\rho_j(x)\varphi(x)\,dx = \int_{\mathbb{R}^n}\rho(y)\varphi(y/j)\,dy.$$

Since φ is bounded and continuous, the integrands on the right converge pointwise as $j \to \infty$ to $\varphi(0)\rho$ and are uniformly bounded by a constant multiple of the integrable function ρ. The dominated convergence theorem thus implies that the integrals converge to $\int_{\mathbb{R}^n}\varphi(0)\rho\,dm = \varphi(0)$. ☐

Lemma 8.16. *There exists a tempered approximate identity $\rho_j \in \mathscr{S}(\mathbb{R}^n)$ with the following properties:*

(1) $\mathscr{F}^*\widehat{\rho}_j = \rho_j$ for every j;
(2) The functions $\widehat{\rho}_j$ satisfy a uniform bound $|\widehat{\rho}_j| \leqslant C$ for all j and converge pointwise to 1.

Proof. Set $\rho(x) := \frac{1}{\sqrt{\pi}}e^{-|x|^2}$ and use this to define ρ_j as in Lemma 8.15. Then ρ_j is a Gaussian for every j, so both ρ_j and $\widehat{\rho}_j$ are in $\mathscr{S}(\mathbb{R}^n)$, and Corollary 8.13 implies $\mathscr{F}^*\widehat{\rho}_j = \rho_j$. Applying Lemma 8.15 with the bounded continuous function $f(x) = e^{-2\pi i p\cdot x}$ for each $p \in \mathbb{R}^n$, we also have

$$\lim_{j\to\infty}\widehat{\rho}_j(p) = \lim_{j\to\infty}\int_{\mathbb{R}^n} e^{-2\pi i p\cdot x}\rho_j(x)\,dx = 1$$

and

$$|\widehat{\rho}_j(p)| \leqslant \|\rho_j\|_{L^1} = \|\rho\|_{L^1} = 1,$$

where a quick computation via change of variables gives $\int_{\mathbb{R}^n}\rho_j\,dm = \int_{\mathbb{R}^n}\rho\,dm$. Alternatively, these last two statements also follow from the explicit computation of $\widehat{\rho}_j$ in Proposition 8.12. ☐

8.5. **The Fourier inversion formula.** We can now prove that the operators $\mathscr{F}$ and $\mathscr{F}^*$ on $\mathscr{S}(\mathbb{R}^n)$ are inverse to each other.

The "physicist's proof" that $\mathscr{F}^*\widehat{f} = f$ works via the following adventurous application of Fubini's theorem:

$$\begin{aligned}
(\mathscr{F}^*\widehat{f})(x) &= \int_{\mathbb{R}^n} e^{2\pi i p\cdot x}\widehat{f}(p)\,dp = \int_{\mathbb{R}^n} e^{2\pi i p\cdot x}\left(\int_{\mathbb{R}^n} e^{-2\pi i p\cdot y}f(y)\,dy\right)dp \\
&= \int_{\mathbb{R}^n\times\mathbb{R}^n} e^{2\pi i p\cdot(x-y)}f(y)\,dy\,dp = \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} e^{2\pi i p\cdot(x-y)}f(y)\,dp\right)dy \\
&= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} e^{2\pi i p\cdot(x-y)}\,dp\right)f(y)\,dy = \int_{\mathbb{R}^n}\delta(x-y)f(y)\,dy = \int_{\mathbb{R}^n} f(x-y)\delta(y)\,dy \\
&= f(x).
\end{aligned}$$

Here the δ-function appears due to the formal relation (8.5), and something clearly must be modified to justify the use of Fubini's theorem since $(y,p) \mapsto e^{2\pi i p\cdot(x-y)}f(y)$ is not an integrable function on $\mathbb{R}^n \times \mathbb{R}^n$ for any $x \in \mathbb{R}^n$. In analogy with §7.6, the remedy is to multiply this function by the Fourier transform of a tempered approximate identity $\rho_j \in \mathscr{S}(\mathbb{R}^n)$, and then take the limit of the resulting relation as $j \to \infty$.

Proof of Theorem 8.4, conclusion. Given $f \in \mathscr{S}(\mathbb{R}^n)$, we need to show $\mathscr{F}^* \widehat{f} = f$. Choose a tempered approximate identity ρ_j with the properties listed in Lemma 8.16. We then have $\widehat{\rho}_j \in L^1(\mathbb{R}^n)$, and for every $x \in \mathbb{R}^n$, the function $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$ given by

$$F(y,p) := e^{2\pi i p\cdot x} e^{-2\pi i p\cdot y} \widehat{\rho}_j(p) f(y)$$

is therefore integrable. Applying Fubini's theorem to the integral of F over $\mathbb{R}^n \times \mathbb{R}^n$ now gives

$$\int_{\mathbb{R}^n} e^{2\pi i p\cdot x} \widehat{\rho}_j(p) \left(\int_{\mathbb{R}^n} e^{-2\pi i p\cdot y} f(y) dy \right) dp = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} e^{2\pi i p\cdot(x-y)} \widehat{\rho}_j(p)\, dp \right) f(y)\, dy.$$

The left hand side is $\int_{\mathbb{R}^n} e^{2\pi i p\cdot x} \widehat{\rho}_j(p) \widehat{f}(p)\, dp$, which converges via the dominated convergence theorem as $j \to \infty$ to $\int_{\mathbb{R}^n} e^{2\pi i p\cdot x} \widehat{f}(p)\, dp$ since $\widehat{f} \in \mathscr{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$ while (by Lemma 8.16) $\widehat{\rho}_j$ is uniformly bounded and converges pointwise to 1. The right hand side is likewise $\int_{\mathbb{R}^n} \rho_j(x - y) f(y)\, dy = (\rho_j * f)(x)$, which converges to $f(x)$ by (8.7). We've proved:

$$\mathscr{F}^* \mathscr{F} f = f \quad \text{for all} \quad f \in \mathscr{S}(\mathbb{R}^n).$$

An almost identical argument proves $\mathscr{F}\mathscr{F}^* f = f$ for all $f \in \mathscr{S}(\mathbb{R}^n)$. □

8.6. **Plancherel's theorem.** With the Fourier inversion formula in hand, Plancherel's theorem will follow easily from the observation that $\mathscr{F}$ and $\mathscr{F}^*$ are adjoints:

Lemma 8.17. *For every $f, g \in \mathscr{S}(\mathbb{R}^n)$,*

$$\langle g, \mathscr{F} f \rangle_{L^2} = \langle \mathscr{F}^* g, f \rangle_{L^2}.$$

Proof. Since f and g are both in $\mathscr{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$, the function $(x, p) \mapsto |e^{-2\pi i p\cdot x} \langle g(p), f(x) \rangle| \leqslant |g(p)| \cdot |f(x)|$ is integrable on $\mathbb{R}^n \times \mathbb{R}^n$, so Fubini's theorem gives

$$\begin{aligned}
\langle g, \mathscr{F} f \rangle_{L^2} &= \int_{\mathbb{R}^n} \left\langle g(p), \left(\int_{\mathbb{R}^n} e^{-2\pi i p\cdot x} f(x)\, dx \right) \right\rangle dp = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \langle g(p), e^{-2\pi i p\cdot x} f(x) \rangle\, dx \right) dp \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} e^{-2\pi i p\cdot x} \langle g(p), f(x) \rangle\, d(m(x) \otimes m(p)) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \langle e^{2\pi i p\cdot x} g(p), f(x) \rangle\, dp \right) dx \\
&= \int_{\mathbb{R}^n} \left\langle \int_{\mathbb{R}^n} e^{2\pi i p\cdot x} g(p)\, dp, f(x) \right\rangle dx = \int_{\mathbb{R}^n} \langle \check{g}(x), f(x) \rangle\, dx = \langle \mathscr{F}^* g, f \rangle_{L^2}.
\end{aligned}$$

□

Proof of Theorem 8.5. For $f, g \in \mathscr{S}(\mathbb{R}^n)$, the Fourier transforms $\widehat{f}, \widehat{g}$ are also in $\mathscr{S}(\mathbb{R}^n)$, so Lemma 8.17 together with the relation $\mathscr{F}\mathscr{F}^* = \mathrm{Id}$ gives

$$\langle \widehat{f}, \widehat{g} \rangle_{L^2} = \langle \mathscr{F}^* f, \mathscr{F}^* g \rangle_{L^2} = \langle f, \mathscr{F}\mathscr{F}^* g \rangle_{L^2} = \langle f, g \rangle_{L^2}.$$

□

8.7. **Convolutions.** If f and g are functions of class L^1 and L^2 respectively on $\mathbb{R}^n$ and at least one of them is assumed to be scalar valued (so that pointwise products $f(x)g(x)$ are well defined), then Young's inequality (Theorem 5.8) implies that the convolution $f * g$ is a well-defined function in $L^2(\mathbb{R}^n)$. Since $\widehat{f}, \check{f} \in C^0_b(\mathbb{R}^n)$ and $\widehat{g}, \check{g} \in L^2(\mathbb{R}^n)$ in this situation, the pointwise products $\widehat{f}\widehat{g}$ and $\check{f}\check{g}$ are also well-defined functions in $L^2(\mathbb{R}^n)$, and the formulas in the following result therefore make sense:

Theorem 8.18. *If $f \in L^1(\mathbb{R}^n)$ and $g \in L^2(\mathbb{R}^n)$, then $\mathscr{F}(f * g) = \widehat{f}\widehat{g}$ and $\mathscr{F}^*(f * g) = \check{f}\check{g}$ almost everywhere.*

Proof. We focus on the formula for $\mathscr{F}(f * g)$, as the same argument works for $\mathscr{F}^*(f * g)$. If $f, g \in L^1(\mathbb{R}^n)$, then the formula is a straightforward application of Fubini's theorem, which we leave as a (highly recommended!) exercise. To extend this result to general $g \in L^2(\mathbb{R}^n)$, one can choose a sequence $g_k \in \mathscr{S}(\mathbb{R}^n)$ with $g_k \to g$ in L^2: then $\widehat{f * g_k} = \hat{f}\hat{g}_k$ since $\mathscr{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$, and Young's inequality implies $f * g_k \to f * g$ in L^2, hence by Plancherel's theorem, $\hat{f}\hat{g}_k = \widehat{f * g_k} \xrightarrow{L^2} \widehat{f * g}$. At the same time, $\hat{f} \in C^0_b(\mathbb{R}^n)$ and Plancherel's theorem also implies $\hat{g}_k \to \hat{g}$ in L^2, thus $\hat{f}\hat{g}_k$ also converges in L^2 to $\hat{f}\hat{g}$. $\square$

There are two analogues of Theorem 8.18 for fully periodic functions and Fourier series. We defined in Remark 7.12 the convolution of two periodic functions f and g as an integral over the torus $\mathbb{T}^n$. There is a similar definition for functions on $\mathbb{Z}^n$, with Lebesgue integration replaced by summation (i.e. integration with respect to the counting measure): for two functions f, g on $\mathbb{Z}^n$ such that at least one is scalar valued, we write

$$(f * g)(k) := \sum_{j \in \mathbb{Z}^n} f(k - j) g(j).$$

This is considered well-defined for a given $k \in \mathbb{Z}^n$ if and only if the sum on the right hand side converges absolutely.

Exercise 8.19. Adapt the proof of Theorem 5.8 to show that Young's inequality also holds for functions on $\mathbb{T}^n$ and $\mathbb{Z}^n$, that is:

(a) For any $f \in L^1(\mathbb{T}^n)$ and $g \in L^p(\mathbb{T}^n)$ with $1 \leqslant p \leqslant \infty$, $(f * g)(x)$ is defined for almost every $x \in \mathbb{T}^n$ and determines a function $f * g \in L^p(\mathbb{T}^n)$ such that $\|f * g\|_{L^p} \leqslant \|f\|_{L^1} \cdot \|g\|_{L^p}$.

(b) For any $f \in \ell^1(\mathbb{Z}^n)$ and $g \in \ell^p(\mathbb{Z}^n)$ with $1 \leqslant p \leqslant \infty$, $(f * g)(k)$ is defined for all $k \in \mathbb{Z}^n$ and satisfies $\|f * g\|_{\ell^p} \leqslant \|f\|_{\ell^1} \cdot \|g\|_{\ell^p}$.

Exercise 8.20. Prove the following analogues of Theorem 8.18 for Fourier series:

(a) For any $f, g \in L^1(\mathbb{T}^n)$, the Fourier coefficients of $f * g \in L^1(\mathbb{T}^n)$ are given by $\widehat{f * g}_k = \hat{f}_k \hat{g}_k$.[18]

(b) If $f \in C^0(\mathbb{T}^n)$ and $g \in L^2(\mathbb{T}^n)$ have Fourier coefficients $\hat{f} \in \ell^1(\mathbb{Z}^n)$ and $\hat{g} \in \ell^2(\mathbb{Z}^n)$, then the Fourier coefficients of fg are given by $\widehat{fg}_k = (\hat{f} * \hat{g})_k$.

The following amusing variation on Theorem 8.18 will be useful in our discussion of nowhere differentiable functions in the next subsection. Suppose f is a fully periodic function on $\mathbb{R}^n$, expressed as a Fourier series $f(x) = \sum_{k \in \mathbb{Z}^n} e^{2\pi i k \cdot x} \hat{f}_k$. This function does not belong to $L^p(\mathbb{R}^n)$ for any $p < \infty$ unless it is zero almost everywhere, thus we cannot define a Fourier transform for f in the usual sense.[19] In the following paragraph, we shall ignore this difficulty as we did in the initial "physicist's proofs" of Theorems 7.5 and 8.4, thus the reader is asked to temporarily suspend all skepticism about issues like convergence, interchange of summation and integration, and the existence of the Dirac δ-function. The logical gaps will be filled in subsequently.

With this understood, let us pretend that $\hat{f}$ is a well-defined function on $\mathbb{R}^n$ given by the usual formula $\hat{f}(p) = \int_{\mathbb{R}^n} e^{-2\pi i p \cdot x} f(x)\, dx$. To write it down more precisely, observe that the inverse Fourier transform of the (fictional) Dirac δ-function is given by $\check{\delta}(x) = \int_{\mathbb{R}^n} e^{2\pi i p \cdot x} \delta(p)\, dp = 1$, so applying the Fourier inversion formula gives the formal relation $\hat{1} = \delta$, or in verbose form (cf. (8.5)),

$$\delta(p) = \int_{\mathbb{R}^n} e^{-2\pi i p \cdot x}\, dx. \tag{8.8}$$

[18] We are not mentioning the case $g \in L^2(\mathbb{T}^n)$ here because it is redundant: since $\mathbb{T}^2$ has finite measure, $L^2(\mathbb{T}^n) \subset L^1(\mathbb{T}^n)$.

[19] A function $f \notin L^2(\mathbb{R}^n)$ may nonetheless have a well-defined Fourier transform that is not a function but a *tempered distribution*; see §10.6. This notion can be used to give rigorous meaning to formulas like $\hat{1} = \delta$, though it is not required for the present discussion.

This suggests the formula

$$(8.9)\qquad \widehat{f}(p) = \int_{\mathbb{R}^n} e^{-2\pi i p\cdot x} \sum_{k\in\mathbb{Z}^n} e^{2\pi i k\cdot x} \widehat{f}_k \, dx = \sum_{k\in\mathbb{Z}^n} \left(\int_{\mathbb{R}^n} e^{-2\pi i (p-k)\cdot x} \, dx \right) \widehat{f}_k = \sum_{k\in\mathbb{Z}^n} \delta(p-k) \widehat{f}_k.$$

The support of this "function" is $\mathbb{Z}^n$ since $\delta(p) = 0$ for all $p \neq 0$. Now for a given $k \in \mathbb{Z}^n$, choose a smooth compactly supported function $\widehat{\psi} : \mathbb{R}^n \to [0,1]$ that satisfies $\widehat{\psi}(k) = 1$ and has no other points of $\mathbb{Z}^n$ in its support. We have labeled it $\widehat{\psi}$ because, as an element of $\mathscr{S}(\mathbb{R}^n)$, $\widehat{\psi}$ is the Fourier transform of another function $\psi \in \mathscr{S}(\mathbb{R}^n)$. The product of $\widehat{\psi}$ with the right hand side of (8.9) is $\delta(p-k)\widehat{f}_k$, which is formally the Fourier transform of $e^{2\pi i k\cdot x}\widehat{f}_k$, i.e. a single term in the Fourier series for f. Since products of Fourier transforms are Fourier transforms of convolutions according to Theorem 8.18, we can take this formal discussion as motivation for the formula

$$(8.10)\qquad (\psi * f)(x) = e^{2\pi i k\cdot x} \widehat{f}_k.$$

In contrast with several other questionable things that have been said in this paragraph, (8.10) does not look at all implausible, e.g. both sides are smooth bounded functions on $\mathbb{R}^n$ (for the left hand side this follows from Theorem 5.7 and Young's inequality since $\psi \in \mathscr{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$ and $f \in L^\infty(\mathbb{R}^n)$). Let us now give a rigorous proof.

Lemma 8.21. *Suppose f is a continuous fully periodic function on $\mathbb{R}^n$ with absolutely summable Fourier coefficients $\widehat{f} \in \ell^1(\mathbb{Z}^n)$, and $\psi : \mathbb{R}^n \to \mathbb{C}$ is the inverse Fourier transform of a function $\widehat{\psi} \in \mathscr{S}(\mathbb{R}^n)$ with $\widehat{\psi}(k) = 1$ for some $k \in \mathbb{Z}^n$ and $\widehat{\psi}(k') = 0$ for all $k' \in \mathbb{Z}^n \backslash \{k\}$. Then* (8.10) *holds.*

Proof. The reversal of summation and integration in the following computation is justified by the dominated convergence theorem since $\left| \psi(y) \sum_{j\in\mathbb{Z}^n} e^{2\pi i j\cdot(x-y)} \widehat{f}_j \right| \leqslant |\psi(y)| \cdot \|\widehat{f}\|_{\ell^1}$ and $\psi \in \mathscr{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$:

$$\begin{aligned} (\psi * f)(x) = (f * \psi)(x) &= \int_{\mathbb{R}^n} f(x-y)\psi(y)\, dy = \int_{\mathbb{R}^n} \psi(y) \Big(\sum_{j\in\mathbb{Z}^n} e^{2\pi i j\cdot(x-y)} \widehat{f}_j \Big)\, dy \\ &= \sum_{j\in\mathbb{Z}^n} e^{2\pi i j\cdot x} \left(\int_{\mathbb{R}^n} \psi(y) e^{-2\pi i j\cdot y}\, dy \right) \widehat{f}_j = \sum_{j\in\mathbb{Z}^n} e^{2\pi i j\cdot x} \widehat{\psi}(j) \widehat{f}_j = e^{2\pi i k\cdot x} \widehat{f}_k. \end{aligned}$$

□

8.8. **Nowhere differentiable functions.** Fix constants $a, b > 1$ and consider the function $f : \mathbb{R} \to \mathbb{C}$ defined by

$$(8.11)\qquad f(x) := \sum_{k=0}^{\infty} \frac{1}{a^k} e^{2\pi i b^k x}.$$

Since $\sum_{k=0}^\infty \frac{1}{a^k} < \infty$, the partial sums of this series converge uniformly to a continuous function. If $b \in \mathbb{N}$, then f is periodic, and (8.11) is an expression of its Fourier series. Differentiating it term by term gives

$$(8.12)\qquad f'(x) = 2\pi i \sum_{k=0}^{\infty} \frac{b^k}{a^k} e^{2\pi i b^k x},$$

a formula that should be taken with a grain of salt until we have investigated whether the right hand side converges. In fact, the series converges absolutely and uniformly if $b < a$, and it follows in this case that f is indeed continuously differentiable. The interesting question is what happens when $b \geqslant a$.

Theorem 8.22. *If $b \geqslant a > 1$, then the function f in* (8.11) *is not differentiable at any point.*

Up to unimportant details such as the factor of 2π in the exponent, the real part of f is the function that was introduced by Weierstrass in 1872 as the first published example of a continuous but nowhere differentiable function. It was later [Ban31, Maz31] shown that, while such functions are typically not so easy to write down, they are not at all unusual, e.g. the subset of $C^0([0,1])$ consisting of nowhere differentiable functions is dense, and even better, it is *comeager*, meaning it is a countable intersection of open and dense subsets.[20] In other words, "almost all" continuous functions are nowhere differentiable in some quantifiable sense.

The version of Theorem 8.22 proved by Weierstrass included the extra conditions that b is an odd integer and $b/a > 1 + \frac{3}{2}\pi$, which are not necessary. In the form stated here, Theorem 8.22 is due to Hardy [Har16], and our proof below is adapted from [Joh10].

Some initial intuition for Theorem 8.22 comes from (8.12), as we have already learned to expect some correspondence between the differentiability of a function and the rate at which its Fourier coefficients decay. This correspondence typically goes in only one direction, e.g. absolute summability of the series $\sum_k \widehat{f}_k$ or $\sum_k |k|\widehat{f}_k$ implies continuity of f or f' respectively, but not every continuous function has summable Fourier coefficients. The challenge in Theorem 8.22 is similar, as we need to show that if f is differentiable at some point, then the coefficients on the right hand side of (8.12) must indeed be absolutely summable. The Weierstrass function has a special property that makes proving results like this more feasible: its Fourier series is **lacunary**, meaning that most of its Fourier coefficients are zero, and the gaps between its nonzero terms become wider (at an exponential rate) as the series continues. We will not give a more precise definition of this property here, nor mention it explicitly in the proof below, but you may recognize where it is used implicitly if you pay careful attention. A similar result worth mentioning is that for any function g on $S^1 := \mathbb{T}^1$ with a lacunary Fourier series, g is bounded if *and only if* its Fourier coefficients are absolutely summable; see [Kat04, §V.1.4].

Proof of Theorem 8.22. As already mentioned, the absolute summability of $\sum_{k=0}^{\infty} \frac{1}{a^k}$ for $a > 1$ implies that f is continuous and bounded. Let us assume that for some $x_0 \in \mathbb{R}$, the difference quotients

$$F(h) := D_h f(x_0) := \frac{f(x_0 + h) - f(x_0)}{h} \qquad \text{for} \qquad h \in \mathbb{R}\backslash\{0\}$$

have a well-defined limit $f'(x_0) = \lim_{h\to 0} F(h)$. Since f is bounded, it follows that F extends to a bounded continuous function on $\mathbb{R}$. We will show that this assumption implies $\lim_{k\to\infty}(b/a)^k = 0$, and thus $b < a$.

In order to estimate $(b/a)^k$ for large $k \in \mathbb{N}$, we will use the convolution formula (8.10). Choose a smooth function $\widehat{\psi} : \mathbb{R} \to [0,1]$ with $\widehat{\psi}(1) = 1$ and compact support in the interval $(1/b, b)$, and for each $k \in \mathbb{Z}$, let

$$\widehat{\psi}_k(p) := \widehat{\psi}(p/b^k),$$

which satisfies

$$\operatorname{supp}(\widehat{\psi}_k) \subset (b^{k-1}, b^{k+1}), \qquad \text{thus} \qquad \widehat{\psi}_k(b^n) = \begin{cases} 1 & \text{if } n = k, \\ 0 & \text{if } n \in \mathbb{Z}\backslash\{k\}. \end{cases}$$

Since $\widehat{\psi}_k \in C_0^\infty(\mathbb{R}) \subset \mathscr{S}(\mathbb{R})$, these functions are Fourier transforms of Schwartz-class functions $\psi_k \in \mathscr{S}(\mathbb{R})$, and an easy change of variables in the Fourier inversion formula gives

$$\psi_k(x) = \int_{-\infty}^{\infty} e^{2\pi i p x} \widehat{\psi}(p/b^k)\, dp = b^k \int_{-\infty}^{\infty} e^{2\pi i b^k p x} \widehat{\psi}(p)\, dp = b^k \psi(b^k x).$$

Notice also that since $0 \in \mathbb{R}$ lies outside the support of $\widehat{\psi}_k$ for each $k \in \mathbb{Z}$, we have

$$0 = \widehat{\psi}_k(0) = \int_{-\infty}^{\infty} \psi_k(x)\, dx \tag{8.13}$$

[20]Comeager sets are the complements of *meager* sets, which are countable unions of nowhere dense sets. Since there is no meaningful notion of "Lebesgue measure" on an infinite-dimensional vector space, meager sets often play the role of "sets of measure zero" in the infinite-dimensional context.

and

$$0 = \widehat{\psi}_k'(0) = -2\pi i \int_{-\infty}^{\infty} x\psi_k(x)\,dx. \tag{8.14}$$

The first of these two relations implies $\int_{-\infty}^{\infty} f(x_0)\psi_k(x)\,dx = 0$, so we now plug in (8.10) and compute:

$$\begin{aligned}\frac{1}{a^k}e^{2\pi i b^k x_0} &= (f * \psi_k)(x_0) = \int_{-\infty}^{\infty} f(x_0 - x)\psi_k(x)\,dx = \int_{-\infty}^{\infty} [f(x_0 - x) - f(x_0)]\,\psi_k(x)\,dx \\ &= -\int_{-\infty}^{\infty} xF(-x)\psi_k(x)\,dx = -b^k \int_{-\infty}^{\infty} xF(-x)\psi(b^k x)\,dx = -\int_{-\infty}^{\infty} \frac{x}{b^k} F(-x/b^k)\psi(x)\,dx,\end{aligned}$$

implying

$$\left(\frac{b}{a}\right)^k e^{2\pi i b^k x_0} = -\int_{-\infty}^{\infty} F(-x/b^k)x\psi(x)\,dx.$$

Since $\psi \in \mathscr{S}(\mathbb{R})$ and F is bounded, the integrand on the right hand side is bounded for every $k \geqslant 0$ by a constant times $|x|\psi \in L^1(\mathbb{R})$, and it converges pointwise as $k \to \infty$ to $F(0)x\psi(x) = f'(x_0)x\psi(x)$. Applying the dominated convergence theorem and the $k = 0$ case of (8.14), we conclude

$$\lim_{k\to\infty} \left(\frac{b}{a}\right)^k e^{2\pi i b^k x_0} = -f'(x_0)\int_{-\infty}^{\infty} x\psi(x)\,dx = 0,$$

thus $b < a$. $\square$

Exercise 8.23. Show that the Weierstrass function (8.11) with arbitrary constants $a, b > 1$ is of class C^m but has no derivative of order $m+1$ at any point, where $m \geqslant 0$ is the unique integer such that $m < \log_b a \leqslant m + 1$.

9. Sobolev spaces via Fourier analysis

9.1. **The general idea of Sobolev spaces.** In order to study PDEs via functional-analytic methods, one needs function spaces on which derivatives can be defined as bounded linear operators. For instance, the spaces $C^m(\mathbb{R}^n)$ and $C^m(\mathbb{T}^n)$ of bounded functions on $\mathbb{R}^n$ (or in the latter case fully periodic functions on $\mathbb{R}^n$) that have bounded partial derivatives up to order m is a Banach space with respect to the norm

$$\|f\|_{C^m} := \sum_{0 \leqslant |\alpha| \leqslant m} \sup_x |\partial^\alpha f(x)|, \tag{9.1}$$

where the sum ranges over all multi-indices of order at most m. For each $j = 1, \ldots, n$, the operation of taking the partial derivative with respect to coordinate x_j then defines a bounded linear operator

$$\partial_j : C^1(\mathbb{R}^n) \to C^0(\mathbb{R}^n) \quad \text{or} \quad \partial_j : C^1(\mathbb{T}^n) \to C^0(\mathbb{T}^n),$$

and similarly, any multi-index α of order $|\alpha| = m$ defines $\partial^\alpha : C^m(\mathbb{R}^n) \to C^0(\mathbb{R}^n)$ or $\partial^\alpha : C^m(\mathbb{T}^n) \to C^0(\mathbb{T}^n)$. That is all fine, but unfortunately the Banach spaces $C^m(\mathbb{R}^n)$ and $C^m(\mathbb{T}^n)$ do not have enough nice properties to be very useful in technical arguments. They are, for example, not reflexive, and their dual spaces are not easy to describe, e.g. by the *Riesz-Markov theorem* (see [Sal16, §3.3]), the dual of the space of continuous functions on a compact domain can be identified with a space of *measures*, which is inconveniently much larger than a space of functions. In this sense, the L^p-spaces are much nicer, but they have the obvious drawback that functions of class L^p are typically not even continuous, much less differentiable, so operators like ∂_j cannot be defined on $L^p(\mathbb{R}^n)$ or $L^p(\mathbb{T}^n)$.

The theory of Sobolev spaces, which is indispensable for the modern theory of PDEs, provides a means of keeping the good properties of the L^p-spaces while also permitting differentiation to be a bounded linear operator. Let us suppose first that we want to be able to handle first-order differential operators for functions on an open domain $\Omega \subset \mathbb{R}^n$. There are a few ways that one can imagine defining a suitable generalization of $L^p(\Omega)$ for this purpose:

Idea 1. Define $X_1(\Omega)$ to be the space of functions $f \in L^p(\Omega)$ that are differentiable almost everywhere and satisfy $\partial_j f \in L^p(\Omega)$ for every $j = 1, \ldots, n$. A natural choice of norm on this space is

$$\|f\|_{X_1} := \|f\|_{L^p} + \sum_{j=1}^n \|\partial_j f\|_{L^p}. \tag{9.2}$$

Unfortunately, it will turn out that this space is not complete, i.e. it is a reasonable normed vector space, but not a Banach space.

Idea 2. Since $X_1(\Omega)$ as defined above is not complete, one could define $X_2(\Omega)$ to be the closure of $X_1(\Omega) \subset L^p(\Omega)$ with respect to the X_1-norm. This is a reasonable definition, but not convenient to work with—we would prefer to be able to say precisely what the elements of $X_2(\Omega)$ are, rather than just calling it the closure of a dense subspace whose elements we can explicitly describe.

Idea 3. In the case $n = 1$ with $\Omega = (a, b) \subset \mathbb{R}$, one can consider the space $X_3(\Omega)$ of functions that have absolutely continuous extensions to $[a, b]$ such that their (almost everywhere defined) derivatives are of class L^p on (a, b). This is also a reasonable definition, but it only makes sense for functions of one real variable—on domains in $\mathbb{R}^n$, the notion of absolute continuity can be defined for measures, but not functions. It also doesn't give much of a hint how we should handle higher-order derivatives.

The general solution to these problems will be to generalize the notion of the derivative and thus talk about "weakly differentiable" functions; we will do this in §10 by introducing the theory of distributions. But before that, we observe that in the setting of $\Omega = \mathbb{R}^n$ with $p = 2$, a simpler solution is available using the properties of the Fourier transform.

9.2. **The spaces $H^m(\mathbb{R}^n)$ and $H^m(\mathbb{T}^n)$.** Let us start by writing down a norm that measures derivatives up to order $m \geqslant 0$ by integrating them instead of taking suprema (as the C^m-norm does). The case $m = 1$ appeared already in (9.2), and it generalizes naturally to

$$\|f\|_{W^{m,p}} := \sum_{|\alpha| \leqslant m} \|\partial^\alpha f\|_{L^p}, \tag{9.3}$$

where the summation ranges over all multi-indices α of order at most m; note that this includes the trivial multi-index with $|\alpha| = 0$, so the L^p-norm of f is one of the terms in the sum. If $p = 2$, we can use Plancherel's theorem and (8.4) to rewrite this norm as

$$\sum_{|\alpha| \leqslant m} \|(2\pi i p)^\alpha \hat{f}\|_{L^2}.$$

Up to equivalence of norms, the factors of $2\pi i$ in this expression clearly make no difference, and every monomial of order at most m satisfies $|p^\alpha| \leqslant c(1 + |p|^2)^{m/2}$ for some constant $c > 0$, thus an equivalent norm is given by the simpler expression

$$\|f\|_{H^m} := \left\|(1 + |p|^2)^{m/2} \hat{f}\right\|_{L^2} = \left(\int_{\mathbb{R}^n} (1 + |p|^2)^m |\hat{f}(p)|^2 \, dp\right)^{1/2} \in [0, \infty]. \tag{9.4}$$

Notice that this formula does not require f to be differentiable, nor even continuous; it is defined for *all* L^2-functions on $\mathbb{R}^n$, though we have no guarantee in general that it will be finite. Finiteness of this norm determines a subspace

$$H^m(\mathbb{R}^n) := \left\{ f \in L^2(\mathbb{R}^n) \;\middle|\; \|f\|_{H^m} < \infty \right\}.$$

We now observe two interesting things about this definition: first, it does not actually mention any derivatives of f, so $\|f\|_{H^m}$ might potentially be finite even if f is not differentiable or continuous. We plan to interpret $H^m(\mathbb{R}^n)$ nonetheless as the space of L^2-functions whose derivatives up to order m are also of class L^2, and this interpretation will turn out to be correct as soon as we enlarge our notion of what the word "derivative" can mean in §10. Second, the stated definition of the H^m-norm does not actually require m to be a nonnegative integer. It makes sense in fact for any $m \in \mathbb{R}$ as long as f belongs to a class of functions whose Fourier transforms can be defined, e.g. one can even allow $m < 0$ and drop the condition $f \in L^2$ by allowing f to be a so-called *tempered distribution* (see §10.6). We will not discuss the case $m < 0$ here, but the case of nonnegative real numbers other than integers gives rise to a notion of *fractional differentiability* that is sometimes useful in applications.

Theorem 9.1. *For every $m \geqslant 0$, $H^m(\mathbb{R}^n)$ is a Hilbert space with respect to the inner product*

$$\langle f, g\rangle_{H^m} := \int_{\mathbb{R}^n} (1 + |p|^2)^m \langle \hat{f}(p), \hat{g}(p)\rangle \, dp.$$

Proof. The map $H^m(\mathbb{R}^n) \to L^2(\mathbb{R}^n) : f \mapsto (1 + |p|^2)^{m/2} \hat{f}$ is a bijective isometry, so completeness of $H^m(\mathbb{R}^n)$ follows from completeness of $L^2(\mathbb{R}^n)$.[21] □

For fully periodic functions, there is a natural analogue of the space $H^m(\mathbb{R}^n)$ whose definition uses Fourier series instead of the Fourier transform. We define for each $f \in L^2(\mathbb{T}^n)$ the norm

$$\|f\|_{H^m} := \left\|(1 + |k|^2)^{m/2} \hat{f}\right\|_{\ell^2} = \left(\sum_{k \in \mathbb{Z}^n} (1 + |k|^2)^m |\hat{f}_k|^2\right)^{1/2} \in [0, \infty],$$

and set

$$H^m(\mathbb{T}^n) := \left\{ f \in L^2(\mathbb{T}^n) \;\middle|\; \|f\|_{H^m} < \infty \right\}.$$

The proof of the next statement is an easy adaptation of Theorem 9.1.

[21]Theorems 9.1 and 9.2 are also true and can be proved in the same way for $m < 0$, but we are not stating them for that case because our definition of H^m as a subspace of L^2 is only correct for $m \geqslant 0$. For a more general discussion, see e.g. [Tay96].

Theorem 9.2. *For every $m \geqslant 0$, $H^m(\mathbb{T}^n)$ is a Hilbert space with respect to the inner product*
$$\langle f, g\rangle_{H^m} := \sum_{k\in\mathbb{Z}^n} (1+|k|^2)^m \langle \widehat{f}_k, \widehat{g}_k\rangle.$$
□

Exercise 9.3. Show that $\mathscr{S}(\mathbb{R}^n) \subset H^m(\mathbb{R}^n)$ and $C^\infty(\mathbb{T}^n) \subset H^m(\mathbb{T}^n)$ for every $m \geqslant 0$.

By construction, any L^2-function f on $\mathbb{R}^n$ or $\mathbb{T}^n$ that has continuous derivatives up to order $m \geqslant 0$ which are also of class L^2 belongs to $H^s(\mathbb{R}^n)$ or $H^s(\mathbb{T}^n)$ respectively for every $s \leqslant m$. In particular, every smooth function with compact support is of class H^m for every m. If α is a multi-index with $|\alpha| \leqslant m$, then for f of class C_0^∞, (7.10) and (8.4) determine formulas for $\partial^\alpha f$ in terms of the Fourier series or transform of f. These formulas also make sense if f is not smooth but is of class H^m, and in this way one also obtains a bound on $\|\partial^\alpha f\|_{L^2}$ in terms of $\|f\|_{H^m}$, proving:

Proposition 9.4. *For any multi-index α of order $|\alpha| = m \in \mathbb{N}$, the operator ∂^α on smooth functions with compact support has a natural extension to a bounded linear map $\partial^\alpha : H^m(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ or $\partial^\alpha : H^m(\mathbb{T}^n) \to L^2(\mathbb{T}^n)$.* □

Exercise 9.5. Extend Proposition 9.4 to define ∂^α as a bounded linear map $H^{s+m}(\mathbb{R}^n) \to H^s(\mathbb{R}^n)$ or $H^{s+m}(\mathbb{T}^n) \to H^s(\mathbb{T}^n)$ for every $s \geqslant 0$ whenever $|\alpha| = m$.

Exercise 9.6. Show that for $a, b > 1$ with $b \in \mathbb{N}$, the Weierstrass function $f(x) = \sum_{k=0}^\infty \frac{1}{a^k} e^{2\pi i b^k x}$ belongs to $H^m(S^1) := H^m(\mathbb{T}^1)$ if and only if $m < \log_b a$.

For the Weierstrass functions, comparing Exercises 8.23 and 9.6 reveals a fairly straightforward correspondence: for each integer $m \geqslant 0$, $f \in C^m(S^1) := C^m(\mathbb{T}^1)$ if and only if $f \in H^m(S^1)$, and C^m-functions must also belong to $H^s(S^1)$ for some $s \in (m, m+1)$. In particular, f can be nowhere differentiable but will still belong to $H^s(S^1)$ for $s \in (0,1)$ sufficiently small. But the simplicity of this correspondence is slightly misleading. Beyond the Weierstrass functions, it cannot be true in general that every function of class H^m for an integer $m \geqslant 0$ is also of class C^m; this is clearly false for $m = 0$, since not all L^2-functions are continuous. The following exercises exhibit some less obvious examples.

Exercise 9.7. Two simple examples of discontinuous real-valued periodic functions on $\mathbb{R}$ are the *square* and *sawtooth* waves, defined respectiely as the obvious periodic extensions of
$$f(x) := \begin{cases} 1 & \text{if } 0 \leqslant x < 1/2, \\ -1 & \text{if } 1/2 \leqslant x < 1, \end{cases} \qquad \text{and} \qquad g(x) := x \text{ for } 0 \leqslant x < 1.$$
Show that both belong to $H^m(S^1)$ for all $m < 1/2$ but not for $m \geqslant 1/2$.

Exercise 9.8. The goal of this exercise is to show that the improper integral
$$f(x) := \int_2^\infty \frac{e^{2\pi i p x}}{p \ln p}\, dp := \lim_{N\to\infty} \int_2^N \frac{e^{2\pi i p x}}{p\ln p}\, dp, \qquad x \in \mathbb{R}\backslash\{0\} \tag{9.5}$$
defines a discontinuous function in $H^{1/2}(\mathbb{R})$. Note that the integrand $\frac{e^{2\pi i p x}}{p \ln p}$ is not a Lebesgue-integrable function of $p \in \mathbb{R}$, so the limit is necessary in order to define the integral, and its convergence is not obvious.

(a) Show that there exists a function $g \in L^2(\mathbb{R})$ whose Fourier transform is given almost everywhere by
$$\widehat{g}(p) = \begin{cases} \frac{1}{p \ln p} & \text{if } p \geqslant 2, \\ 0 & \text{if } p < 2, \end{cases}$$
and that this function belongs to $H^m(\mathbb{R})$ if and only if $m \leqslant 1/2$.

(b) Show that the function g in part (a) is the L^2-limit of the functions $f_N(x) := \int_2^N \frac{e^{2\pi i p x}}{p \ln p}\, dp$ as $N \to \infty$.

(c) Use integration by parts to prove that for every $M \geqslant 2$ and $x \in \mathbb{R}\backslash\{0\}$, the limit $\int_M^\infty \frac{e^{2\pi ipx}}{p\ln p}\,dp := \lim_{N\to\infty}\int_M^N \frac{e^{2\pi ipx}}{p\ln p}\,dp$ exists, depends continuously on x, and satisfies

$$\left|\int_M^\infty \frac{e^{2\pi ipx}}{p\ln p}\,dp\right| \leqslant \frac{1}{\pi|x|\cdot M\ln M}.$$

Deduce from this that the function g in part (a) matches (almost everywhere) the function f defined in (9.5), which is continuous on $\mathbb{R}\backslash\{0\}$.
Hint: Recall that L^2-convergence implies pointwise almost everywhere convergence of a subsequence.

(d) Prove that $\lim_{x\to 0}|f(x)| = \infty$.
Hint: Break up the integral over the intervals $[2,\epsilon/|x|]$ and $[\epsilon/|x|,\infty)$ for some small $\epsilon > 0$ with $|x| < \epsilon/2$. The estimate in part (c) will bound it on the second interval, while on the first, its absolute value should be larger than some positive multiple of $\int_2^{\epsilon/|x|}\frac{dp}{p\ln p}$ whenever ϵ is sufficiently small. Now let $|x|\to 0$ and use the fact that $\int_2^\infty \frac{dp}{p\ln p} = \infty$.

Exercise 9.9. Adapt the argument of Exercise 9.8 to show that the L^2-convergent Fourier series $f(x) := \sum_{k=2}^\infty \frac{e^{2\pi ikx}}{k\ln k}$ defines a discontinuous function in $H^{1/2}(S^1) := H^{1/2}(\mathbb{T}^1)$.
Hint: Proving a bound on $\left|\sum_{k=M}^\infty \frac{e^{2\pi ikx}}{k\ln k}\right|$ for $x \neq 0$ requires an analogue of integration by parts for summations, which is easy to prove if you regard the "derivative" of a sequence a_k as the sequence $a_k' := a_{k+1} - a_k$. If you need more inspiration, see [Rud76, pp. 70–71].

9.3. **The Sobolev embedding theorem.** Exercises 9.7, 9.8 and 9.9 demonstrate that functions of class H^m for $m \leqslant 1/2$ on S^1 or $\mathbb{R}$ need not be continuous, though it seems that discontinuous examples for the case $m = 1/2$ are not so easy to construct. We will now show that it becomes impossible for $m > 1/2$, and in fact, such a threshold also exists for functions on $\mathbb{T}^n$ or $\mathbb{R}^n$ and depends on the dimension n. Recall that $H^m(\mathbb{R}^n)$ and $H^m(\mathbb{T}^n)$ were defined as subspaces of $L^2(\mathbb{R}^n)$ and $L^2(\mathbb{T}^n)$ respectively, so their elements are not actually functions, but rather *equivalence classes* of functions defined almost everywhere. This is different from the Banach spaces $C^m(\mathbb{R}^n)$ and $C^m(\mathbb{T}^n)$, whose elements are actual functions. We will say that there exists a **continuous inclusion**

$$H^s(\mathbb{R}^n) \hookrightarrow C^m(\mathbb{R}^n)$$

whenever the following is true: every $f \in H^s(\mathbb{R}^n)$ is equal almost everywhere to a unique function $\tilde{f} \in C^m(\mathbb{R}^n)$, and the resulting map $H^s(\mathbb{R}^n) \to C^m(\mathbb{R}^n) : f \mapsto \tilde{f}$ is a bounded linear operator. The existence of a continuous inclusion thus comes with an estimate of the form

$$\|f\|_{C^m} \leqslant c\|f\|_{H^s} \quad \text{for some constant } c > 0 \text{ independent of } f,$$

where we abuse notation by forgetting the distinction between the C^m-function f and the equivalence class in $H^s(\mathbb{R}^n)$ that it represents. There is an obvious similar definition for the spaces of fully periodic functions $H^s(\mathbb{T}^n)$ and $C^m(\mathbb{T}^n)$.

Theorem 9.10 (Sobolev embedding theorem, case $p = 2$)**.** *Assume $n \in \mathbb{N}$ and $s > 0$ satisfy $2s > n$. Then there exist continuous inclusions*

$$H^{s+m}(\mathbb{R}^n) \hookrightarrow C^m(\mathbb{R}^n) \quad \text{and} \quad H^{s+m}(\mathbb{T}^n) \hookrightarrow C^m(\mathbb{T}^n)$$

for every integer $m \geqslant 0$.

Proof. We first consider functions $f \in H^s(\mathbb{R}^n)$ with $2s > n$. The main step is to establish a bound on $\|\hat{f}\|_{L^1}$, as f is then equal almost everywhere to $\mathscr{F}^*\hat{f}$, which is continuous since $\mathscr{F}^*$

defines a bounded linear operator $L^1(\mathbb{R}^n) \to C^0(\mathbb{R}^n)$. We use the Cauchy-Schwarz inequality:

$$\begin{aligned} \|\widehat{f}\|_{L^1} &= \int_{\mathbb{R}^n} \frac{1}{(1+|p|^2)^{s/2}} \cdot \left|(1+|p|^2)^{s/2}\widehat{f}\right| dp \leqslant \left\|\frac{1}{(1+|p|^2)^{s/2}}\right\|_{L^2} \cdot \left\|(1+|p|^2)^{s/2}\widehat{f}\right\|_{L^2} \\ &\leqslant \left(\int_{\mathbb{R}^n} \frac{1}{(1+|p|^2)^s}\, dp\right)^{1/2} \cdot \|f\|_{H^s} \end{aligned}$$

Using n-dimensional polar coordinates, we see that the integral in the second line converges if and only if $\int_1^\infty \frac{r^{n-1}}{(1+r^2)^s}\, dr < \infty$. For large $r > 0$, the latter integrand behaves like $r^{n-1}/r^{2s} = r^{n-2s-1}$, so the integral converges if and only if $n - 2s < 0$, which is exactly the condition $2s > n$. This proves the continuous inclusion of $H^s(\mathbb{R}^n)$ into $C^0(\mathbb{R}^n)$.

If $f \in H^{s+m}(\mathbb{R}^n)$ with $m \in \mathbb{N}$, then the same argument bounds the L^1-norm of the function $p \mapsto p^\alpha \widehat{f}(p)$ for each multi-index α with $|\alpha| \leqslant m$ in terms of $\|f\|_{H^{s+m}}$, so the argument of Exercise 8.11 shows that the partial derivatives $\partial^\alpha f$ up to order m exist and are continuous. Moreover, their C^0-norms are bounded in terms of the L^1-norm of $p^\alpha \widehat{f}$, which gives a bound for $\|f\|_{C^m}$ in terms of $\|f\|_{H^{s+m}}$.

The result for fully periodic functions follows by essentially the same argument, except that the version of the Cauchy-Schwarz inequality one needs is $\|fg\|_{\ell^1} \leqslant \|f\|_{\ell^2} \cdot \|g\|_{\ell^2}$ for functions $f, g : \mathbb{Z}^n \to [0, \infty)$. The crucial detail is then the convergence of the series

$$\sum_{k \in \mathbb{Z}^n} \frac{1}{(1+|k|^2)^s} < \infty \quad \text{for } 2s > n,$$

which can be established by comparing it with the integral $\int_{\mathbb{R}^n} \frac{1}{(1+|p|^2)^s}\, dp$. □

Corollary 9.11. *Any function belonging to $H^s(\mathbb{R}^n)$ for all $s \geqslant 0$ is (after changing its values on a set of measure zero) smooth, and its derivatives of all orders are bounded. Similarly, $\bigcap_{s \geqslant 0} H^s(\mathbb{T}^n) = C^\infty(\mathbb{T}^n)$.* □

Theorem 9.10 leads to the intuition that functions of class H^s have "$s - \frac{n}{2}$ continuous derivatives," where in general the number $s - n/2$ need not be an integer, but should be assumed positive in order for the statement to carry any meaning. We will make this more precise for the case $0 < s - n/2 < 1$ in §9.6.

9.4. **Compact inclusions.** A much more obvious fact than the Sobolev embedding theorem is that for every $t > s \geqslant 0$, there are continuous inclusions $H^t(\mathbb{R}^n) \hookrightarrow H^s(\mathbb{R}^n)$ and $H^t(\mathbb{T}^n) \hookrightarrow H^s(\mathbb{T}^n)$. If we think of functions of class H^s as being $(s - n/2)$-times differentiable, then these inclusions are analogous to the obvious continuous inclusions $C^m \hookrightarrow C^k$ for $m > k$. Let us focus for this subsection on fully periodic functions, which can be regarded as functions on the compact metric space $\mathbb{T}^n$. One interesting fact about the inclusion $C^m(\mathbb{T}^n) \hookrightarrow C^k(\mathbb{T}^n)$ for $m > k$ is that it is a compact operator. A bounded linear operator $A : X \to Y$ between Banach spaces is called a **compact operator** if it maps every bounded subset of X to a precompact subset of Y, or equivalently, for every bounded sequence $x_n \in X$, the sequence $Ax_n \in Y$ has a convergent subsequence. Such compactness properties furnish a favorite source of existence results in applications, e.g. if one can find a sequence of functions that approximate solutions to a PDE arbitrarily well, then a convergent subsequence can be expected to have a limit that is an exact solution.

Exercise 9.12. Use the Arzelà-Ascoli theorem to show that for any integers $m > k \geqslant 0$, the inclusion $C^m(\mathbb{T}^n) \hookrightarrow C^k(\mathbb{T}^n)$ is a compact operator.
Hint: Compositions of compact operators are also compact, thus it suffices to prove that $C^{k+1}(\mathbb{T}^n) \hookrightarrow C^k(\mathbb{T}^n)$ is compact for every $k \geqslant 0$. Start with $k = 0$, and notice that any C^1-bounded sequence in $C^1(\mathbb{T}^n)$ is equicontinuous.

Exercise 9.13. Find a bounded sequence in $C^1(\mathbb{R}^n)$ that converges pointwise to 0 but does not have any C^0-convergent subsequence.
Hint: Translations!

The analogue of Exercise 9.12 for Sobolev spaces is known as the *Rellich-Kondrachov compactness theorem*:

Theorem 9.14 (Rellich-Kondrachov for $p = 2$)**.** *For every $t > s \geqslant 0$, the natural inclusion $H^t(\mathbb{T}^n) \hookrightarrow H^s(\mathbb{T}^n)$ is compact.*

Exercise 9.15. Adapt Exercise 9.13 to show that the inclusion $H^s(\mathbb{R}^n) \hookrightarrow L^2(\mathbb{R}^n)$ for $s > 0$ is not compact; in particular, there exists an H^s-bounded sequence that converges pointwise to 0 but stays a fixed positive distance away from 0 in the L^2-norm, thus it has no L^2-convergent subsequence.

To prove Theorem 9.14, we will appeal to two very useful general facts about compact operators. The first concerns bounded linear operators with finite rank:

Proposition 9.16. *If X and Y are Banach spaces and $A : X \to Y$ is a bounded linear operator with finite-dimensional image, then A is compact.*

Proof. The image of a bounded sequence $x_n \in X$ is a bounded sequence $Ax_n \in Y$, but by assumption it also belongs to a finite-dimensional subspace $\operatorname{im} A \subset Y$. The result thus follows from the fact that all bounded sequences in finite-dimensional vector spaces have convergent subsequences. □

Proposition 9.17. *If X and Y are Banach spaces and $A_n : X \to Y$ is a sequence of compact operators that converge in the operator norm to an operator $A : X \to Y$, then A is also compact.*

Proof. Suppose $x_n \in X$ is a bounded sequence. Since $A_1 : X \to Y$ is compact, x_n has a subsequence $x_n^{(1)}$ such that $A_1 x_n^{(1)}$ converges. We can then use the compactness of A_2 to extract from $x_n^{(1)}$ a further subsequence $x_n^{(2)}$ such that $A_2 x_n^{(2)}$ converges. Continuing in this manner, one obtains a sequence of subsequences $x_n^{(j)}$ such that $A_j x_n^{(j)}$ converges as $n \to \infty$ for every $j \in \mathbb{N}$. The diagonal subsequence

$$x_n^{(\infty)} := x_n^{(n)}$$

then has the property that $A_j x_n^{(\infty)}$ converges as $n \to \infty$ for every j.

We claim now that $Ax_n^{(\infty)}$ also converges, which will imply that $A : X \to Y$ is compact. Since Y is complete, it suffices to show that $Ax_n^{(\infty)}$ is a Cauchy sequence. Given $\epsilon > 0$, choose $M \in \mathbb{N}$ such that

$$\|A - A_M\| < \frac{\epsilon}{3} \sup_{n \in \mathbb{N}} \|x_n\|,$$

and then choose $N \in \mathbb{N}$ such that $\|A_M x_n^{(\infty)} - A_M x_m^{(\infty)}\| < \epsilon/3$ for all $m, n \geqslant N$; the latter is possible since $A_M x_n^{(\infty)}$ is a Cauchy sequence. It follows that for all $m, n \geqslant N$,

$$\|Ax_n^{(\infty)} - Ax_m^{(\infty)}\| \leqslant \|(A - A_M)x_n^{(\infty)}\| + \|A_M(x_n^{(\infty)} - x_m^{(\infty)})\| + \|(A_M - A)x_m^{(\infty)}\| \leqslant \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,$$

thus proving the claim. □

Corollary 9.18. *Any bounded linear operator in the closure (with respect to the operator norm) of the space of finite-rank operators is compact.* □

Remark 9.19. We will not need this at present, but if X is a separable Hilbert space, then the converse of Corollary 9.18 is also true for bounded linear operators $X \to X$, i.e. they are compact if and only if they can be approximated arbitrarily well in the operator norm by operators with finite rank. The proof is not hard; see [RS80, Theorem VI.13].

Proof of Theorem 9.14. Fix $t > s \geqslant 0$, and consider for each $N \in \mathbb{N}$ the operator

$$A_N : H^t(\mathbb{T}^n) \to H^s(\mathbb{T}^n) : f \mapsto \sum_{|k| \leqslant N} e^{2\pi i k \cdot x} \widehat{f}_k.$$

The image of A_N is finite dimensional since there are only finitely many lattice points $k \in \mathbb{Z}^n$ satisfying $|k| \leqslant N$. The goal is now to show that A_N converges in the operator norm as $N \to \infty$

to the inclusion $A : H^t(\mathbb{T}^n) \hookrightarrow H^s(\mathbb{T}^n)$, hence the latter is a limit of finite-rank operators and is therefore compact.

To prove $\|A - A_N\| \to 0$, we observe that for each $f \in H^t(\mathbb{T}^n)$, the functions $(A - A_N)f$ have the same Fourier coefficients as f except that every coefficient for $k \in \mathbb{Z}^n$ with $|k| \leqslant N$ is set to zero, hence

$$\begin{aligned}\|(A - A_N)f\|_{H^s}^2 &= \sum_{|k|>N} (1 + |k|^2)^s |\widehat{f}_k|^2 = \sum_{|k|>N} \frac{1}{(1 + |k|^2)^{t-s}} (1 + |k|^2)^t |\widehat{f}_k|^2 \\ &\leqslant \frac{1}{(1 + N^2)^{t-s}} \sum_{k \in \mathbb{Z}^n} (1 + |k|^2)^t |\widehat{f}_k|^2 = \frac{1}{(1 + N^2)^{t-s}} \cdot \|f\|_{H^t}^2.\end{aligned}$$

This proves $\|A - A_N\|^2 \leqslant \frac{1}{(1+N^2)^{t-s}}$, and the latter converges to 0 as $N \to \infty$ since $t > s$. ☐

9.5. **Approximation by smooth functions.** The following result says that $H^s(\mathbb{T}^n)$ and $H^s(\mathbb{R}^n)$ could just as well have been defined as the closures of the subspaces $C^\infty(\mathbb{T}^n) \subset L^2(\mathbb{T}^n)$ and $\mathscr{S}(\mathbb{R}^n) \subset L^2(\mathbb{R}^n)$ with respect to the H^s-norm. As a first application, it implies that for each $s \geqslant 0$ and each multi-index α with order $|\alpha| = m$, the extension of the classical differential operator ∂^α to a bounded linear operator $H^{s+m}(\mathbb{T}^n) \to H^s(\mathbb{T}^n)$ or $H^{s+m}(\mathbb{R}^n) \to H^s(\mathbb{R}^n)$ is unique (cf. Proposition 9.4 and Exercise 9.5).

Theorem 9.20. *The subspaces $C^\infty(\mathbb{T}^n) \subset H^s(\mathbb{T}^n)$ and $\mathscr{S}(\mathbb{R}^n) \subset H^s(\mathbb{R}^n)$ are dense for every $s \geqslant 0$.*

Proof. We begin with the easiest case: suppose $f \in H^s(\mathbb{T}^n)$, and for $j \in \mathbb{N}$, let

$$f_j(x) := \sum_{|k| \leqslant j} e^{2\pi i k \cdot x} \widehat{f}_j.$$

Since f is a finite sum of smooth functions, it is smooth, and we have

$$\|f - f_j\|_{H^s}^2 = \sum_{|k|>j} (1 + |k|^2)^s |\widehat{f}_k|^2 \to 0 \quad \text{as} \quad j \to \infty$$

since $\sum_{k \in \mathbb{Z}^n} (1 + |k|^2)^s |\widehat{f}_k|^2 = \|f\|_{H^s}^2 < \infty$.

For $f \in H^s(\mathbb{R}^n)$, we have $(1+|p|^2)^{s/2} \widehat{f} \in L^2(\mathbb{R}^n)$, and the density of $C_0^\infty(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$ implies that there exists a sequence $h_j \in C_0^\infty(\mathbb{R}^n)$ with

$$h_j \xrightarrow{L^2} (1 + |p|^2)^{s/2} \widehat{f}.$$

The functions $g_j(p) := \frac{h_j(p)}{(1+|p|^2)^{s/2}}$ are then also in $C_0^\infty(\mathbb{R}^n)$, and they satisfy

$$(1 + |p|^2)^{s/2} g_j \xrightarrow{L^2} (1 + |p|^2)^{s/2} \widehat{f}. \tag{9.6}$$

Since $C_0^\infty(\mathbb{R}^n) \subset \mathscr{S}(\mathbb{R}^n)$, each g_j is then the Fourier transform of a unique function $f_j \in \mathscr{S}(\mathbb{R}^n)$, and (9.6) implies $f_j \to f$ in H^s. ☐

Remark 9.21. A stronger result is true for functions on $\mathbb{R}^n$: the space of smooth *compactly supported* functions $C_0^\infty(\mathbb{R}^n)$, which is a subspace of $\mathscr{S}(\mathbb{R}^n)$, is also dense in $H^s(\mathbb{R}^n)$. For a proof of this in the more general setting of $W^{m,p}$-spaces (assuming $m \in \mathbb{Z}$), see [AF03, Theorem 3.22].

Recall from §5 that the density of smooth functions in L^p is proved by taking convolutions of $f \in L^p(\mathbb{R}^n)$ with an approximate identity ρ_j, a trick often referred to as **mollification**. For most purposes, Theorem 9.20 can also be placed into this context: for instance, the approximating sequence $f_j \to f \in H^s(\mathbb{T}^n)$ in the proof above was constructed by defining its Fourier coefficients to be $\widehat{f}_j = \chi_{\bar{B}_j} \widehat{f} : \mathbb{Z}^n \to V$, where $\chi_{\bar{B}_j} : \mathbb{Z}^n \to [0, 1]$ denotes the characteristic function of the intersection of $\mathbb{Z}^n$ with the closed ball of radius j in $\mathbb{R}^n$. Clearly $\chi_{\bar{B}_j} \in \mathscr{S}(\mathbb{Z}^n)$, so $\chi_{\bar{B}_j}$ defines the Fourier coefficients of a smooth function, namely

$$\rho_j(x) := \sum_{|k| \leqslant j} e^{2\pi i k \cdot x}.$$

Since this function belongs to $L^1(\mathbb{T}^n)$, Exercise 8.20 implies

$$f_j = \rho_j * f.$$

For $f \in H^s(\mathbb{R}^n)$, if we wanted to approximate f with smooth functions in $H^s(\mathbb{R}^n)$ but did not care whether they are rapidly decreasing, we could use a similar trick:

Exercise 9.22. Suppose $\rho \in \mathscr{S}(\mathbb{R}^n)$ satisfies $\int_{\mathbb{R}^n} \rho(x)\, dx = 1$, and define $\rho_j(x) := j^n \rho(jx)$.

(a) Show that for any $s \geqslant 0$ and $f \in H^s(\mathbb{R}^n)$, the sequence $\rho_j * f \in C^\infty(\mathbb{R}^n)$ satisfies

$$\|\rho_j * f\|_{H^s} \leqslant \|f\|_{H^s} \qquad \text{and} \qquad \rho_j * f \xrightarrow{H^s} f \text{ as } j \to \infty.$$

Hint: Compute $\widehat{\rho}_j$ in terms of $\widehat{\rho}$, then use change of variables and dominated convergence to prove $\|f - \rho_j * f\|_{H^s} \to 0$.

(b) Show that the same result holds if $\rho_j \in \mathscr{S}(\mathbb{R}^n)$ is instead defined as $\check{\psi}_j$ for a sequence of smooth functions $\psi_j : \mathbb{R}^n \to [0,1]$ with compact support in B_{j+1} and $\psi_j|_{B_j} \equiv 1$.

Exercise 9.23. Suppose α is a multi-index of order $|\alpha| = m \in \mathbb{N}$.

(a) Use the definition of $\partial^\alpha : H^m(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ in Proposition 9.4 to prove that for every $\psi \in L^1(\mathbb{R}^n)$ and $f \in H^m(\mathbb{R}^n)$, $\partial^\alpha(\psi * f) = \psi * \partial^\alpha f \in L^2(\mathbb{R}^n)$.

(b) Use the result of part (a) to give an alternative proof that for any $f \in H^m(\mathbb{R}^n)$ with $m \in \mathbb{N}$ and any approximate identity ρ_j as in §5.4, $\rho_j * f \to f$ in H^m.

9.6. **Hölder estimates.** The compactness of the inclusions $H^t(\mathbb{T}^n) \hookrightarrow H^s(\mathbb{T}^n)$ has an interesting consequence related to the Sobolev embedding theorem: if $2s > n$, then there also exists some $t < s$ such that $2t > n$, and the continuous inclusion $H^{s+m}(\mathbb{T}^n) \hookrightarrow C^m(\mathbb{T}^n)$ thus factors into a composition of two inclusions,

$$H^{s+m}(\mathbb{T}^n) \hookrightarrow H^{t+m}(\mathbb{T}^n) \hookrightarrow C^m(\mathbb{T}^n).$$

The first of these is compact, and therefore so is the composition:[22]

Corollary 9.24. *For $2s > n$, the continuous inclusions $H^{s+m}(\mathbb{T}^n) \hookrightarrow C^m(\mathbb{T}^n)$ in Theorem 9.10 are also compact.* □

There is a second way to see the compactness of $H^{s+m}(\mathbb{T}^n) \hookrightarrow C^m(\mathbb{T}^n)$ that provides more information, while also yielding a practical interpretation of the motto that functions in $H^s(\mathbb{T}^n)$ are "$(s - n/2)$-times differentiable".

Assume Ω is a measurable subset of either $\mathbb{R}^n$ or $\mathbb{T}^n$, regarded in each case as a metric space with metric denoted by $\operatorname{dist}(x,y) = |x-y|$. Recall that a function $f : \Omega \to V$ is called **Lipschitz continuous** if there exists a constant $C > 0$ such that

$$|f(x) - f(y)| \leqslant C|x-y| \quad \text{for all } x, y \in \Omega.$$

For example, a continuously differentiable function on an open domain $\mathcal{U} \subset \mathbb{R}^n$ is Lipschitz continuous on every subset $\Omega \subset \mathcal{U}$ on which the partial derivatives are bounded. Classic examples of non-Lipschitz continuous functions include $f(x) := |x|^\alpha$ for $0 < \alpha < 1$ on any neighborhood of $0 \in \mathbb{R}$. These instead satisfy the following condition, which is the same as Lipschitz contintuity for $\alpha = 1$, but weaker for $0 < \alpha < 1$:

[22] Lemma: Any composition of a compact operator with a bounded linear operator is compact. Proof: Easy exercise.

Definition 9.25. A function f on $\Omega \subset \mathbb{R}^n$ is called **Hölder continuous** if there exists a number $\alpha \in (0, 1]$ and a constant $C > 0$ such that

$$|f(x) - f(y)| \leqslant C|x - y|^\alpha \quad \text{for all } x, y \in \Omega.$$

The space of Hölder continuous functions on Ω with fixed **Hölder exponent** $\alpha \in (0, 1]$ is denoted by $C^{0,\alpha}(\Omega)$.

Hölder continuity can be quantified by the **Hölder seminorms**, defined for each $\alpha \in (0, 1]$ by

$$|f|_{C^{0,\alpha}} := \sup_{x \neq y \in \Omega} \frac{|f(x) - f(y)|}{|x - y|^\alpha},$$

thus a continuous function is α-Hölder continuous if and only if $|f|_{C^{0,\alpha}} < \infty$. This is a seminorm rather than a norm since it vanishes for constant functions, even if they are nonzero. A norm on the space $C^{0,\alpha}(\Omega)$ can then be defined by

$$\|f\|_{C^{0,\alpha}} := \|f\|_{C^0} + |f|_{C^{0,\alpha}}.$$

Exercise 9.26. Prove:

(a) $|\cdot|_{C^{0,\alpha}}$ is a seminorm.
(b) If $f_n \in C^{0,\alpha}(\Omega)$ converges uniformly to $f \in C^0(\Omega)$ and satisfies a uniform bound $|f_n|_{C^{0,\alpha}} \leqslant C$ for all n, then $f \in C^{0,\alpha}(\Omega)$.
(c) The norm $\|\cdot\|_{C^{0,\alpha}}$ on $C^{0,\alpha}(\Omega)$ is complete, i.e. $C^{0,\alpha}(\Omega)$ is a Banach space.
Hint: Show that if f_n is C^0-convergent to f and $|f_n - f_m|_{C^{0,\alpha}} < \epsilon$ holds for all $m, n \geqslant N$, then $|f - f_n|_{C^{0,\alpha}} \leqslant \epsilon$ holds for all $n \geqslant N$. Here is a start:

$$|(f - f_n)(x) - (f - f_n)(y)| \leqslant |(f - f_m)(x)| + |(f_m - f_n)(x) - (f_m - f_n)(y)| + |(f_m - f)(y)|.$$

Keep in mind that after fixing $n \geqslant N$ and $x \neq y$, m can be chosen arbitrarily large.

For functions that can be written down in simple formulas, it is typically easy to prove a $C^{0,1}$-bound by differentiating and bounding the derivative. As the example of the Weierstrass function in §8.8 shows, this trick cannot be relied upon for functions that arise as uniform limits of sequences. This is precisely the situation in which one often encounters functions that are Hölder but not necessarily Lipschitz continuous, and the following lemma provides a useful tool to recognize this.

Lemma 9.27. *Suppose f_k is a sequence of continuous functions on $\Omega \subset \mathbb{R}^n$ converging uniformly to f, and there exist constants $a > 1$, $b \geqslant 1$, $C > 0$ and $\beta \in (0, 1]$ such that*

$$\|f - f_k\|_{C^0} \leqslant \frac{C}{a^k} \qquad \textit{and} \qquad |f_k|_{C^{0,\beta}} \leqslant Cb^k.$$

Then $f \in C^{0,\alpha}(\Omega)$ for $\alpha := \dfrac{\beta}{1 + \log_a b}$.

Exercise 9.28. Fill in the gaps in the following proof of Lemma 9.27. The estimate $|f(x) - f(y)| \leqslant C|x - y|^\alpha$ only needs to be proved for all $x, y \in \Omega$ with $0 < |x - y| \leqslant c$ for some constant $c > 0$. For any $k \in \mathbb{N}$, we have

$$|f(x) - f(y)| \leqslant |f(x) - f_k(x)| + |f_k(x) - f_k(y)| + |f_k(y) - f(y)| \leqslant \frac{2C}{a^k} + Cb^k|x - y|^\beta$$

for all $x, y \in \Omega$. Assuming $0 < |x - y| \leqslant c$ for some $c > 0$ sufficiently small, choose $k \in \mathbb{N}$ such that $\frac{1}{(ab)^{k+1}} \leqslant |x - y|^\beta \leqslant \frac{1}{(ab)^k}$. Use this to show $|f(x) - f(y)| \leqslant \frac{3aC}{a^{k+1}}$, and then use the identity $a^{1 + \log_a b} = ab$.

Exercise 9.29. The Cantor function $f : [0, 1] \to \mathbb{R}$ from Example 6.2 satisfies $f(1/3^n) = 1/2^n$ for every $n \in \mathbb{N}$. Use this to prove $f \notin C^{0,\alpha}([0, 1])$ for $\alpha > \log_3 2$. Then show that the C^0-convergent sequence f_n in Example 6.2 satisfies $|f_n|_{C^{0,1}} = (3/2)^n$, and use it to prove $f \in C^{0,\alpha}([0, 1])$ for all $\alpha \leqslant \log_3 2$.

Exercise 9.30. For any $\theta \in (0,1)$, there is a distinguished set $C_\theta \subset [0,1]$ of full measure such that $C_{1/3}$ is the usual Cantor ternary set: it is constructed by an inductive procedure in which at step $n \in \mathbb{N}$, one removes from the middle of each of 2^{n-1} intervals of identical lengths l_n an open interval of length θl_n. Follow this idea to its logical conclusion in order to prove the following statement: for every $\alpha_0 \in (0,1)$, there exists a surjective increasing function $f : [0,1] \to [0,1]$ such that $f \in C^{0,\alpha}([0,1])$ if and only if $\alpha \leqslant \alpha_0$, and f has vanishing derivative almost everywhere (and is therefore not absolutely continuous).

Exercise 9.31. Show that for $b \geqslant a > 1$, the Weierstrass function $f(x) = \sum_{k=0}^\infty \frac{1}{a^k} e^{2\pi i b^k x}$ belongs to $C^{0,\alpha}(\mathbb{R})$ for every $\alpha \in (0,1)$ with $\alpha \leqslant \log_b a$.
Remark: f is nowhere differentiable by Theorem 8.22, so it cannot be absolutely continuous and therefore (by Exercise 6.6) cannot be Lipschitz, even if $\log_b a = 1$.

Exercise 9.32. Suppose $g : [0,\infty) \to \mathbb{R}$ is a strictly increasing smooth function with $g^{(k)}(0) = 0$ for all $k \geqslant 0$, e.g. one can take $g(x) = e^{-1/x^2}$ for $x > 0$ and $g(0) = 0$. There is a unique extension of g to an odd function $\mathbb{R} \to \mathbb{R}$, which is also strictly increasing and continuous, so it admits a continuous inverse $f := g^{-1} : I \to \mathbb{R}$ on a sufficiently small interval $I = [-a,a]$, $a > 0$. Prove that f is absolutely continuous on I, but does not belong to $C^{0,\alpha}(I)$ for any $\alpha \in (0,1]$.
Hint: The vanishing of $g^{(k)}(0)$ implies an estimate of the form $|x|^{1/k} \leqslant c_k |f(x)|$ for some constant $c_k > 0$. For absolute continuity, prove directly that f satisfies the fundamental theorem of calculus, starting from the fact that this is clearly true on any interval not containing 0.

If the domain $\Omega \subset \mathbb{R}^n$ is open, then we can also discuss differentiability of functions on Ω and define for C^m-functions the norm

$$\|f\|_{C^{m,\alpha}} := \|f\|_{C^m} + \sum_{|\beta|=m} |\partial^\beta f|_{C^{0,\alpha}},$$

where the sum ranges over all multi-indices β of order m. This norm is finite if and only if f is of class C^m with bounded and α-Hölder continuous partial derivatives up to order m. (Note that the Hölder continuity of derivatives of order less than m follows already from the fact that derivatives of higher order are bounded, so the norm does not need to include any terms $|\partial^\beta f|_{C^{0,\alpha}}$ with $|\beta| < m$.) The space of functions satisfying this condition is denoted by

$$C^{m,\alpha}(\Omega) \subset C^m(\Omega).$$

Exercise 9.33. Prove that $C^{m,\alpha}(\Omega)$ is a Banach space for every integer $m \geqslant 0$ and $\alpha \in (0,1]$.

Exercise 9.34. Use the Arzelà-Ascoli theorem to prove that if Ω is an open subset of $\mathbb{R}^n$ or $\mathbb{T}^n$ with compact closure, then for every $\alpha \in (0,1]$, the obvious continuous inclusion

$$C^{0,\alpha}(\Omega) \hookrightarrow C^0(\Omega)$$

is compact. Then generalize by induction to the statement that for each integer $m \geqslant 0$ and $\alpha \in (0,1]$, the inclusion

$$C^{m,\alpha}(\Omega) \hookrightarrow C^m(\Omega)$$

is compact.

Exercise 9.35. Extend Exercise 9.34 to show that under the same assumption on Ω, for every integer $m \geqslant 0$ and $0 < \alpha < \beta \leqslant 1$, the obvious inclusion

$$C^{m,\beta}(\Omega) \hookrightarrow C^{m,\alpha}(\Omega)$$

is compact.
Hint: For $m = 0$, Exercise 9.34 guarantees for any $C^{0,\beta}$-bounded sequence a C^0-convergent subsequence, and Exercise 9.26 then implies that the limit is also of class $C^{0,\beta}$, though the convergence need not be in the $C^{0,\beta}$-topology. To show that the subsequence is $C^{0,\alpha}$-convergent for $\alpha < \beta$, the following relation can help:

$$\frac{|f(x) - f(y)|}{|x-y|^\alpha} = \left(\frac{|f(x) - f(y)|}{|x-y|^\beta} \right)^{\alpha/\beta} \cdot |f(x) - f(y)|^{1-\frac{\alpha}{\beta}}.$$

Exercise 9.34 holds in particular for fully periodic functions on $\mathbb{R}^n$ since $\mathbb{T}^n$ is compact. Thus Corollary 9.24 can now be seen as a consequence of the following enhancement of the Sobolev embedding theorem (Theorem 9.10):

Theorem 9.36. *Assume $n \in \mathbb{N}$, $s > 0$ and $\alpha \in (0,1)$ satisfy $\alpha \leqslant s - \frac{n}{2}$. Then there exist continuous inclusions*

$$H^{s+m}(\mathbb{R}^n) \hookrightarrow C^{m,\alpha}(\mathbb{R}^n) \quad \text{and} \quad H^{s+m}(\mathbb{T}^n) \hookrightarrow C^{m,\alpha}(\mathbb{T}^n)$$

for every integer $m \geqslant 0$.

Remark 9.37. Note that Theorem 9.36 only gives us something new when $0 < s - n/2 \leqslant 1$, as the case $s - n/2 > 1$ is already handled by Theorem 9.10, which gives an inclusion $H^{s+m} \hookrightarrow C^{m+1}$ and therefore also into $C^{m,\alpha}$ for every $\alpha \in (0,1]$. In the case $s - n/2 = 1$, one should be careful to note that α is not allowed to equal 1, so we are not claiming anything about an inclusion $H^{s+m} \hookrightarrow C^{m,1}$. We will point out the specific step in the proof below that would fail if $\alpha = 1$, and an actual counterexample to the statement for this case may be found in Example 9.38.

Proof of Theorem 9.36. We will establish the inclusion $H^s(\mathbb{R}^n) \hookrightarrow C^{0,\alpha}(\mathbb{R}^n)$ for $\alpha \in (0,1)$ with $\alpha \leqslant s - n/2$ and leave the remaining cases as exercises. In light of the inclusions $H^t \hookrightarrow H^s$ for $t > s$, we can assume

$$0 < s - n/2 = \alpha < 1$$

without loss of generality. Then Theorem 9.10 already implies that $f \in H^s(\mathbb{R}^n)$ is continuous and satisfies an estimate of the form $\|f\|_{C^0} \leqslant C\|f\|_{H^s}$, so our remaining task is to find a similar bound for its Hölder seminorm $|f|_{C^{0,\alpha}}$. In other words, we need to find a constant $C > 0$ independent of $f \in H^s(\mathbb{R}^n)$ such that

$$|f(x+y) - f(x)| \leqslant C\|f\|_{H^s} \cdot |y|^\alpha \quad \text{for all} \quad x, y \in \mathbb{R}^n \text{ with } y \neq 0.$$

The proof of Theorem 9.10 shows that $\widehat{f} \in L^1(\mathbb{R}^n)$, thus we can write down the usual integral formula for f in terms of $\widehat{f}$ and use the assumption $\|f\|_{H^s} = \|(1+|p|^2)^{s/2}\widehat{f}\|_{L^2} < \infty$ to apply the Cauchy-Schwarz inequality:

(9.7)
$$\begin{aligned}
|f(x+y) - f(x)| &= \left| \int_{\mathbb{R}^n} e^{2\pi i p\cdot(x+y)} \widehat{f}(p)\,dp - \int_{\mathbb{R}^n} e^{2\pi i p\cdot x}\widehat{f}(p)\,dp \right| \leqslant \int_{\mathbb{R}^n} \left|e^{2\pi i p\cdot y} - 1\right| \cdot |\widehat{f}(p)|\,dp \\
&= \int_{\mathbb{R}^n} \frac{\left|e^{2\pi i p\cdot y} - 1\right|}{(1+|p|^2)^{s/2}} \cdot (1+|p|^2)^{s/2}|\widehat{f}|\,dp \leqslant \left(\int_{\mathbb{R}^n} \frac{\left|e^{2\pi i p\cdot y} - 1\right|^2}{(1+|p|^2)^s}\,dp \right)^{1/2} \cdot \|f\|_{H^s} \\
&\leqslant \left(\int_{\mathbb{R}^n} \frac{\left|e^{2\pi i p\cdot y} - 1\right|^2}{|p|^{2s}}\,dp \right)^{1/2} \cdot \|f\|_{H^s}
\end{aligned}$$

To estimate the integral in the last line, we first observe that the function $\mathbb{R} \to \mathbb{C} : t \mapsto e^{2\pi i t}$ has globally bounded derivative $2\pi i e^{2\pi i t}$ and thus satisifes $|e^{2\pi i t} - 1| \leqslant 2\pi|t|$ for all $t \in \mathbb{R}$, implying

$$\left|e^{2\pi i p\cdot y} - 1\right| \leqslant 2\pi|p \cdot y| \leqslant 2\pi|p| \cdot |y|. \tag{9.8}$$

Now partition $\mathbb{R}^n$ into the domains

$$E_0 := \left\{p \in \mathbb{R}^n \;\middle|\; |p| \leqslant 1/|y|\right\} \quad \text{and} \quad E_\infty := \left\{p \in \mathbb{R}^n \;\middle|\; |p| > 1/|y|\right\},$$

and let $\mathrm{Vol}(S^{n-1})$ denote the $(n-1)$-dimensional volume of the unit sphere in $\mathbb{R}^n$. Integrating in n-dimensional polar coordinates then gives

$$\begin{aligned}\int_{E_0} \frac{\left|e^{2\pi i p\cdot y} - 1\right|^2}{|p|^{2s}}\, dp &\leqslant 4\pi^2|y|^2 \int_{E_0} \frac{1}{|p|^{2s-2}}\, dp = 4\pi^2 \mathrm{Vol}(S^{n-1}) \cdot |y|^2 \int_0^{1/|y|} \frac{r^{n-1}}{r^{2s-2}}\, dr \\ &= 4\pi^2 \mathrm{Vol}(S^{n-1}) \cdot |y|^2 \int_0^{1/|y|} r^{n-2s+1}\, dr = \frac{4\pi^2 \mathrm{Vol}(S^{n-1})}{n-2s+2} \cdot |y|^2 \frac{1}{|y|^{n-2s+2}} \\ &= \frac{2\pi^2 \mathrm{Vol}(S^{n-1})}{1-\alpha} \cdot |y|^{2\alpha},\end{aligned}$$

where the convergence of $\int_0^{1/|y|} r^{n-2s+1}\, dr$ relies on the assumption $n - 2s + 2 = 2(1-\alpha) > 0$. (This step in the proof would fail if we allowed $\alpha = 1$.) On E_∞, the estimate (9.8) is not useful since $|p|$ may be large, so instead we use the simpler estimate $|e^{2\pi i p\cdot y} - 1| \leqslant 2$ arising from the triangle inequality, and the convergence of the integral will depend on the assumption $n - 2s = -2\alpha < 0$:

$$\begin{aligned}\int_{E_\infty} \frac{\left|e^{2\pi i p\cdot y} - 1\right|^2}{|p|^{2s}}\, dp &\leqslant 4 \int_{E_\infty} \frac{1}{|p|^{2s}}\, dp = 4\,\mathrm{Vol}(S^{n-1}) \int_{1/|y|}^\infty \frac{r^{n-1}}{r^{2s}}\, dr = 4\,\mathrm{Vol}(S^{n-1}) \int_{1/|y|}^\infty r^{n-1-2s}\, dr \\ &= \frac{4\,\mathrm{Vol}(S^{n-1})}{n-2s} r^{n-2s}\Big|_{r=1/|y|}^{r=\infty} = \frac{4\,\mathrm{Vol}(S^{n-1})}{2\alpha}\frac{1}{|y|^{-2\alpha}} = \frac{2\,\mathrm{Vol}(S^{n-1})}{\alpha}|y|^{2\alpha}.\end{aligned}$$

Putting both pieces of the integral together gives an estimate $\int_{\mathbb{R}^n} \frac{\left|e^{2\pi i p\cdot y} - 1\right|^2}{|p|^{2s}}\, dp \leqslant c|y|^{2\alpha}$ for a suitable constant $c > 0$, so plugging this into (9.7) gives the result we were hoping for. ☐

Example 9.38. Let $f(x) := \sum_{k=2}^\infty \frac{e^{2\pi i k x}}{k^2 \ln k}$. Up to multiplication by a constant, differentiating this series term by term gives the Fourier series of Exercise 9.9, so $f \in H^{3/2}(S^1)$, and Theorem 9.36 thus implies $f \in C^{0,\alpha}(S^1)$ for every $\alpha \in (0,1)$. One can also show as in Exercise 9.8 that the differentiated series converges uniformly on compact subsets of $\{x \neq 0\}$, thus f is continuously differentiable on $S^1 \backslash \{0\}$. But its derivative blows up at $x = 0$, showing that $f \notin C^{0,1}(S^1)$.

Remark 9.39. One should not assume that the constants in Theorem 9.36 are always optimal. Consider for instance the Weierstrass function $f(x) = \sum_{k=0}^\infty \frac{1}{a^k} e^{2\pi i b^k x}$ for $b \in \mathbb{N}$ with $1 < \sqrt{b} < a \leqslant b$. According to Exercise 9.6, $f \in H^s(S^1)$ if and only if $s < \log_b a$. Since $a > \sqrt{b}$, this range includes values $s > 1/2$, so Theorem 9.36 implies $f \in C^{0,\alpha}(S^1)$ for all $\alpha < \log_b a - \frac{1}{2}$. But in fact, Exercise 9.31 shows that $f \in C^{0,\alpha}(S^1)$ for all $\alpha < \log_b a$.

9.7. **Elliptic regularity.** To demonstrate the power of the Fourier transform and Sobolev spaces, in this section we shall give a brief taste of the theory of elliptic PDEs.

To understand the goal, consider first a second-order ordinary differential equation of the form

$$\ddot{x}(t) = F(x(t), \dot{x}(t)) \tag{9.9}$$

for paths $x : (-\epsilon, \epsilon) \to \mathbb{R}^n$, where $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is a function of class C^m, $1 \leqslant m \leqslant \infty$. A solution to this equation must by definition be twice differentiable at every point, but it is easy to see that it must in fact be better, i.e. more "regular," which for this discussion you can take to be a synonym for "smoother". Indeed, if $\ddot{x}$ always exists, then x and $\dot{x}$ are both continuous, and (9.9) thus implies that $\ddot{x}$ is continuous, hence x is of class C^2 and $\dot{x}$ is of class C^1. Since we also assumed $F \in C^1$, thus implies $F \circ (x, \dot{x}) \in C^1$ and therefore $\ddot{x} \in C^1$, so x is of class C^3. One can repeat this argument until F runs out of derivatives. The conclusion is that if the data in the equation is of class C^m, then any solution must be at least two steps more regular, namely of class C^{m+2}; in particular, if F is smooth, then so is x. This is true even though the equation itself makes sense for any function x that is everywhere twice differentiable.

The following example shows that *partial* differential equations, by contrast, do not always have this "regularizing" property.

Example 9.40. The simplest version of the **wave equation** is the second-order PDE

$$\partial_t^2 u - \partial_x^2 u = 0$$

for a function $u : \mathbb{R}^2 \to \mathbb{R}$ of two variables $(t, x) \in \mathbb{R}^2$. For any C^2-function $f : \mathbb{R} \to \mathbb{R}$, the wave equation has solutions given by

$$u(t, x) := f(t \pm x).$$

Notice that although the wave equation is linear with constant (and thus smooth) coefficients, its solutions need not be smooth; the function $f \in C^2(\mathbb{R})$ can be chosen arbitrarily, and the solution u will then have only as many derivatives as f does.

There is a special class of PDEs, called *elliptic*, that do exhibit the same regularizing behavior as ODEs. For this discussion, we shall only consider the simplest and most popular example: the **Poisson equation**

$$\Delta f := \sum_{j=1}^{n} \partial_j^2 f = g,$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a given function, and the solution is meant to be a function $f : \mathbb{R}^n \to \mathbb{R}$. The second-order differential operator $-\Delta := -\sum_{j=1}^n \partial_j^2$ is called the **Laplacian**, and arises often in physics (e.g. in the study of electrostatic or gravitational potentials), as well as in differential geometry.[23] We shall consider the Poisson equation on the torus $\mathbb{T}^n$, that is, we assume $g : \mathbb{R}^n \to \mathbb{R}$ is a fully periodic function and consider solutions $f : \mathbb{R}^n \to \mathbb{R}$ that are also fully periodic.

Theorem 9.41. *For any integer $m \geqslant 0$ and smooth function $g : \mathbb{T}^n \to \mathbb{R}$, all C^2-solutions $f : \mathbb{T}^n \to \mathbb{R}$ to the equation $\Delta f = g$ are also smooth.*

I would encourage the reader at this point to take out a piece of paper and consider whether Theorem 9.41 might be proved by some trick as simple as the ODE discussion at the top of this subsection. You will quickly run into difficulties, because the Laplace operator Δ gives us information about a particular linear combination of second partial derivatives of a solution f, but we cannot deduce from this anything about any individual partial derivative. From this perspective, Theorem 9.41 is a very surprising result. It follows from the next theorem, which is of a slightly more technical nature since it involves Sobolev spaces. To prepare the statement, observe that by Proposition 9.4, Δ defines a bounded linear operator

$$\Delta : H^2(\mathbb{T}^n) \to L^2(\mathbb{T}^n),$$

which is defined in the obvious way on the dense subspace $C^\infty(\mathbb{T}^n)$ but requires Fourier transforms in order to define Δf for $f \in H^2(\mathbb{T}^n) \backslash C^2(\mathbb{T}^n)$. Recall that functions in $H^2(\mathbb{T}^n)$ need not be twice differentiable in general; when $n > 3$, they need not even be continuous (cf. Theorem 9.10).

Theorem 9.42. *If $m \in \mathbb{N}$ and $f \in H^2(\mathbb{T}^n)$ satisfies $\Delta f \in H^m(\mathbb{T}^n)$, then $f \in H^{m+2}(\mathbb{T}^n)$.*

To prove Theorem 9.41 from this statement, observe that if $g \in C^m(\mathbb{T}^n)$, then $g \in H^m(\mathbb{T}^n)$ since g has derivatives up to order m that are continuous, and therefore also in $L^2(\mathbb{T}^n)$. If $f \in C^2(\mathbb{T}^n)$ satisfies $\Delta f = g \in C^\infty(\mathbb{T}^n)$, it follows that $f \in H^2(\mathbb{T}^n)$ and $\Delta f \in H^m(\mathbb{T}^n)$ for every $m \in \mathbb{N}$. Theorem 9.42 then implies $f \in H^{m+2}(\mathbb{T}^n)$, thus f belongs to *all* of the Sobolev spaces $H^s(\mathbb{T}^n)$ for $s \geqslant 0$, and is therefore smooth according to the Sobolev embedding theorem (Theorem 9.10).

Proof of Theorem 9.42. Suppose $f \in H^2(\mathbb{T}^n)$ and $\Delta f = g \in H^m(\mathbb{T}^n)$ for $m \in \mathbb{N}$. The Fourier coefficients of f and g are then related by

$$\widehat{\Delta f}_k = \sum_{j=1}^{n} \widehat{\partial_j^2 f}_k = \sum_{j=1}^{n} (2\pi i k_j)^2 \hat{f}_k = -4\pi^2 |k|^2 \hat{f}_k = \hat{g}_k$$

[23]The minus sign in the definition of the Laplace operator appears in some sources and not in others. It is appropriate if one wants to consider the *spectrum* of the operator: the minus sign ensures that all of its eigenvalues are positive. For our present discussion this makes no difference.

for all $k \in \mathbb{Z}^n$. For $s \leqslant m+2$, this implies

$$\sum_{k\in\mathbb{Z}^n\setminus\{0\}} |k|^{2s}|\widehat{f}_k|^2 = C \sum_{k\in\mathbb{Z}^n\setminus\{0\}} \frac{|k|^{2s}}{|k|^4}|\widehat{g}_k|^2 = C \sum_{k\in\mathbb{Z}^n\setminus\{0\}} |k|^{2(s-2)}|\widehat{g}_k|^2 \leqslant C \sum_{k\in\mathbb{Z}^n\setminus\{0\}} |k|^{2m}|\widehat{g}_k|^2$$
$$\leqslant C\|g\|^2_{H^m}$$

for a suitable constant $C > 0$, thus $f \in H^{m+2}(\mathbb{T}^n)$. □

10. Distributions

Throughout this section, we assume unless stated otherwise that

$$\Omega \subset \mathbb{R}^n$$

is an open subset, and we again consider functions on Ω with values in an arbitrary finite-dimensional inner product space $(V, \langle\ ,\ \rangle)$ over the field $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. At the beginning of §9, we heuristically sketched the definition of a Banach space $W^{m,p}(\Omega)$ consisting of functions in L^p that have derivatives up to order m also in L^p. Here we assume $m \geqslant 0$ is an integer and $1 \leqslant p \leqslant \infty$. The quickest rigorous definition of this space is as the closure with respect to the $W^{m,p}$**-norm**

$$\|f\|_{W^{m,p}} := \sum_{|\alpha| \leqslant m} \|\partial^\alpha f\|_{L^p}, \tag{10.1}$$

of the space of all smooth functions $f : \Omega \to V$ with $\|f\|_{W^{m,p}} < \infty$. There is nothing wrong with defining $W^{m,p}(\Omega)$ in this way, but it leaves open the question of precisely which functions actually belong to $W^{m,p}(\Omega)$. For $p = 2$ and $\Omega = \mathbb{R}^n$, we found an elegant solution to the this problem in §9 by using the Fourier transform to identify differentiation with the operation of multiplication by a polynomial, so that the space $H^m(\mathbb{R}^n) := W^{m,2}(\mathbb{R}^n)$ could be defined without having to explicitly differentiate its elements. We also saw that functions of class H^m really need not be m times differentiable, e.g. Example 9.38 describes a function in $H^1(S^1)$ that is of class C^1 on the complement of one point but has its derivative blowing up at that point. To understand this phenomenon properly in the cases $p \ne 2$ or $\Omega \subsetneq \mathbb{R}^n$ where the Fourier transform is not available, we need a new trick for talking about derivatives of functions that might not be classically differentiable.

10.1. Weak derivatives.

The trick we have in mind arises from the following straightforward exercise combining Fubini's theorem with integration by parts:

Exercise 10.1. Show that if $f : \Omega \to V$ and $\varphi : \Omega \to \mathbb{K}$ are functions of class C^1 and φ has compact support in Ω, then for each $j = 1, \dots, n$,

$$\int_\Omega \varphi \cdot \partial_j f \, dm = -\int_\Omega \partial_j \varphi \cdot f \, dm.$$

Hint: The function φf has an obvious extension to a C^1-function on $\mathbb{R}^n$ that vanishes outside of Ω. Compute $\int_{\mathbb{R}^n} \partial_j(\varphi f)\, dm$.

In this exercise, requiring φ to have compact support ensures on the one hand that $\varphi \cdot \partial_j f$ and $\partial_j \varphi \cdot f$ are both Lebesgue-integrable functions, and it also eliminates the boundary terms that would otherwise appear when carrying out integration by parts. The resulting formula can be used to uniquely characterize the partial derivatives of f: namely, if $f : \Omega \to V$ and $g : \Omega \to V$ are functions of class C^1 and C^0 respectively such that

$$\int_\Omega \varphi g \, dm = -\int_\Omega \partial_j \varphi \cdot f \, dm \qquad \text{for all} \qquad \varphi \in C_0^\infty(\Omega), \tag{10.2}$$

then $g = \partial_j f$. Indeed, $h := g - \partial_j f$ is then a continuous function on Ω satisfying

$$\int_\Omega \varphi h = 0 \qquad \text{for all} \qquad \varphi \in C_0^\infty(\Omega),$$

and if $h(x) \ne 0$ for some $x \in \Omega$, then the latter relation is contradicted by any smooth bump function φ that satisfies $\varphi(x) = 1$ and vanishes outside a sufficiently small neighborhood of x.

Notice: the condition (10.2) does not explicitly mention any derivative of f. In fact, both sides of the relation are well defined as soon as f and g are locally integrable functions on Ω.

Definition 10.2. A function $f \in L^1_{\mathrm{loc}}(\Omega)$ is said to be **weakly differentiable** if there exist functions $g_1, \dots, g_n \in L^1_{\mathrm{loc}}(\Omega)$ such that for each $j = 1, \dots, n$,

$$\int_\Omega \varphi g_j \, dm = -\int_\Omega \partial_j \varphi \cdot f \, dm \qquad \text{for all} \qquad \varphi \in C_0^\infty(\Omega).^{24}$$

We then call g_j a **weak partial derivative** of f with respect to the variable x_j, and write $\partial_j f := g_j$.

Three important remarks should be understood immediately:

(1) If f is of class C^1, then its classical partial derivatives are also weak partial derivatives, thus for this class of functions there is no ambiguity in denoting weak derivatives by $\partial_j f$. (There will occasionally be ambiguity if we talk about functions that are differentiable almost everywhere—these sometimes also have weak derivatives, but sometimes they do not.)

(2) In contrast with classical derivatives, weak derivatives are well defined only up to equality almost everywhere, i.e. if $g = \partial_j f$ weakly and $h = g$ almost everywhere, then h is also a weak derivative of f. Similarly, f can be changed on a set of measure zero without changing its weak derivatives.

(3) Related to the second point: weak differentiability is a property of the whole function $f \in L^1_{\mathrm{loc}}(\Omega)$, and it is not purely *local*, i.e. it generally makes no sense to ask whether f is weakly differentiable at an individual point $x \in \Omega$, nor what the value of $\partial_j f(x)$ is, though one *can* ask what $\int_E \partial_j f \, dm$ is for any given measurable subset $E \subset \Omega$.

Since weak derivatives of locally integrable functions are also locally integrable functions, one can iterate the definition in obvious ways to define higher-order weak differentiability and weak derivatives $\partial^\alpha f$, which will be uniquely characterized by the relation

$$\int_\Omega \varphi \cdot \partial^\alpha f \, dm = (-1)^{|\alpha|} \int_\Omega \partial^\alpha \varphi \cdot f \, dm \qquad \text{for all} \qquad \varphi \in C_0^\infty(\Omega).$$

There is again no problem in making sense of this condition since φ is always assumed to be infinitely differentiable with compact support; we only need f and $\partial^\alpha f$ to be of class L^1_{loc}.

Let us clarify that a function may indeed have a weak derivative without being classically differentiable:

Exercise 10.3. Show that the function $f : \mathbb{R} \to \mathbb{R} : x \mapsto |x|$ has weak derivative $f'(x) := x/|x|$. (It is not necessary to specify a value for $f'(0)$ since $\{0\} \subset \mathbb{R}$ is a set of measure zero.) Then show that $f' \in L^1_{\mathrm{loc}}(\mathbb{R})$ is not weakly differentiable.

For functions that are not of class C^1, we have not yet shown that weak derivatives are uniquely defined almost everywhere, but this is true, and follows from:

Lemma 10.4. *If $f \in L^1_{\mathrm{loc}}(\Omega)$ satisfies $\int \varphi f \, dm = 0$ for every smooth compactly supported function $\varphi : \Omega \to \mathbb{R}$, then $f = 0$ almost everywhere.*

Proof. Given $x_0 \in \Omega$, choose $\epsilon > 0$ small enough so that the closed ϵ-ball $\bar{B}_\epsilon(x_0)$ about x_0 lies in Ω, and consider the function $g \in L^1(\mathbb{R}^n)$ defined by

$$g := \begin{cases} f & \text{on } \overline{B_\epsilon(x_0)}, \\ 0 & \text{everywhere else.} \end{cases}$$

Choose an approximate identity $\rho_j : \mathbb{R}^n \to [0, \infty)$ with shrinking support. For $x \in B_{\epsilon/2}(x_0)$ and j sufficiently large so that $\mathrm{supp}(\rho_j) \subset B_{\epsilon/2}(0)$, the function $\rho_j(x - \cdot) : \mathbb{R}^n \to [0, \infty)$ then has

[24] The function $\varphi \in C_0^\infty(\Omega)$ in this definition can be taken to have either real or complex values; it does not matter.

compact support in $B_{\epsilon/2}(x) \subset B_\epsilon(x_0)$ and can therefore be regarded as an element of $C_0^\infty(\Omega)$, implying that the convolution

$$(\rho_j * g)(x) = \int_{\mathbb{R}^n} \rho_j(x-y)g(y)\,dy = \int_{B_{\epsilon/2}(x)} \rho_j(x-y)g(y)\,dy = \int_\Omega \rho_j(x-\cdot)f\,dm$$

vanishes for $x \in B_{\epsilon/2}(x_0)$. By Theorem 5.14, $\rho_j * g \to g$ in $L^1(\mathbb{R}^n)$ as $j \to \infty$, so we conclude that g (and therefore f) vanishes almost everywhere on $B_{\epsilon/2}(x_0)$. Since Ω can be covered by countably many subsets of the form $B_{\epsilon/2}(x_0)$ with $x_0 \in \Omega$ and $\epsilon > 0$, it follows that f vanishes almost everywhere in Ω. ☐

Corollary 10.5. *If $f \in L^1_{\mathrm{loc}}(\Omega)$ is weakly differentiable, then its weak partial derivatives are unique up to equality almost everywhere.* ☐

Exercise 10.6. Consider the Cantor function f from Example 6.2 on the domain $\Omega := (0,1) \subset \mathbb{R}$, which has classical derivative $f' = 0$ at almost every point. Show however that f is not weakly differentiable.
Hint: Show first that if a weak derivative existed, it would necessarily vanish almost everywhere on each of the intervals that are removed to form the Cantor set.

10.2. **Test functions and the space of distributions.** Let us fit the notion of weak derivatives into a larger context. We saw in Exercise 10.3 that locally integrable functions can be weakly differentiable without being classically differentiable, but also that not *all* functions in L^1_{loc} have weak derivatives. We will now see that if our notion of what a "function" can be is suitably enlarged, then *every* L^1_{loc} function can be understood to have a derivative in some sense.

The key observation is that for weak differentiation, what matters is not the values of a function $f : \Omega \to V$ at points in Ω, but rather the values of the linear map

$$\Lambda_f : C_0^\infty(\Omega) \to V : \varphi \mapsto \int_\Omega \varphi f.$$

This suggests that instead of talking about functions on Ω, we should talk about linear maps $C_0^\infty(\Omega) \to V$, e.g. in the case $V = \mathbb{K}$, we are talking about the dual space of $C_0^\infty(\Omega)$. To do this properly, we should consider only linear functionals that are continuous, which requires endowing $C_0^\infty(\Omega)$ with a topology.

Definition 10.7. A **test function** on Ω is defined to be a smooth function $\varphi : \Omega \to \mathbb{K}$ with compact support, and the vector space of all such functions is denoted by $\mathscr{D}(\Omega)$. A sequence $\varphi_j \in \mathscr{D}(\Omega)$ is said to **converge** to $\varphi_\infty \in \mathscr{D}(\Omega)$ if there exists a compact subset $K \subset \Omega$ such that φ_j has support contained in K for every $j \in \mathbb{N} \cup \{\infty\}$ and $\partial^\alpha \varphi_j$ converges uniformly to $\partial^\alpha \varphi_\infty$ for every multi-index α. A $\mathbb{K}$-linear map $\Lambda : \mathscr{D}(\Omega) \to V$ is then said to be **continuous** if and only if $\Lambda(\varphi_j) \to \Lambda(\varphi_\infty)$ for every convergent sequence $\varphi_j \to \varphi_\infty \in \mathscr{D}(\Omega)$.

Putting Definition 10.7 on firm mathematical footing requires the following result, whose proof is outsourced to §10.8 in order to avoid too much of a digression into abstract topology:

Proposition 10.8 (see §10.8)**.** *The space of test functions $\mathscr{D}(\Omega)$ admits a natural topology that induces the notions of convergence and continuity described in Definition 10.7.*

Definition 10.9. A scalar-valued **distribution** on Ω is a continuous $\mathbb{K}$-linear functional $\Lambda : \mathscr{D}(\Omega) \to \mathbb{K}$. Similarly, a vector-valued distribution with values in the finite-dimensional vector space V over $\mathbb{K}$ is a continuous $\mathbb{K}$-linear map $\Lambda : \mathscr{D}(\Omega) \to V$. We shall generally assume that all our distributions take values in a fixed vector space V, and denote the the space of vector-valued distributions by

$$\mathscr{D}'(\Omega) = \left\{ \Lambda : \mathscr{D}(\Omega) \to V \;\middle|\; \Lambda \text{ is } \mathbb{K}\text{-linear and continuous} \right\}.$$

The space $\mathscr{D}'(\Omega)$ is endowed with the weak*-topology, i.e. the locally convex topology generated by the seminorms $\|\Lambda\|_\varphi := |\Lambda(\varphi)|$ for all $\varphi \in \mathscr{D}(\Omega)$. In particular, a sequence $\Lambda_j \in \mathscr{D}'(\Omega)$ converges to $\Lambda_\infty \in \mathscr{D}'(\Omega)$ if and only if $\Lambda_j(\varphi) \to \Lambda_\infty(\varphi)$ for every $\varphi \in \mathscr{D}(\Omega)$.

Remark 10.10. If V is a complex vector space, then one can regard it as a real vector space (of twice the dimension) and set $\mathbb{K} = \mathbb{R}$ without changing any result in the theory of distributions. The reason is that every real-linear map from the space of real-valued test functions to a complex vector space has a unique complex-linear extension to the space of complex-valued test functions. Thus for most purposes, it makes no difference whether we set $\mathbb{K}$ to be $\mathbb{R}$ or $\mathbb{C}$, and many books on distributions treat only the case $\mathbb{K} = \mathbb{R}$. We will need to set $\mathbb{K} = \mathbb{C}$ however when we discuss Fourier transforms in §10.6.

Note that choosing a basis of V identifies each vector-valued distribution with a finite tuple of scalar-valued distributions, just as for vector-valued functions. Since the choice of the space V almost never plays an important role in our discussion, we shall suppress it from the notation whenever possible.

Example 10.11. There is a natural linear map

$$L^1_{\mathrm{loc}}(\Omega) \to \mathscr{D}'(\Omega) : f \mapsto \Lambda_f, \qquad \Lambda_f(\varphi) := \int_\Omega \varphi f \, dm,$$

and by Lemma 10.4, this map is injective. (Exercise: check that $\Lambda_f : \mathscr{D}(\Omega) \to V$ is continuous.) In this way, every locally integrable function determines a distribution, and we shall often abuse terminology by identifying one with the other, e.g. when we say "$\Lambda \in \mathscr{D}'(\Omega)$ is a function," we mean that there exists a (necessarily unique up to equality almost everywhere) function $f \in L^1_{\mathrm{loc}}(\Omega)$ such that $\Lambda = \Lambda_f$.

Convention. We will sometimes also denote the action of a distribution $\Lambda \in \mathscr{D}'(\Omega)$ on test functions $\varphi \in \mathscr{D}(\Omega)$ by

$$(\Lambda, \varphi) := \Lambda(\varphi),$$

and abbreviate the case of a locally integrable function $f \in L^1_{\mathrm{loc}}(\Omega)$ by

$$(f, \varphi) := \Lambda_f(\varphi) := \int_\Omega \varphi f \, dm.$$

Exercise 10.12. Show that the map $L^1_{\mathrm{loc}}(\Omega) \to \mathscr{D}'(\Omega)$ in Example 10.11 is continuous, where $L^1_{\mathrm{loc}}(\Omega)$ is endowed with the Fréchet space topology defined in §0.3.

Example 10.13. The most popular scalar-valued distribution that is not a function is what physicists call the **Dirac delta function**: for each $x \in \Omega$, we define $\delta_x \in \mathscr{D}'(\Omega)$ by

$$\delta_x(\varphi) := \varphi(x).$$

On $\Omega = \mathbb{R}^n$, one typically abbreviates $\delta := \delta_0$ for the δ-function centered at the origin, so that pretending δ is an actual function on $\mathbb{R}^n$ gives rise to the usual formula $\int_{\mathbb{R}^n} \varphi(x)\delta(x)\, dx = \varphi(0)$. A formal change of variables transforms this into $\delta_x(\varphi) = \varphi(x) = \int_{\mathbb{R}^n} \varphi(y + x)\delta(y)\, dy = \int_{\mathbb{R}^n} \varphi(u)\delta(u - x)\, du$, motivating the notation

$$\delta(\cdot - x) := \delta_x \in \mathscr{D}'(\Omega).$$

Example 10.14. Suppose μ is a measure defined on the Borel subsets of $\Omega \subset \mathbb{R}^n$ such that $\mu(K) < \infty$ whenever $K \subset \Omega$ is compact. Then $\Lambda(\varphi) := \int_\Omega \varphi \, d\mu$ defines a real-valued distribution. The distributions Λ_f in Examples 10.11 (with $f : \Omega \to [0, \infty)$) and δ_x in Example 10.13 are both special cases of this, with measures defined by

$$\mu(E) := \int_E f \, dm \qquad \text{and} \qquad \mu(E) := \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{otherwise} \end{cases}$$

respectively. The latter is of course also known as the **Dirac measure** centered at x.

Example 10.15. Here is a distribution that is not a special case of Example 10.14: for $k \in \mathbb{N}$ and $x \in \Omega \subset \mathbb{R}$, define

$$\Lambda(\varphi) := \varphi^{(k)}(x).$$

Exercise 10.16. Verify that the linear maps $\mathscr{D}(\Omega) \to V$ described in Examples 10.11, 10.13, 10.14 and 10.15 are all continuous.

The trick via integration by parts in the definition of weak differentiation now generalizes as follows.

Definition 10.17. Given $\Lambda \in \mathscr{D}'(\Omega)$ and $j = 1, \ldots, n$, the **distributional derivative** (or derivative "in the sense of distributions") of Λ with respect to the variable x_j is a distribution $\partial_j \Lambda \in \mathscr{D}'(\Omega)$ defined by

$$(\partial_j \Lambda)(\varphi) := -\Lambda(\partial_j \varphi).$$

More generally, if α is any multi-index with order $|\alpha| \geqslant 0$, one defines $\partial^\alpha \Lambda \in \mathscr{D}'(\Omega)$ by

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

It is easy to check that the distributions in Definition 10.17 are always well defined continuous linear maps, so every distribution is infinitely differentiable, and the operators $\partial^\alpha : \mathscr{D}'(\Omega) \to \mathscr{D}'(\Omega)$ are continuous linear maps. In this language, a function is weakly differentiable if and only if its derivative in the sense of distributions is also represented by a function. For functions of class C^1, the distributional derivatives can always be represented by the classical derivatives.

Exercise 10.18. The function $f(x) := x/|x|$ appeared in Exercise 10.3 as the weak derivative of the function $|x|$. Show that the derivative of f in the sense of distributions (meaning the derivative of the distribution Λ_f) is 2δ.

Example 10.19. Up to a sign, the distribution in Example 10.15 is the kth derivative of the δ-function: concretely, $\Lambda = (-1)^k \delta_x^{(k)}$.

When we talk about distributions represented by functions, we typically assume these functions to be locally integrable so that expressions like $\int_\Omega \varphi f \, dm$ make sense for all $\varphi \in \mathscr{D}(\Omega)$. This is not always strictly necessary, however: the next exercise exhibits a locally integrable function that is not weakly differentiable in the sense of Definition 10.2, but its distributional derivative can (with a little care) be represented by a function that is not of class L^1_{loc}.

Exercise 10.20. Show that the function $f(x) := \ln |x|$ is locally integrable on $\mathbb{R}$, and its derivative in $\mathscr{D}'(\mathbb{R})$ is given by

$$\Lambda_f'(\varphi) = \mathrm{p.\,v.} \int_{\mathbb{R}} \frac{\varphi(x)}{x} \, dx := \lim_{\epsilon \to 0^+} \int_{|x| \geqslant \epsilon} \frac{\varphi(x)}{x} \, dx.$$

Here the notation p. v. stands for "Cauchy principal value" and is defined as the limit on the right. Check that this expression gives a well-defined distribution even though $1/x$ is not a locally integrable function on $\mathbb{R}$.

The product of a distribution $\Lambda \in \mathscr{D}'(\Omega)$ with a smooth scalar-valued function $f \in C^\infty(\Omega)$ defines a distribution $f\Lambda \in \mathscr{D}'(\Omega)$ via the obvious formula

$$(f\Lambda)(\varphi) := \Lambda(f\varphi).$$

This is well defined because $\varphi \mapsto f\varphi$ defines a continuous map $\mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$; note that this depends on f having derivatives of all orders (though it does not need to have compact support), so the product of an arbitrary distribution Λ with a non-smooth function is not well defined in general. It is easy to check that for every $f \in C^\infty(\Omega)$, the linear map $\mathscr{D}'(\Omega) \to \mathscr{D}'(\Omega) : \Lambda \mapsto f\Lambda$ is also continuous.

Exercise 10.21. Show that for $f \in C^\infty(\Omega)$ and $\Lambda \in \mathscr{D}'(\Omega)$, distributional derivatives satisfy the Leibniz rule

$$\partial_j(f\Lambda) = (\partial_j f)\Lambda + f \partial_j \Lambda,$$

where on the right hand side, ∂_j denotes a classical derivative in the first term and a distributional derivative in the second.

10.3. **Smoothness of distributions.** For applications of distributions in the theory of PDEs, we need a more concrete understanding of the relationship between classical and distributional derivatives. This includes the answers to two questions:

- How well can an arbitrary distribution be approximated by a C^m-function? (see Corollary 10.32)
- How can one recognize whether a given distribution is representable by a C^m-function? (see Theorem 10.33)

The most useful tool toward these ends is a generalization of the convolution operator.

10.3.1. *The convolution.* Recall from §5.2 that for any locally integrable function $f : \mathbb{R}^n \to V$ and any test function $\varphi \in \mathscr{D}(\mathbb{R}^n)$, the convolution $\varphi * f : \mathbb{R}^n \to V$ is a well-defined function at every point $x \in \mathbb{R}^n$. It can be expressed in terms of the distribution $\Lambda_f \in \mathscr{D}'(\mathbb{R}^n)$ if we introduce two natural operations on the space of test functions: one is the translation operator

$$\tau_v : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}(\mathbb{R}^n), \qquad \tau_v\varphi(x) := \varphi(x+v) \quad \text{for} \quad v \in \mathbb{R}^n,$$

which we considered on L^p-spaces in §5.1. The other is the antipodal reflection operator

$$\sigma : \mathscr{D}(\mathbb{R}^n) \to \mathscr{D}(\mathbb{R}^n), \qquad \sigma\varphi(x) := \varphi(-x).$$

Both τ_v and σ extend naturally to operations on the space of distributions on $\mathbb{R}^n$. For $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ and $\varphi \in \mathscr{D}(\mathbb{R}^n)$, we have

$$\Lambda_{\tau_v f}(\varphi) = \int_{\mathbb{R}^n} \varphi(x) f(x+v)\, dx = \int_{\mathbb{R}^n} \varphi(x-v) f(x)\, dx = \Lambda_f(\tau_{-v}\varphi),$$

which motivates defining

$$\tau_v : \mathscr{D}'(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n), \qquad \tau_v\Lambda := \Lambda \circ \tau_{-v}.$$

Similarly,

$$\Lambda_{\sigma f}(\varphi) = \int_{\mathbb{R}^n} \varphi(x) f(-x)\, dx = \int_{\mathbb{R}^n} \varphi(-x) f(x)\, dx = \Lambda_f(\sigma\varphi),$$

and we therefore define

$$\sigma : \mathscr{D}'(\mathbb{R}^n) \to \mathscr{D}'(\mathbb{R}^n), \qquad \sigma\Lambda := \Lambda \circ \sigma.$$

One verifies easily that τ_v and σ are each continuous linear maps on both $\mathscr{D}(\mathbb{R}^n)$ and $\mathscr{D}'(\mathbb{R}^n)$.

The convolution of $\varphi \in \mathscr{D}(\mathbb{R}^n)$ with $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ can now be expressed as

$$(\varphi * f)(x) = \int_{\mathbb{R}^n} \varphi(x-y) f(y)\, dy = \int_{\mathbb{R}^n} \sigma\varphi(y-x) f(y)\, dy = \int_{\mathbb{R}^n} \tau_{-x}\sigma\varphi(y) f(y)\, dy = \Lambda_f(\tau_{-x}\sigma\varphi).$$

It is therefore sensible to define the **convolution** of any distribution $\Lambda \in \mathscr{D}'(\mathbb{R}^n)$ with a test function $\varphi \in \mathscr{D}(\mathbb{R}^n)$ as the function $\varphi * \Lambda : \mathbb{R}^n \to V$ given by

$$(\varphi * \Lambda)(x) := \Lambda(\tau_{-x}\sigma\varphi) = \tau_x\Lambda(\sigma\varphi). \tag{10.3}$$

With a little care, this definition can be extended to include distributions that are defined only on an open subset $\Omega \subset \mathbb{R}^n$. Given subsets $A, B \subset \mathbb{R}^n$ and $v \in \mathbb{R}^n$, let us denote

$$\begin{aligned} A \pm v &:= \left\{ x \pm v \in \mathbb{R}^n \;\middle|\; x \in A \right\}, \\ A \pm B &:= \left\{ x \pm y \in \mathbb{R}^n \;\middle|\; x \in A \text{ and } y \in B \right\}, \\ -A &:= \left\{ -x \in \mathbb{R}^n \;\middle|\; x \in A \right\}. \end{aligned}$$

Then for any function $\varphi : \mathbb{R}^n \to \mathbb{K}$ with support in a subset $K \subset \mathbb{R}^n$, and for any $v \in \mathbb{R}^n$,

$$\operatorname{supp}(\varphi) \subset K \quad \Rightarrow \quad \operatorname{supp}(\tau_v\varphi) \subset K - v \quad \text{and} \quad \operatorname{supp}(\sigma\varphi) \subset -K.$$

It follows that for open sets $\Omega, \Omega' \subset \mathbb{R}^n$ and $v \in \mathbb{R}^n$, there is a well-defined continuous linear operator

$$\tau_v : \mathscr{D}(\Omega) \to \mathscr{D}(\Omega') \qquad \text{whenever} \qquad \Omega - v \subset \Omega',$$

and similarly

$$\sigma : \mathscr{D}(\Omega) \to \mathscr{D}(\Omega') \qquad \text{whenever} \qquad -\Omega \subset \Omega'.$$

Now if $\Lambda \in \mathscr{D}'(\Omega)$ and $K \subset \mathbb{R}^n$ is any compact set containing the support of $\varphi \in \mathscr{D}(\mathbb{R}^n)$, then (10.3) defines $\varphi * \Lambda$ as a function on the open set

$$\Omega' := \left\{ x \in \mathbb{R}^n \;\middle|\; -K + x \subset \Omega \right\}.$$

One must keep in mind that this set may be empty, but we will mostly be interested in situations where K is an arbitrarily small compact neighborhood of the origin, in which case Ω' is a nonempty subset of Ω. Since convolutions of functions are symmetric, we define

$$\Lambda * \varphi := \varphi * \Lambda.$$

Exercise 10.22. Prove that for $v \in \mathbb{R}^n$ and $k = 1, \dots, n$, the operators τ_v, σ and ∂_k, acting on either the space of test functions or the space of distributions, are related to each other by

$$\tau_v \circ \partial_k = \partial_k \circ \tau_v, \qquad \sigma \circ \partial_k = -\partial_k \circ \sigma, \qquad \tau_v \circ \sigma = \sigma \circ \tau_{-v}.$$

We will see below that even in cases where Λ is not a function, the function $\varphi * \Lambda$ inherits the smoothness of φ. The proof of this rests on the smoothness of the translation operator τ_x as a function of $x \in \mathbb{R}^n$, i.e. the fact that for any fixed $\varphi \in \mathscr{D}(\mathbb{R}^n)$ and $\Lambda \in \mathscr{D}'(\Omega')$, the function $x \mapsto (\tau_x \Lambda)(\varphi) = \Lambda(\tau_{-x}\varphi)$ is smooth on a suitable open domain in $\mathbb{R}^n$. This follows in turn from a more general result related to differentiation under the integral sign.

The setting for the result we need is as follows. Assume $\mathcal{U} \subset \mathbb{R}^m$ and $\Omega \subset \mathbb{R}^n$ are open subsets, $\varphi : \mathcal{U} \times \Omega \to \mathbb{K}$ is a smooth function such that $\varphi_x := \varphi(x, \cdot) \in \mathscr{D}(\Omega)$ for every $x \in \mathcal{U}$, and $f \in L^1_{\text{loc}}(\Omega)$. One can then consider the function F on $\mathcal{U}$ defined via the parameter-dependent integral

$$F(x) := \int_\Omega \varphi(x, y) f(y)\, dy = \Lambda_f(\varphi_x).$$

If φ satisfies sufficient hypotheses for the application of Theorem 0.4, then one should expect this function to be smooth and satisfy

$$\frac{\partial^{|\alpha|} F}{\partial x^\alpha}(x) = \int_\Omega \frac{\partial^{|\alpha|}\varphi}{\partial x^\alpha}(x, y) f(y)\, dy = \Lambda_f\left(\frac{\partial^{|\alpha|}\varphi}{\partial x^\alpha}(x, \cdot) \right)$$

for every multi-index α in the variables $x = (x_1, \dots, x_m) \in \mathcal{U} \subset \mathbb{R}^m$. It turns out that under a mild assumption about the support of φ, this also works when Λ_f is replaced by an arbitrary distribution:

Proposition 10.23. *Assume $\mathcal{U} \subset \mathbb{R}^m$ and $\Omega \subset \mathbb{R}^n$ are open subsets and $\varphi : \mathcal{U} \times \Omega \to \mathbb{K}$ is a smooth function such that for every compact set $K \subset \mathcal{U}$, $\varphi|_{K \times \Omega}$ has compact support. Then for any $\Lambda \in \mathscr{D}'(\Omega)$, the function*

$$F : \mathcal{U} \to V : x \mapsto \Lambda(\varphi(x, \cdot))$$

is smooth and satisfies

$$\frac{\partial^{|\alpha|} F}{\partial x^\alpha}(x) = \Lambda\left(\frac{\partial^{|\alpha|}\varphi}{\partial x^\alpha}(x, \cdot) \right)$$

for all multi-indices α in the variables $x = (x_1, \dots, x_m) \in \mathcal{U} \subset \mathbb{R}^m$.

The proof requires two preparatory lemmas about the space of test functions.

Lemma 10.24. *Under the assumptions of Proposition 10.23, the map $\mathcal{U} \to \mathscr{D}(\Omega) : x \mapsto \varphi_x := \varphi(x, \cdot)$ is continuous.*

Proof. Given a convergent sequence $x_j \to x_\infty$ in $\mathcal{U}$, choose a compact set $C \subset \mathcal{U}$ containing an open neighborhood of x_∞. By assumption, there then exists a compact set $K \subset \Omega$ such that φ_x vanishes outside K for all $x \in C$, thus $\text{supp}(\varphi_{x_j}) \subset K$ for all j sufficiently large. It thus remains only to prove C^∞-convergence of φ_{x_j} to φ_{x_∞}. Uniform convergence follows from the fact that since $C \times K$ is compact, φ is uniformly continuous on $C \times K$. The same argument proves uniform convergence $\partial^\alpha \varphi_{x_j} \to \partial^\alpha \varphi_{x_\infty}$ for all multi-indices α in the variables $y = (y_1, \dots, y_n) \in \Omega \subset \mathbb{R}^n$, since $\frac{\partial^{|\alpha|}\varphi}{\partial y^\alpha}$ is also continuous. $\square$

Lemma 10.25. *Under the assumptions of Proposition 10.23, the functions* $\varphi_x := \varphi(x, \cdot) : \Omega \to \mathbb{K}$ *satisfy*

$$\lim_{h \to 0} \frac{\varphi_{x+he_k} - \varphi_x}{h} = \frac{\partial \varphi}{\partial x_k}(x, \cdot)$$

for every $x \in \mathcal{U}$ *and* $k = 1, \dots, m$*, where* $e_1, \dots, e_m \in \mathbb{R}^m$ *denotes the standard Euclidean basis, and the convergence of the limit is in the topology of* $\mathscr{D}(\Omega)$*.*

Proof. Fix $x \in \mathcal{U}$ and $k \in \{1, \dots, m\}$. For all $h \in \mathbb{R}\backslash\{0\}$ close enough to 0, we can assume $x + he_k$ belongs to a compact subset in $\mathcal{U}$ such that all the functions $\varphi_{x+he_k} : \Omega \to \mathbb{R}$ have support contained in some fixed compact subset $K \subset \Omega$. Now use the fundamental theorem of calculus to write

$$\varphi_{x+he_k}(y) - \varphi_x(y) = h \int_0^1 \frac{\partial \varphi}{\partial x_k}(x + the_k, y)\, dt,$$

and note that for any multi-index α in the variables $y = (y_1, \dots, y_n) \in \Omega \subset \mathbb{R}^n$, the operator $\frac{\partial^{|\alpha|}}{\partial y^\alpha}$ can be passed under the integral sign on the right hand side since φ is smooth. We thus have

$$\left| \frac{\partial^{|\alpha|}}{\partial y^\alpha} \left(\frac{\varphi_{x+he_k}(y) - \varphi_x(y)}{h} \right) - \frac{\partial^{|\alpha|}}{\partial y^\alpha} \frac{\partial \varphi}{\partial x_k}(y) \right| \leqslant \int_0^1 \left| \frac{\partial^{|\alpha|}}{\partial y^\alpha} \frac{\partial \varphi}{\partial x_k}(x + the_k, y) - \frac{\partial^{|\alpha|}}{\partial y^\alpha} \frac{\partial \varphi}{\partial x_k}(x, y) \right| dt.$$

Since $\frac{\partial^{|\alpha|}}{\partial y^\alpha} \frac{\partial \varphi}{\partial x_k}(x + the_k, y)$ can be assumed to vanish for all $y \notin K$ and $|h|$ sufficiently small, uniform continuity implies that the integrand on the right hand side becomes arbitrarily small uniformly in $y \in \Omega$ as $h \to 0$. $\square$

Proof of Proposition 10.23. The continuity of F follows immediately from Lemma 10.24 and the continuity of Λ. The main task is thus to prove that F has first partial derivatives given by

$$\frac{\partial F}{\partial x_k}(x) = \Lambda \left(\frac{\partial \varphi}{\partial x_k}(x, \cdot) \right),$$

since a similar application of Lemma 10.24 will then imply that these derivatives are also continuous, and the argument can be repeated inductively for all higher-order derivatives. For the computation of $\frac{\partial F}{\partial x_k}(x)$, one can again appeal to the continuity of Λ, together with Lemma 10.25, which gives

$$\frac{F(x + he_k) - F(x)}{h} = \frac{\Lambda(\varphi_{x+he_k}) - \Lambda(\varphi_x)}{h} = \Lambda \left(\frac{\varphi_{x+he_k} - \varphi_x}{h} \right) \to \Lambda \left(\frac{\partial \varphi}{\partial x_k}(x, \cdot) \right)$$

as $h \to 0$. $\square$

Corollary 10.26. *For an open set* $\Omega \subset \mathbb{R}^n$ *and a compact set* $K \subset \mathbb{R}^n$*, consider the open set*

$$\mathcal{U} := \left\{ x \in \mathbb{R}^n \;\middle|\; K + x \subset \Omega \right\}.$$

For any $\varphi \in \mathscr{D}(\mathbb{R}^n)$ *with* $\operatorname{supp}(\varphi) \subset K$*, associate to each* $\Lambda \in \mathscr{D}'(\Omega)$ *the function* F_Λ *defined on* $\mathcal{U}$ *by*

$$F_\Lambda(x) := (\tau_x \Lambda)(\varphi).$$

Then F_Λ *is smooth and satisfies* $\partial^\alpha F_\Lambda = F_{\partial^\alpha \Lambda}$ *for every multi-index* α*.*

Proof. Apply Proposition 10.23 with the smooth function $\mathcal{U} \times \Omega \to \mathbb{R} : (x, y) \mapsto \varphi(y - x)$. $\square$

This is enough preparation to prove the first main result about the convolution.

Theorem 10.27. *Suppose* $\Lambda \in \mathscr{D}(\Omega)$ *for an open set* $\Omega \subset \mathbb{R}^n$*, and* $\varphi \in \mathscr{D}(\mathbb{R}^n)$ *has support contained in the compact set* $K \subset \mathbb{R}^n$*. Then* (10.3) *defines a smooth function* $\varphi * \Lambda$ *on the open domain* $\Omega' := \left\{ x \in \mathbb{R}^n \;\middle|\; -K + x \subset \Omega \right\}$*, and it satisfies*

$$\partial^\alpha(\varphi * \Lambda) = (\partial^\alpha \varphi) * \Lambda = \varphi * (\partial^\alpha \Lambda)$$

for every multi-index α*, where the operator* ∂^α *denotes a classical derivative in the first formula and a distributional derivative in the second.*

Proof. The second formula is immediate from Corollary 10.26 and the definition of the convolution. Since ∂_k commutes with translation operators and anticommutes with σ, we also have

$$\begin{aligned}(\varphi * \partial_k \Lambda)(x) &= \tau_x \partial_k \Lambda(\sigma\varphi) = \partial_k \Lambda(\tau_{-x}\sigma\varphi) = -\Lambda(\partial_k \tau_{-x}\sigma\varphi) = \Lambda(\tau_{-x}\sigma\partial_k\varphi) = \tau_x\Lambda(\sigma\partial_k\varphi) \\ &= (\partial_k\varphi * \Lambda)(x)\end{aligned}$$

for all $x \in \Omega'$ and $k = 1, \ldots, n$. The relation $\varphi * \partial^\alpha \Lambda = \partial^\alpha\varphi * \Lambda$ follows from this by induction on the order of differentiation. $\square$

Since $\varphi * \Lambda$ is always a smooth function on Ω', it also defines an element of $\mathscr{D}'(\Omega')$. We would next like to give an alternative characterization of this distribution. For the case $\Lambda = \Lambda_f$ with $f \in L^1(\Omega)$, f can be extended to a function on $\mathbb{R}^n$ vanishing outside of Ω without changing the values of $\varphi * \Lambda_f = \varphi * f$ on Ω'. For any $\psi \in \mathscr{D}(\Omega')$, we can similarly extend ψ as 0 on $\mathbb{R}^n \backslash \Omega'$, and then use Fubini's theorem to show

$$\begin{aligned}(\varphi * f, \psi) &= \int_{\mathbb{R}^n} \psi(x)(\varphi * f)(x)\, dx = \int_{\mathbb{R}^n\times\mathbb{R}^n} \psi(x)\varphi(x-y)f(y)\, dx\, dy \\ &= \int_{\mathbb{R}^n\times\mathbb{R}^n} (\sigma\varphi)(y-x)\psi(x)f(y)\, dx\, dy = \int_{\mathbb{R}^n} (\sigma\varphi * \psi)(y)f(y)\, dy = (f, \sigma\varphi * \psi).\end{aligned}$$

It turns out that this formula remains valid when f is replaced by an arbitrary distribution. The proof requires a preparatory exercise.

Exercise 10.28. Show that for any $\varphi, \psi \in \mathscr{D}(\mathbb{R}^n)$ with $\operatorname{supp}(\varphi) \subset K \subset \mathbb{R}^n$ and $\operatorname{supp}(\psi) \subset K' \subset \mathbb{R}^n$, $\varphi * \psi$ is also in $\mathscr{D}(\mathbb{R}^n)$ and has $\operatorname{supp}(\varphi * \psi) \subset K + K'$. Moreover, if ψ_j is a sequence converging to ψ in $\mathscr{D}(\mathbb{R}^n)$, then $\varphi * \psi_j \to \varphi * \psi$ in $\mathscr{D}(\mathbb{R}^n)$.
Hint: Focus on proving uniform convergence of $\varphi * \psi_j$ to $\varphi * \psi$. Everything involving derivatives then follows easily from the formula $\partial^\alpha(\varphi * \psi) = \partial^\alpha\varphi * \psi = \varphi * \partial^\alpha\psi$.

Proposition 10.29. *For any Λ and φ satisfying the assumptions of Theorem 10.27 and any $\psi \in \mathscr{D}(\Omega')$, the smooth function $\sigma\varphi * \psi$ has compact support in Ω, and*

$$(\varphi * \Lambda, \psi) = (\Lambda, \sigma\varphi * \psi).$$

Proof. Since $\operatorname{supp}(\sigma\varphi) \subset -K$ and ψ has compact support in Ω', Exercise 10.28 together with the definition of Ω' in Theorem 10.27 imply $\sigma\varphi * \psi \in \mathscr{D}(\Omega)$.

To prove the stated formula, we shall exploit the linearity of Λ by approximating the integral defining $F(x) := (\sigma\varphi * \psi)(x) = \int_{\mathbb{R}^n} \sigma\varphi(x-y)\psi(y)\, dy$ with Riemann sums. For $\epsilon > 0$ and any given $x \in \mathbb{R}^n$, the compact support of ψ implies that the function $y \mapsto \sigma\varphi(x-y)\psi(y)$ is nonzero on at most finitely many points in the lattice $\epsilon\mathbb{Z}^n \subset \mathbb{R}^n$, thus we can define a function $F_\epsilon : \mathbb{R}^n \to \mathbb{K}$ by

$$F_\epsilon(x) := \epsilon^n \sum_{y\in\epsilon\mathbb{Z}^n} \sigma\varphi(x-y)\psi(y) = \epsilon^n \sum_{y\in\epsilon\mathbb{Z}^n} \tau_{-y}\sigma\varphi(x)\psi(y).$$

In fact, this is a finite linear combination of smooth functions with compact supports contained in $-K + \operatorname{supp}(\psi) \subset -K + \Omega' \subset \Omega$, thus it belongs to $\mathscr{D}(\Omega)$ and its support is contained in a compact subset of Ω independent of ϵ. The function $F_\epsilon(x)$ can also be written as $\int_{\mathbb{R}^n} f_{\epsilon,x}(y)\, dy$ for a step function $f_{\epsilon,x} : \mathbb{R}^n \to \mathbb{K}$ whose value at each y is the value of $f_{0,x}(y) := \sigma\varphi(x-y)\psi(y)$ at the nearest lattice point $y \in \epsilon\mathbb{Z}^n$. Since φ and ψ are both uniformly continuous, for every $\delta > 0$ there exists $\epsilon_0 > 0$ such that $\|f_{\epsilon,x} - f_{0,x}\|_{C^0} < \delta$ for all $x \in \mathbb{R}^n$ and $\epsilon < \epsilon_0$, thus F_ϵ converges uniformly to F as $\epsilon \to 0$. The same is then true for all derivatives: since $\partial^\alpha F(x) = (\partial^\alpha(\sigma\varphi) * \psi)(x)$ and $\partial^\alpha F_\epsilon(x) = \epsilon^n \sum_{y\in\epsilon\mathbb{Z}^n} \partial^\alpha(\sigma\varphi)(x-y)\psi(y)$ for all multi-indices α, the same arguments imply that all derivatives of F_ϵ converge uniformly to F, hence $F_\epsilon \to F$ in $\mathscr{D}(\Omega)$. The continuity and linearity of Λ then imply

$$(\Lambda, \sigma\varphi*\psi) = \Lambda(F) = \lim_{\epsilon\to 0^+} \Lambda(F_\epsilon) = \lim_{\epsilon\to 0^+} \epsilon^n \sum_{y\in\epsilon\mathbb{Z}^n} \psi(y)\Lambda(\tau_{-y}\sigma\varphi(x)) = \lim_{\epsilon\to 0^+} \epsilon^n \sum_{y\in\epsilon\mathbb{Z}^n} \psi(y)(\varphi*\Lambda)(y).$$

This last expression is a Riemann sum approximating the integral $\int_{\Omega'} \psi(y)(\varphi * \Lambda)(y)\, dy$, whose integrand is also a smooth function with compact support, so the sum converges to the integral as $\epsilon \to 0$. $\square$

10.3.2. *Approximation of distributions by smooth functions.*

Example 10.30. The Dirac δ-function $\delta \in \mathscr{D}'(\mathbb{R}^n)$ satisfies $(\varphi*\delta)(x) = \delta(\tau_{-x}\sigma\varphi) = \tau_{-x}\sigma\varphi(0) = \sigma\varphi(-x) = \varphi(x)$, i.e. $\varphi * \delta = \delta * \varphi = \varphi$ for every $\varphi \in \mathscr{D}(\mathbb{R}^n)$.

The definition of the term *approximate identity* in §5.4 can now be restated as follows: a sequence of smooth functions $\rho_j : \mathbb{R}^n \to [0, \infty)$ is an approximate identity if and only if

$$\rho_j \to \delta \quad \text{in} \quad \mathscr{D}'(\mathbb{R}^n),$$

where we are of course identifying the functions ρ_j with the distributions $\Lambda_{\rho_j} \in \mathscr{D}'(\mathbb{R}^n)$ that they determine. If ρ_j also has shrinking support, then we can assume for any given open neighborhood $\Omega \subset \mathbb{R}^n$ of the origin that ρ_j belongs to $\mathscr{D}(\Omega)$ for large j.

Now suppose $\Lambda \in \mathscr{D}'(\Omega)$ is an arbitrary distribution on some open set $\Omega \subset \mathbb{R}^n$, and ρ_j is an approximate identity with $\operatorname{supp}(\rho_j) \subset B_{r_j}$ for some sequence $r_j \to 0$. The convolutions $\Lambda_j := \rho_j * \Lambda$ are then defined on the subsets

$$\Omega_j := \left\{ x \in \Omega \;\middle|\; \operatorname{dist}(x, \mathbb{R}^n \backslash \Omega) > r_j \right\}, \tag{10.4}$$

whose union for all j is Ω. It follows that any $\varphi \in \mathscr{D}(\Omega)$ has support contained in Ω_j for all j sufficiently large, so that the integrals $\int_\Omega \varphi\Lambda_j\, dm := \int_{\Omega_j} \varphi\Lambda_j\, dm$ can be defined for large j by regarding the integrand as 0 wherever φ vanishes. The statement of the following result should be understood in these terms.

Theorem 10.31. *Suppose $\rho_j : \mathbb{R}^n \to [0, \infty)$ is an approximate identity with shrinking support, $\Lambda \in \mathscr{D}'(\Omega)$ is a distribution defined on some open set $\Omega \subset \mathbb{R}^n$, and $\Lambda_j := \rho_j * \Lambda$. Then for every $\varphi \in \mathscr{D}(\Omega)$, $\int_\Omega \varphi\Lambda_j\, dm \to \Lambda(\varphi)$.*

Proof. Assume j is large enough for $\operatorname{supp}(\varphi)$ to be contained in the domain of Λ_j. Then according to Proposition 10.29,

$$\int_\Omega \varphi\Lambda_j\, dm = (\rho_j * \Lambda, \varphi) = \Lambda(\sigma\rho_j * \varphi).$$

The functions $\sigma\rho_j$ are also an approximate identity with shrinking support, so the result follows via the continuity of Λ and the following claim: for any approximate identity ρ_j with shrinking support and any $\varphi \in \mathscr{D}(\Omega)$, the functions $\rho_j * \varphi$ have compact support in Ω for all j sufficiently large and converge in $\mathscr{D}(\Omega)$ to φ as $j \to \infty$. Indeed, Exercise 10.28 implies that $\operatorname{supp}(\rho_j * \varphi)$ lives in an arbitrarily small compact neighborhood of $\operatorname{supp}(\varphi)$ for large j, and Theorem 5.17 gives convergence $\rho_j * \varphi \to \varphi$ in $C^\infty_{\text{loc}}(\Omega)$. In light of the supports, C^∞_{loc}-convergence in this situation implies uniform convergence of all derivatives and thus convergence in $\mathscr{D}(\Omega)$. $\square$

Corollary 10.32. *For every open set $\Omega \subset \mathbb{R}^n$, $C^\infty_0(\Omega)$ is dense in $\mathscr{D}'(\Omega)$.*

Proof. Given an approximate identity ρ_j with shrinking support, define $\Lambda_j := \rho_j * \Lambda$, a sequence of smooth functions defined on the nested sequence of open subsets $\Omega_1 \subset \Omega_2 \subset \ldots \subset \bigcup_{j\in\mathbb{N}} \Omega_j$ described in (10.4). Choose a corresponding sequence of smooth functions $\beta_j : \Omega \to [0, 1]$ with $\operatorname{supp}(\beta_j) \subset \Omega_j$ and $\beta_j|_{\Omega_{j-1}} \equiv 1$. Then $\beta_j\Lambda_j$ can be extended to smooth functions on Ω that vanish outside of Ω_j, and since every $\varphi \in \mathscr{D}(\Omega)$ has support in Ω_j for j large, Theorem 10.31 implies $\beta_j\Lambda_j \to \Lambda$ in $\mathscr{D}'(\Omega)$. $\square$

10.3.3. *Distributions of class C^m.*

Theorem 10.33. *For a distribution $\Lambda \in \mathscr{D}'(\Omega)$ on an open set $\Omega \subset \mathbb{R}^n$ and integers $m, k \geqslant 0$, the following conditions are equivalent:*

1. *Λ is represented by a function of class C^{k+m};*
2. *$\partial^\alpha\Lambda$ is represented by a function of class C^k for each multi-index α of order m.*

Proof. The main step is to prove the special case with $k = 0$ and $m = 1$, as the rest then follows by a straightforward inductive argument. Let us therefore assume $\Lambda \in \mathscr{D}'(\Omega)$ has the property that $\partial_k \Lambda = \Lambda_{g_k}$ for every $k = 1, \dots, n$, with continuous functions $g_k \in C^0(\Omega)$. The goal is then to show that $\Lambda = \Lambda_f$ for some $f \in C^1(\Omega)$.

Choose an approximate identity ρ_j with shrinking support, and consider the sequence of smooth functions $f_j := \rho_j * \Lambda$, which are defined on a nested sequence of open subdomains $\Omega_j \subset \Omega$ whose union is Ω. By Theorem 10.27, $\partial_k f_j = \rho_j * g_k$ for each $k = 1, \dots, n$, and since the g_k are continuous, it follows via Theorem 5.17 that $\partial_k f_j \to g_k$ in $C^0_{\text{loc}}(\Omega)$. We claim that f_j also converges in $C^1_{\text{loc}}(\Omega)$ to a function $f \in C^1(\Omega)$. Indeed, by the fundamental theorem of calculus, every $x_0 \in \Omega$ has a convex neighborhood $\mathcal{U}_{x_0} \subset \Omega$ in which for $x = x_0 + h \in \mathcal{U}_{x_0}$ with $h = (h_1, \dots, h_n) \in \mathbb{R}^n$,

$$f_j(x) - f_j(x_0) = \sum_{k=1}^n h_k \int_0^1 \partial_k f_j(x_0 + th)\, dt, \tag{10.5}$$

and the right hand side converges uniformly in x to $\sum_{k=1}^n h_k \int_0^1 g_k(x_0 + th)\, dt$. If $f_j(x_0)$ converges, it follows that f_j converges uniformly on a neighborhood of x, and the limiting function will then satisfy

$$f(x) - f(x_0) = \sum_{k=1}^n h_k \int_0^1 g_k(x_0 + th)\, dt,$$

implying that f is of class C^1 on this neighborhood with $\partial_k f = g_k$. The claim will thus follow if we can prove that $f_j(x_0)$ converges. To this end, choose a test function $\varphi : \mathbb{R}^n \to [0, \infty)$ that is positive at x_0 and has support in a neighborhood $\mathcal{U}_{x_0}$ of x_0 which can be assumed to be arbitrarily small. By Theorem 10.31,

$$\lim_{j \to \infty} \int_{\mathcal{U}_{x_0}} \varphi f_j\, dm \to \Lambda(\varphi). \tag{10.6}$$

Now if $f_j(x_0)$ does not converge, then at least one of the following occurs after passing to a subsequence:

(1) $|f_j(x_0)| \to \infty$. Since (10.5) implies that $|f_j(x) - f_j(x_0)|$ is bounded independently of j for all $x \in \mathcal{U}_{x_0}$, it follows if $\operatorname{supp}(\varphi)$ is sufficiently concentrated around x_0 that $\left| \int_{\mathcal{U}_{x_0}} \varphi f_j\, dm \right| \to \infty$, contradicting (10.6).
(2) $f_{2j-1}(x_0)$ and $f_{2j}(x_0)$ each converge to different limits. A similar argument via (10.5) then implies that if φ has support sufficiently concentrated near x_0, then $\int_{\mathcal{U}_{x_0}} \varphi f_{2j-1}\, dm$ and $\int_{\mathcal{U}_{x_0}} \varphi f_{2j}\, dm$ each converge to different limits, giving another contradiction to (10.6).

These contradictions prove the claim.

We've now proved that f_j converges in $C^1_{\text{loc}}(\Omega)$ to a function $f \in C^1(\Omega)$, and it follows that for every $\varphi \in \mathscr{D}(\Omega)$, $\int_\Omega \varphi f_j\, dm \to \int_\Omega \varphi f\, dm$. The latter equals $\Lambda(\varphi)$ according to Theorem 10.31, so $\Lambda = \Lambda_f$. □

Here is a consequence that is much less obvious than it looks:

Corollary 10.34. *If f and g are two functions on a connected open set $\Omega \subset \mathbb{R}^n$ that have the same weak first-order partial derivatives almost everywhere, then $f - g$ is equal to a constant almost everywhere.*

Proof. The assumptions imply that $h := f - g$ satisfies $h' = 0$ in the sense of distributions. Since 0 is a continuous function, Theorem 10.33 then implies that h is equal almost everywhere to a C^1-function whose classical gradient is zero; since Ω is connected, that function is a constant. □

Exercise 10.35. Consider a linear differential operator of the form $L = \sum_\alpha c_\alpha \partial^\alpha$ acting on scalar-valued functions on $\mathbb{R}^n$, where the coefficients c_α are scalars and the sum runs over finitely many multi-indices, which may be of various orders. A distribution $K \in \mathscr{D}'(\mathbb{R}^n)$ is called a **fundamental solution**[25] for the operator L if it satisfies $LK = \delta$.

(a) Show that if K is a fundamental solution for L, then for every smooth compactly supported function $f : \mathbb{R}^n \to \mathbb{K}$, $u := K * f$ is a smooth solution to the partial differential equation $Lu = f$.
(b) Find a locally integrable function $K : \mathbb{R} \to \mathbb{R}$ that is a fundamental solution for the operator ∂_x^2, and verify explicity that $u := K * f$ satisfies $u'' = f$ for any $f \in C_0^\infty(\mathbb{R})$.

Exercise 10.36. Show that the functions

$$K(x) := -\frac{1}{2\pi} \ln|x| \quad \text{for } n = 2, \qquad K(x) := \frac{1}{(n-2)\operatorname{Vol}(S^{n-1})|x|^{n-2}} \quad \text{for } n \geqslant 3,$$

where $\operatorname{Vol}(S^{n-1}) > 0$ denotes the volume of the unit sphere in $\mathbb{R}^n$, are in $L^1_{\text{loc}}(\mathbb{R}^n)$ and are fundamental solutions for the Laplace operator $\Delta := -\sum_{j=1}^n \partial_j^2$ on $\mathbb{R}^n$ with $n \geqslant 2$. In particular, they have (weak) first derivatives

$$K_j(x) := \partial_j K(x) = -\frac{1}{\operatorname{Vol}(S^{n-1})} \frac{x_j}{|x|^n},$$

and their second derivatives (in the sense of distributions) take the form

$$K_{jk}(x) := \partial_j \partial_k K(x) = \frac{1}{\operatorname{Vol}(S^{n-1})} \frac{x_j x_k}{|x|^{n+2}}, \qquad \text{for } j \neq k,$$

and $\partial_j^2 K = -\frac{1}{n}\delta + K_{jj}$, where

$$K_{jj}(x) := \frac{1}{\operatorname{Vol}(S^{n-1})} \sum_k \frac{x_j^2 - x_k^2}{|x|^{n+2}},$$

and the evaluation of $K_{jk} \in \mathscr{D}'(\mathbb{R}^n)$ on test functions is defined via principal value integrals as in Exercise 10.20, that is,

$$(K_{jk}, \varphi) := \lim_{\epsilon \to 0^+} \int_{\mathbb{R}^n \setminus B_\epsilon^n} K_{jk}(x)\varphi(x)\, dx.$$

10.4. **Product distributions.** In this subsection we assume for simplicity that all distributions are scalar valued, though the discussion can be generalized for vector-valued distributions with minor adjustments (see Remark 10.42).

Recall that for any two σ-finite measure spaces (X, μ) and (Y, ν), there is a *product measure* $\mu \otimes \nu$ on $X \otimes Y$, which is uniquely determined by the condition

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$$

for arbitrary measurable sets $A \subset X$ and $B \subset Y$. Fubini's theorem is essentially the statement that product measures exist and are unique, together with a useful recipe for computing integrals with respect to product measures. We would now like to establish a variation on Fubini's theorem for distributions.

Definition 10.37. If $f : X \to \mathbb{K}$ and $g : Y \to \mathbb{K}$ are two scalar-valued functions on sets X and Y respectively, we define a scalar-valued function $f \otimes g : X \times Y \to \mathbb{K}$ by

$$(f \otimes g)(x, y) := f(x)g(y).$$

Given two open sets $\Omega_1 \subset \mathbb{R}^m$, $\Omega_2 \subset \mathbb{R}^n$ and distributions $\Lambda_1 \in \mathscr{D}'(\Omega_1)$ and $\Lambda_2 \in \mathscr{D}'(\Omega_2)$, a distribution on $\Omega_1 \times \Omega_2 \subset \mathbb{R}^{m+n}$ is called a **product distribution** for Λ_1 and Λ_2, and denoted by $\Lambda_1 \otimes \Lambda_2 \in \mathscr{D}'(\Omega_1 \times \Omega_2)$, if it satisfies

$$(\Lambda_1 \otimes \Lambda_2)(\varphi_1 \otimes \varphi_2) = \Lambda_1(\varphi_1)\Lambda_2(\varphi_2) \quad \text{for all} \quad \varphi_1 \in \mathscr{D}(\Omega_1) \text{ and } \varphi_2 \in \mathscr{D}(\Omega_2).$$

[25]Fundamental solutions are also often called **Green's functions**.

Example 10.38. If Λ_1 and Λ_2 are given by measures as in Example 10.14, then the product measure defines a product distribution $\Lambda_1 \otimes \Lambda_2$. (Note that a measure satisfying the condition stated in Example 10.14 is always σ-finite.)

Exercise 10.39. Use Fubini's theorem to show that for any locally integrable scalar-valued functions $f \in L^1_{\text{loc}}(\Omega_1)$ and $g \in L^1_{\text{loc}}(\Omega_2)$, $f \otimes g$ belongs to $L^1_{\text{loc}}(\Omega_1 \times \Omega_2)$ and $\Lambda_{f\otimes g} = \Lambda_f \otimes \Lambda_g \in \mathscr{D}'(\Omega_1 \times \Omega_2)$.

In the setting of Exercise 10.39, Fubini's theorem provides the following recipe for evaluating $\Lambda_f \otimes \Lambda_g$ on an arbitrary test function $\varphi \in \mathscr{D}(\Omega_1 \otimes \Omega_2)$: extending f and g to functions on $\mathbb{R}^m$ and $\mathbb{R}^n$ that vanish outside Ω and Ω' respectively, the compact support of φ in $\Omega \times \Omega'$ makes $(x, y) \mapsto \varphi(x, y)f(x)g(y)$ a well-defined function in $L^1(\mathbb{R}^{m+n})$ and thus implies

$$
\begin{aligned}
(\Lambda_f \otimes \Lambda_g, \varphi) &= \int_{\Omega\times\Omega'} \varphi(x, y)f(x)g(y)\, dx\, dy = \int_{\mathbb{R}^{m+n}} \varphi(x, y)f(x)g(y)\, dx\, dy \\
&= \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} \varphi(x, y)g(y)\, dy \right) f(x)\, dx = \Lambda_f \left(x \mapsto \Lambda_g(\varphi(x, \cdot)) \right) \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} \varphi(x, y)f(x)\, dx \right) g(y)\, dy = \Lambda_g \left(y \mapsto \Lambda_f(\varphi(\cdot, y)) \right).
\end{aligned}
$$

Implicit in our notation in the last two lines is that $x \mapsto \Lambda_g(\varphi(x, \cdot))$ and $y \mapsto \Lambda_f(\varphi(\cdot, y))$ define smooth compactly supported scalar-valued functions on Ω and Ω' respectively, so they can be regarded as test functions and fed into distributions for evaluation. As an easy consequence of Proposition 10.23, the same holds when Λ_f and Λ_g are replaced by arbitrary distributions:

Exercise 10.40 (cf. Proposition 10.23)**.** Show that if $\Omega_1 \subset \mathbb{R}^m$ and $\Omega_2 \subset \mathbb{R}^n$ are open sets, $\varphi \in \mathscr{D}(\Omega_1 \times \Omega_2)$ and $\Lambda \in \mathscr{D}'(\Omega_1)$, then $\psi(y) := \Lambda(\varphi(\cdot, y))$ defines a smooth compactly supported function on Ω_2.

Theorem 10.41 (Fubini's theorem for distributions)**.** *In the setting of Definition 10.37, there exists a unique product distribution $\Lambda_1 \otimes \Lambda_2 \in \mathscr{D}'(\Omega_1 \times \Omega_2)$, and its evaluation on arbitrary test functions $\varphi \in \mathscr{D}(\Omega_1 \times \Omega_2)$ is given by*

$$
(\Lambda_1 \otimes \Lambda_2)(\varphi) = \Lambda_1 \left(x \mapsto \Lambda_2(\varphi(x, \cdot)) \right) = \Lambda_2 \left(y \mapsto \Lambda_1(\varphi(\cdot, y)) \right). \tag{10.7}
$$

Proof. We first prove the uniqueness of $\Lambda_1 \otimes \Lambda_2$. Given two product distributions for Λ_1 and Λ_2, their difference is a distribution $\Lambda \in \mathscr{D}'(\Omega_1 \times \Omega_2)$ such that $\Lambda(\varphi \otimes \psi) = 0$ for all $\varphi \in \mathscr{D}(\Omega_1)$ and $\psi \in \mathscr{D}(\Omega_2)$. The idea is now to use an approximate identity to approximate Λ with smooth functions that vanish. For $k = 1, 2$, let $\rho_j^{(1)}$ and $\rho_j^{(2)} : \mathbb{R}^n \to [0, \infty)$ denote approximate identities on $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, both with shrinking support. The functions $\rho_j := \rho_j^{(1)} \otimes \rho_j^{(2)} : \mathbb{R}^{m+n} \to [0, \infty)$ then also have shrinking support, and by Fubini's theorem, they satisfy

$$
\int_{\mathbb{R}^{m+n}} \rho_j\, dm = \left(\int_{\mathbb{R}^m} \rho_j^{(1)}\, dm \right) \left(\int_{\mathbb{R}^n} \rho_j^{(2)}\, dm \right) \to 1 \quad \text{as} \quad j \to \infty,
$$

so by Lemma 5.12, ρ_j is an approximate identity on $\mathbb{R}^{m+n}$. Theorem 10.31 then implies that for any $\varphi \in \mathscr{D}(\Omega_1 \times \Omega_2)$, $\rho_j * \Lambda$ is a smooth function defined on a neighborhood of the support of φ for j sufficiently large and satisfying

$$
\int_{\Omega\times\Omega'} \varphi(\rho_j * \Lambda) \to \Lambda(\varphi) \quad \text{as} \quad j \to \infty.
$$

But the function $\rho_j * \Lambda$ is given by

$$
(\rho_j * \Lambda)(x, y) = \tau_{(x,y)}\Lambda(\sigma\rho_j) = \Lambda(\tau_{-x}\sigma\rho_j^{(1)} \otimes \tau_{-y}\sigma\rho_j^{(2)})
$$

for all $(x, y) \in \mathbb{R}^{m+n}$ in its domain of definition, so taking $(x, y) \in \Omega_1 \times \Omega_2$ and j large enough for $\tau_{-x}\sigma\rho_j^{(1)}$ and $\tau_{-y}\sigma\rho_j^{(2)}$ to have support in Ω_1 and Ω_2 respectively, the defining property of Λ implies that $\rho_j * \Lambda$ vanishes, proving $\Lambda(\varphi) = 0$.

It is easy to see that both of the expressions on the right hand side of (10.7) evaluate like a product distribution on test functions of the form $\varphi_1 \otimes \varphi_2 \in \mathscr{D}(\Omega_1 \times \Omega_2)$, thus with uniqueness established, the rest of the theorem will follow if we can show that both of these expressions really define distributions, i.e. they are continuous linear maps on $\mathscr{D}(\Omega_1 \times \Omega_2)$. The proof works the same for both expressions, so let us focus on the first one and consider the linear map $\Lambda : \mathscr{D}(\Omega_1 \times \Omega_2) \to \mathbb{K}$ defined by

$$\Lambda(\varphi) = \Lambda_1 \left(x \mapsto \Lambda_2(\varphi(x, \cdot)) \right).$$

To show that this is continuous, suppose $\varphi_j \to \varphi_\infty$ in $\mathscr{D}(\Omega_1 \times \Omega_2)$, and pick compact subsets $K_1 \subset \Omega_1$ and $K_2 \subset \Omega_2$ such that $\operatorname{supp}(\varphi_j) \subset K := K_1 \times K_2$ for all j. Then the sequence φ_j is also convergent with respect to the C^∞-topology on

$$\mathscr{D}_K(\Omega_1 \times \Omega_2) := \left\{ \varphi \in \mathscr{D}(\Omega_1 \times \Omega_2) \;\middle|\; \operatorname{supp}(\varphi) \subset K \right\},$$

which is a closed subspace of the Fréchet space of C^∞-functions with bounded derivatives of all orders on $\Omega_1 \times \Omega_2$. Since Λ restricts to a continuous linear functional on this subspace, a standard result on continuous linear operators (see Lemmas 10.93 and 10.94 in §10.8, or [RS80, §V.1]) implies that there exists a continuous seminorm $\| \cdot \|$ on $\mathscr{D}_K(\Omega_1 \times \Omega_2)$ such that $|\Lambda(\varphi)| \leqslant \|\varphi\|$ holds for every $\varphi \in \mathscr{D}_K(\Omega_1 \times \Omega_2)$. Since the topology on $\mathscr{D}_K(\Omega_1 \times \Omega_2)$ is generated by the increasing sequence of norms $\| \cdot \|_{C^m}$ for $m \in \mathbb{N}$, this actually means that for sufficiently large constants $C > 0$ and $m \in \mathbb{N}$,

$$|\Lambda(\varphi)| \leqslant C \|\varphi\|_{C^m} \quad \text{for all} \quad \varphi \in \mathscr{D}_K(\Omega_1 \times \Omega_2).$$

This estimate applies in particular to the sequence φ_j and its derivatives $\partial^\alpha \varphi_j$ for every multi-index α. Writing $\psi_j(x) := \Lambda_2(\varphi_j(x, \cdot))$, Proposition 10.23 gives

$$\partial^\alpha \psi_j(x) = \Lambda_2 \left(\frac{\partial^{|\alpha|} \varphi_j}{\partial x^\alpha}(x, \cdot) \right),$$

thus

$$\begin{aligned} |\partial^\alpha \psi_\infty(x) - \partial^\alpha \psi_j(x)| &= \left| \Lambda_2 \left(\frac{\partial^{|\alpha|} \varphi_\infty}{\partial x^\alpha}(x, \cdot) - \frac{\partial^{|\alpha|} \varphi_j}{\partial x^\alpha}(x, \cdot) \right) \right| \leqslant C \left\| \frac{\partial^{|\alpha|} \varphi_\infty}{\partial x^\alpha}(x, \cdot) - \frac{\partial^{|\alpha|} \varphi_j}{\partial x^\alpha}(x, \cdot) \right\|_{C^m} \\ &\leqslant C \left\| \varphi_\infty - \varphi_j \right\|_{C^{m+|\alpha|}} \to 0 \quad \text{as} \quad j \to \infty, \end{aligned}$$

giving C^∞-convergence $\psi_j \to \psi_\infty$. Since $\operatorname{supp}(\varphi_j) \subset K_1 \times K_2$, we also have $\operatorname{supp}(\psi_j) \subset K_1$ for all j, thus $\psi_j \to \psi_\infty$ in $\mathscr{D}(\Omega_1)$, and the continuity of Λ_1 now implies $\Lambda(\varphi_j) = \Lambda_1(\psi_j) \to \Lambda_1(\psi_\infty) = \Lambda(\varphi_\infty)$. □

Remark 10.42. One can also define the notion of a product distribution $\Lambda_1 \otimes \Lambda_2$ if Λ_1 is scalar valued and Λ_2 is vector valued (or the other way around), but in this case an extra definition is needed before one can make sense of (10.7), as $x \mapsto \Lambda_2(\varphi(x, \cdot))$ is now a vector-valued function and thus does not belong to $\mathscr{D}(\Omega_1)$. The quickest way to rectify this is to choose a basis $e_1, \ldots, e_k$ of V and extend $\Lambda_1 : \mathscr{D}(\Omega_1) \to \mathbb{K}$ to a linear map from the space of compactly supported smooth functions $\Omega_1 \to V$ to V by $\Lambda_1(\sum_j \varphi_j e_j) := \sum_j \Lambda_1(\varphi_j) e_j$ for $\varphi_1, \ldots, \varphi_k \in \mathscr{D}(\Omega_1)$. It is easy to check that this definition is independent of the choice of basis, and Theorem 10.41 then becomes valid for the product of a scalar-valued and a vector-valued distribution.

10.5. **The Sobolev spaces $W^{m,p}(\Omega)$.** Let us now explain how to generalize the Sobolev spaces $H^m(\mathbb{R}^n)$ to arbitrary open domains $\Omega \subset \mathbb{R}^n$ and $p \neq 2$. The theory of distributions is not strictly needed for this discussion, but it makes some aspects of it seem easier and more natural.

Definition 10.43. For an open set $\Omega \subset \mathbb{R}^n$, an integer $m \geqslant 0$ and a real number $p \in [1, \infty]$, the space $W^{m,p}(\Omega)$ is defined to consist of all $f \in L^p(\Omega)$ such that for every multi-index α with $|\alpha| \leqslant m$, the weak derivative $\partial^\alpha f$ exists and is also in $L^p(\Omega)$. The norm on $W^{m,p}(\Omega)$ is defined by

$$\|f\|_{W^{m,p}} := \sum_{|\alpha| \leqslant m} \|\partial^\alpha f\|_{L^p}.$$

Remark 10.44. In contrast to §9, we are not considering non-integer values of m in our definition of $W^{m,p}(\Omega)$. Such a notion does exist but is much more complicated to define; details may be found in [AF03].

It is not hard to show that $W^{m,p}(\Omega)$ is a Banach space, as it admits a natural continuous linear inclusion

$$W^{m,p}(\Omega) \hookrightarrow \bigoplus_{|\alpha| \leqslant m} L^p(\Omega)$$

sending each $f \in W^{m,p}(\Omega)$ to a finite tuple of L^p-functions whose "α-coordinate" is $\partial^\alpha f$, and Exercise 10.45 below shows that the image of this inclusion is a closed subspace. More generally, one defines

$$W^{m,p}_{\mathrm{loc}}(\Omega) := \left\{ f \in L^p_{\mathrm{loc}}(\Omega) \;\middle|\; f \text{ has weak derivatives } \partial^\alpha f \in L^p_{\mathrm{loc}}(\Omega) \text{ for all } |\alpha| \leqslant m \right\},$$

which is equivalently the space of functions on Ω (up to equality almost everywhere) whose restrictions to every open subset with compact closure are of class $W^{m,p}$. As with L^p_{loc} (cf. §0.3), one can use the $W^{m,p}$-norms over an exhausting nested sequence of open subsets with compact closures $\Omega_1 \subset \bar{\Omega}_1 \subset \Omega_2 \subset \bar{\Omega}_2 \subset \ldots \subset \bigcup_{j \in \mathbb{N}} \Omega_j = \Omega$ to endow $W^{m,p}_{\mathrm{loc}}(\Omega)$ with the structure of a Fréchet space.

Exercise 10.45. Suppose $f_j \in W^{m,p}(\Omega)$ is a sequence such that for every multi-index α of order at most m, $\partial^\alpha f_j$ is L^p-convergent to some $g_\alpha \in L^p(\Omega)$. Show that the function $f := \lim_{j\to\infty} f_j \in L^p(\Omega)$ is then in $W^{m,p}(\Omega)$ and satisfies $\partial^\alpha f = g_\alpha$ for all $|\alpha| \leqslant m$.
Hint: For any test function $\varphi \in \mathscr{D}(\mathbb{R}^n)$, the L^p-convergence $\partial^\alpha f_j \to g_\alpha$ implies L^1-convergence on the support of φ.

Example 10.46. As shown in Exercise 10.3, the function $f(x) := |x|$ on $\mathbb{R}$ has a bounded weak derivative, thus $f \in W^{1,p}(\Omega)$ for every bounded open interval $\Omega \subset \mathbb{R}$ and $1 \leqslant p \leqslant \infty$. This shows that there is no value of p for which functions of class $W^{1,p}$ must be everywhere differentiable in the classical sense.

One can use approximate identities to show that the subspace

$$W^{m,p}(\Omega) \cap C^\infty(\Omega) \subset W^{m,p}(\Omega)$$

is dense for all $p < \infty$, thus an equivalent definition of $W^{m,p}(\Omega)$ for these cases would be as the closure of the space of smooth functions on Ω with respect to the $W^{m,p}$-norm. The next exercise proves a slightly stronger variant of this result in the case $\Omega = \mathbb{R}^n$.

Exercise 10.47. Prove via the following steps that $C^\infty_0(\mathbb{R}^n)$ is dense in $W^{m,p}(\mathbb{R}^n)$ for every $m \geqslant 0$ and $p < \infty$:

(a) If $f \in W^{m,p}(\mathbb{R}^n)$ and $\rho_j : \mathbb{R}^n \to [0,\infty)$ is an approximate identity with shrinking support, use Theorems 5.14 and 10.27 to show that $f_j := \rho_j * f$ is in $W^{m,p}(\mathbb{R}^n) \cap C^\infty(\mathbb{R}^n)$ and converges in $W^{m,p}$ to f as $j \to \infty$.

(b) Fix a smooth function $\psi : \mathbb{R}^n \to [0,1]$ that equals 1 on the unit ball and has compact support in the ball of radius 2, and let $\psi_\epsilon(x) := \psi(\epsilon x)$ for $\epsilon > 0$. Show that for any $f \in W^{m,p}(\mathbb{R}^n) \cap C^\infty(\mathbb{R}^n)$, $\psi_\epsilon f \to f$ in $W^{m,p}$ as $\epsilon \to 0$.
Hint: You need to estimate $\|\partial^\alpha[(1-\psi_\epsilon)f]\|_{L^p}$ for every multi-index α with $|\alpha| \leqslant m$. Consider separately the terms that either do or do not involve derivatives of ψ_ϵ.

Remark 10.48. While $C^\infty(\Omega) \cap W^{m,p}(\Omega)$ is always dense in $W^{m,p}(\Omega)$, it is not true for arbitrary open domains $\Omega \subset \mathbb{R}^n$ that $C^\infty_0(\Omega)$ is dense in $W^{m,p}(\Omega)$. In general, the $W^{m,p}$-closure of $C^\infty_0(\Omega)$ defines a closed subspace $W^{m,p}_0(\Omega) \subset W^{m,p}(\Omega)$ that is often useful in applications to boundary value problems, as it can be regarded as the space of $W^{m,p}$-functions on Ω that "vanish at the boundary". The proof in Exercise 10.47 that $C^\infty_0(\mathbb{R}^n)$ is dense in $W^{m,p}(\mathbb{R}^n)$ implicitly makes use of the fact that one has an infinite amount of room in $\mathbb{R}^n$ to "stretch out" the cutoff functions ψ_ϵ without losing control of their derivatives. This trick does not work more generally, e.g. when $\Omega \subset \mathbb{R}^n$ is bounded.

We can now clarify the relationship of $W^{m,p}(\Omega)$ to the Sobolev spaces we defined earlier via the Fourier transform.

Proposition 10.49. *For every integer $m \geqslant 0$, $W^{m,2}(\mathbb{R}^n) = H^m(\mathbb{R}^n)$.*

Proof. Both spaces are linear subspaces of $L^2(\mathbb{R}^n)$, and by Theorem 9.20 and Exercise 10.47, both contain the Schwartz space $\mathscr{S}(\mathbb{R}^n)$ as a dense subspace. One can easily show that the $W^{m,2}$-norm and H^m-norm are equivalent on $\mathscr{S}(\mathbb{R}^n)$, thus the two spaces are the closures of $\mathscr{S}(\mathbb{R}^n)$ with respect to equivalent norms, and are therefore identical. □

Exercise 10.50. Prove:

(a) If f is an absolutely continuous function on an interval $[a,b]$, then its classical derivative f' (defined almost everywhere according to Theorem 6.7) is also its weak derivative on the domain (a,b), hence $f \in W^{1,1}((a,b))$.
Hint: For any $\varphi \in \mathscr{D}((a,b))$, φf defines an absolutely continuous function on $[a,b]$ that vanishes at the end points.

(b) If $f \in W^{1,1}_{\mathrm{loc}}(\Omega)$ for an open subset $\Omega \subset \mathbb{R}$, then f is equal almost everywhere to a function that is absolutely continuous on every compact subinterval of Ω.
Hint: On $[a,b] \subset \Omega$, define $g(x) := \int_a^x f'(t)\,dt$ and apply Corollary 10.34.

(c) For any open interval $\Omega \subset \mathbb{R}$, there exists a constant $c > 0$ such that

$$\|f\|_{C^0} \leqslant c\|f\|_{W^{1,1}} \quad \text{for all} \quad f \in W^{1,1}(\Omega).$$

Hint: The fundamental theorem of calculus implies $|f(x) - f(y)| \leqslant \|f'\|_{L^1}$ for all $x, y \in \Omega$, and thus $|f(x)| \geqslant \|f\|_{C^0} - \|f'\|_{L^1}$ for all $x \in \Omega$.

Exercise 10.51. Consider the function $f(x) := \ln\big|\ln|x|\big|$ on the r-ball $B_r \subset \mathbb{R}^n$ about the origin for some $r \in (0,1)$.

(a) Show that the classical first derivatives $\partial_j f$, defined on $B_r \backslash \{0\}$, are also weak derivatives of f on B_r.
Hint: Since f and $\partial_j f$ are both in $L^1(B_r)$, for any $\varphi \in \mathscr{D}(B_r)$ supported in some cube $Q \subset B_r$ around 0, you can approximate $\int_Q \partial_j(\varphi f)\,dm$ by integrating over $Q\backslash\{|x_j| < \epsilon\}$ for small $\epsilon > 0$, and then use integration by parts. There will be a boundary term; you need to show that the singularity of f at 0 is not bad enough to make the boundary term matter as $\epsilon \to 0$.

(b) Show that for $n = 1$, $f \notin W^{1,p}(B_r)$ for any $p \geqslant 1$, but for $n \geqslant 2$, $f \in W^{1,p}(B_r)$ if and only if $p \leqslant n$.

We saw in §9 that in general, functions of class $W^{1,p}$ need not be anywhere differentiable, and on higher-dimensional domains, Exercise 10.51 shows that they need not even be continuous—the continuity result in Exercise 10.50 is special to one-dimensional domains. The Sobolev embedding theorem gives sharp criteria saying to what extent the functions in any given Sobolev space must be classically differentiable. The proof of this important result, which generalizes Theorems 9.10 and 9.36 beyond the case $p = 2$ and $\Omega = \mathbb{R}^n$, belongs more properly to a course on PDEs, so we will not include it, but here is the statement:

Theorem 10.52 (Sobolev embedding theorem)**.** *Suppose $k \in \mathbb{N}$ and $p \in [1,\infty)$ satisfy the relation*

$$0 < k - n/p \leqslant 1,$$

and $\Omega \subset \mathbb{R}^n$ is either $\mathbb{R}^n$ or an open subset whose closure is a compact C^1-smooth manifold with boundary.[26] *Then for every integer $m \geqslant 0$ and every $\alpha \in (0,1)$ with $\alpha \leqslant k - n/p$, there exists a continuous inclusion*

$$W^{k+m,p}(\Omega) \hookrightarrow C^{m,\alpha}(\Omega).$$

[26]The hypothesis on Ω can be generalized considerably; here we are only stating a version that can be understood without too many extra definitions. The theorem as stated remains true for any (bounded or unbounded) open domain $\Omega \subset \mathbb{R}^n$ whose boundary satisfies something called the "strong local Lipschitz condition"; see [AF03, §4.12] for details.

Exercise 10.53. Show that in the situation of Theorem 10.52, whenever Ω is bounded and the strict inequality $\alpha < k - n/p$ is satisfied, the inclusion $W^{k+m,p}(\Omega) \hookrightarrow C^{m,\alpha}(\Omega)$ is compact. In particular, there is a continuous inclusion $W^{k+m,p}(\Omega) \hookrightarrow C^m(\bar{\Omega})$ whenever $kp > n$, and it is compact if $\Omega \subset \mathbb{R}^n$ is bounded. (See §0.3 for the definition of the Banach space $C^m(\bar{\Omega})$.)

Theorem 10.52 motivates thinking of functions in $W^{k,p}(\Omega)$ as functions that have "$k - n/p$ continuous derivatives" whenever $kp > n$, where the number $k-n/p$ need not be an integer. This intuition is further supported by the following generalization of the obvious inclusion $H^t(\mathbb{R}^n) \hookrightarrow H^s(\mathbb{R}^n)$ for $t > s$. The case with Ω bounded is known as the *Rellich-Kondrachov compactness theorem* (cf. Theorem 9.14):

Theorem 10.54. *Under the same assumptions on Ω as in Theorem 10.52, suppose* $1 \leqslant p, q < \infty$ *and $k, m \geqslant 0$ are integers satisfying*

$$k \geqslant m, \quad p \leqslant q, \qquad \text{and} \qquad k - \frac{n}{p} \geqslant m - \frac{n}{q}.$$

Then there exists a continuous inclusion $W^{k,p}(\Omega) \hookrightarrow W^{m,q}(\Omega)$, and this inclusion is compact if the inequality $k - \frac{n}{p} \geqslant m - \frac{n}{q}$ is strict and Ω is bounded.

Exercise 10.55. When Ω is a bounded interval $(a, b) \subset \mathbb{R}$, Theorem 10.52 says that for all integers $m \geqslant 0$, there are continuous inclusions

$$W^{1+m,p}((a,b)) \hookrightarrow C^{m,\alpha}((a,b)) \quad \text{if} \quad 0 < \alpha < 1,\ 1 < p \leqslant \infty \text{ and } \alpha \leqslant 1 - \frac{1}{p}$$
$$W^{2+m,1}((a,b)) \hookrightarrow C^{m,\alpha}((a,b)) \quad \text{if} \quad 0 < \alpha < 1.$$

Prove this as follows:

(a) Deduce the inclusions $W^{2,1} \hookrightarrow C^{0,\alpha}$ for $\alpha \in (0, 1]$ from a continuous inclusion $W^{2,1} \hookrightarrow C^1$ using Exercise 10.50.
(b) Deduce the inclusion $W^{1,p} \hookrightarrow C^0$ for every $p \geqslant 1$ from Exercise 10.50.
(c) For $a \leqslant x < y \leqslant b$, the fundamental theorem of calculus implies $|f(x) - f(y)| \leqslant \|f'\|_{L^1([x,y])}$ for $f \in W^{1,p}((a, b))$ since (by Exercise 10.50) f is absolutely continuous. Use Hölder's inequality to deduce a Hölder-type estimate $|f(x) - f(y)| \leqslant c\|f'\|_{L^p} \cdot |x - y|^\alpha$ for $0 < \alpha \leqslant 1 - 1/p$ whenever $p > 1$. The proof for $m = 0$ is thus complete.
(d) Extend the result to all $m \in \mathbb{N}$ by induction.

According to Theorem 10.52, the condition $kp > n$ guarantees continuity for functions of class $W^{k,p}$ on n-dimensional domains. We saw in Exercise 10.50 that the situation is slightly better when $n = 1$: here the condition $kp = n$ already suffices for continuity, but the function in Exercise 10.51 demonstrates that this is false in dimensions $n \geqslant 2$. The situation with $kp = n$ is often called the *Sobolev borderline case*. Even in dimension one, the borderline case has the disadvantage that functions of class $W^{1,1}$ need not be Hölder continuous, and so in contrast to Exercise 10.53, the inclusion $W^{1,1}(\Omega) \hookrightarrow C^0(\Omega)$ for bounded intervals $\Omega \subset \mathbb{R}$ is not compact.

Exercise 10.56. Find a sequence of smooth functions $f_j : (-1, 1) \to \mathbb{R}$ such that $\|f_j\|_{L^1}$ and $\|f_j'\|_{L^1}$ are bounded but f_j has no C^0-convergent subsequence.
Hint: Construct f_j so that it converges in L^1 to a (discontinuous) characteristic function.

Remark 10.57. The Sobolev embedding theorem furnishes one major reason why it is useful to study the properties of *all* the L^p-spaces for $1 \leqslant p \leqslant \infty$, rather than just L^2, which might otherwise be easier since the latter is a Hilbert space. As a concrete example, suppose you are studying a first-order PDE for functions on 2-dimensional domains. If you want to work only with Hilbert spaces but also want all your functions to be continuous, then Theorem 10.52 requires you to take functions of class $H^m = W^{m,2}$ with $m \geqslant 2$, which involves at least one more (weak) derivative than the PDE itself actually needs. In such a situation, it may be easier to work with functions of class $W^{1,p}$ for some $p > 2$, as these are continuous, and one only needs to compute first-order derivatives in order to verify whether a given function belongs to this space.

10.6. **Tempered distributions and Fourier transforms.** Since we are going to talk about Fourier transforms in this subsection, we need to assume $\mathbb{K} = \mathbb{C}$.

We would now like to define Fourier transforms of functions for which the usual integral formula cannot even approximately make sense, e.g. functions that are not in $L^2(\mathbb{R}^n)$, and ideally, distributions. One can *almost* deduce the correct definition by considering the distribution $\Lambda_f \in \mathscr{D}'(\mathbb{R}^n)$ corresponding to a function $f \in L^1(\mathbb{R}^n)$: by Fubini's theorem, we have

$$(\mathscr{F}f, \varphi) = \int_{\mathbb{R}^n} \varphi(p) \left(\int_{\mathbb{R}^n} e^{-2\pi i p \cdot x} f(x)\, dx \right) dp = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} e^{-2\pi i p \cdot x} \varphi(p)\, dp \right) f(x)\, dx = (f, \mathscr{F}\varphi)$$

for all $\varphi \in \mathscr{D}(\mathbb{R}^n)$. This suggests defining $\mathscr{F}\Lambda \in \mathscr{D}'(\mathbb{R}^n)$ for arbitrary $\Lambda \in \mathscr{D}'(\mathbb{R}^n)$ by $(\mathscr{F}\Lambda)(\varphi) := \Lambda(\mathscr{F}\varphi)$, but this definition as it stands does not quite make sense: $\mathscr{F}\varphi$ might not have compact support, in which case it is not a test function and $\Lambda(\mathscr{F}\varphi)$ will not make sense for arbitrary distributions Λ. The solution is to replace the usual space of test functions with the Schwartz space $\mathscr{S}(\mathbb{R}^n)$, since the latter is closed under the Fourier transform.

Before defining what a continuous linear functional on $\mathscr{S}(\mathbb{R}^n)$ is, we need to define a topology on $\mathscr{S}(\mathbb{R}^n)$. As with $\mathscr{D}(\mathbb{R}^n)$, we would like this topology to be relatively strong, so that as many functionals as possible are continuous, but also to have the property that continuity can be characterized purely in terms of convergent sequences (cf. Proposition 10.8). This turns out to be easier for $\mathscr{S}(\mathbb{R}^n)$ than for $\mathscr{D}(\mathbb{R}^n)$: the natural choice is to endow $\mathscr{S}(\mathbb{R}^n)$ with the topology generated by the countable family of seminorms

$$\|\varphi\|_{\alpha,\beta} := \|x^\alpha \partial^\beta \varphi\|_{C^0}$$

for all multi-indices α, β, so convergence $\varphi_k \to \varphi$ in $\mathscr{S}(\mathbb{R}^n)$ will mean that for every polynomial function $P : \mathbb{R}^n \to \mathbb{R}$ and every multi-index β, the functions $P\partial^\beta \varphi_k$ converge uniformly on $\mathbb{R}^n$ to $P\partial^\beta \varphi$. It follows easily from the completeness of the C^0-norm that sequences that are Cauchy with respect to all of these seminorms must also converge, hence $\mathscr{S}(\mathbb{R}^n)$ is now a Fréchet space. In particular, the topology we have defined on $\mathscr{S}(\mathbb{R}^n)$ is metrizable, thus continuity and sequential continuity of functions defined on $\mathscr{S}(\mathbb{R}^n)$ are equivalent notions.

Exercise 10.58. Show that the natural inclusions $\mathscr{D}(\mathbb{R}^n) \hookrightarrow \mathscr{S}(\mathbb{R}^n)$ and $\mathscr{S}(\mathbb{R}^n) \hookrightarrow W^{m,p}(\mathbb{R}^n)$ for all $m \geqslant 0$ and $p \in [1, \infty]$ are continuous.

Exercise 10.59. Show that the following linear operators $\mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$ are continuous:

(a) ∂^α and $\varphi \mapsto x^\alpha \varphi$ for every multi-index α;
(b) $\mathscr{F}$ and $\mathscr{F}^*$.

Definition 10.60. A complex-valued **tempered distribution** on $\mathbb{R}^n$ is a continuous complex-linear functional $\Lambda : \mathscr{S}(\mathbb{R}^n) \to \mathbb{C}$. Similarly, a vector-valued tempered distribution with values in the finite-dimensional complex vector space V is a continuous complex-linear map $\Lambda : \mathscr{S}(\mathbb{R}^n) \to V$. We shall generally assume that all tempered distributions take values in a fixed vector space V, and denote the the vector space of vector-valued tempered distributions by

$$\mathscr{S}'(\mathbb{R}^n) = \left\{ \Lambda : \mathscr{S}(\mathbb{R}^n) \to V \;\middle|\; \Lambda \text{ is complex linear and continuous} \right\}.$$

The space $\mathscr{S}'(\mathbb{R}^n)$ is endowed with the weak*-topology, i.e. the locally convex topology generated by the seminorms $\|\Lambda\|_\varphi := |\Lambda(\varphi)|$ for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$, hence a sequence $\Lambda_j \in \mathscr{S}'(\mathbb{R}^n)$ converges to $\Lambda_\infty \in \mathscr{S}'(\mathbb{R}^n)$ if and only if $\Lambda_j(\varphi) \to \Lambda_\infty(\varphi)$ for every $\varphi \in \mathscr{S}(\mathbb{R}^n)$.

The inclusion $\mathscr{D}(\mathbb{R}^n) \hookrightarrow \mathscr{S}(\mathbb{R}^n)$ in Exercise 10.58 gives rise to a natural continuous inclusion $\mathscr{S}'(\mathbb{R}^n) \hookrightarrow \mathscr{D}'(\mathbb{R}^n)$, i.e. every tempered distribution is also a distribution in the usual sense. The converse is false, and in fact $\mathscr{S}'(\mathbb{R}^n)$ does not even contain all locally integrable functions, e.g. $f(x) := e^{x^2}$ does not define an element $\Lambda_f \in \mathscr{S}'(\mathbb{R})$ since there exist functions $\varphi \in \mathscr{S}(\mathbb{R})$ for which $\int_{\mathbb{R}} \varphi f\, dm$ is not defined. However, most important examples of distributions are also tempered distributions: these include large classes of functions as in the following two exercises,[27] as well as standard singular examples like the Dirac δ-function and its derivatives. By a slight abuse of notation, we shall write $L^1_{\mathrm{loc}}(\mathbb{R}^n) \cap \mathscr{S}'(\mathbb{R}^n)$ for the space of all locally integrable functions f on $\mathbb{R}^n$ such that $\varphi f \in L^1(\mathbb{R}^n)$ for every $\varphi \in \mathscr{S}(\mathbb{R}^n)$ and the formula $\Lambda_f(\varphi) := \int_{\mathbb{R}^n} \varphi f \, dm$ defines a tempered distribution $\Lambda_f \in \mathscr{S}'(\mathbb{R}^n)$.

Exercise 10.61. A function $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ is said to have **polynomial growth** if it satisfies $|f| \leqslant |P|$ for some polynomial function $P : \mathbb{R}^n \to \mathbb{R}$; equivalently, this is true if and only if there exist constants $C > 0$ and $k \in \mathbb{N}$ such that

$$|f(x)| \leqslant C(1 + |x|^k) \quad \text{for all} \quad x \in \mathbb{R}^n.$$

Show that any function with this property is in $\mathscr{S}'(\mathbb{R}^n)$.

Exercise 10.62. Show that $L^p(\mathbb{R}^n) \subset \mathscr{S}'(\mathbb{R}^n)$ for every $p \in [1, \infty]$, and the inclusions $L^p(\mathbb{R}^n) \hookrightarrow \mathscr{S}'(\mathbb{R}^n)$ are continuous.
Hint: Use the continuity of the inclusions $\mathscr{S}(\mathbb{R}^n) \hookrightarrow L^q(\mathbb{R}^n)$ and the natural injection $L^p \hookrightarrow (L^q)^*$ for $\frac{1}{p} + \frac{1}{q} = 1$.

Partial derivative operators are defined as continuous linear maps on $\mathscr{S}'(\mathbb{R}^n)$ in the same way as $\mathscr{D}'(\mathbb{R}^n)$; continuity in this case follows from the continuity of ∂^α on $\mathscr{S}(\mathbb{R}^n)$ (Exercise 10.59). The product of a smooth function $f \in C^\infty(\mathbb{R}^n)$ with a tempered distribution $\Lambda \in \mathscr{S}'(\mathbb{R}^n)$ is not well defined unless $\varphi \mapsto f\varphi$ is a continuous map $\mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$, which is not true e.g. for $f(x) := e^{x^2}$ on $\mathbb{R}$, but is true if f and its derivatives of all orders have polynomial growth as in Exercise 10.61. Under this assumption, it is straightforward to show that the Leibniz rule in Exercise 10.21 also holds for tempered distributions.

Remark 10.63. For a function $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ that defines a tempered distribution, we now have two potentially inequivalent definitions for the notion of weak derivatives $\partial_j f$, depending whether we want $\partial_j f$ to define an element of $\mathscr{D}'(\mathbb{R}^n)$ or $\mathscr{S}'(\mathbb{R}^n)$. In the latter case, it needs to satisfy a stronger condition involving integration against test functions in $\mathscr{S}(\mathbb{R}^n)$, a larger space than $\mathscr{D}(\mathbb{R}^n)$; it could happen for instance that f has a locally integrable weak derivative $\partial_j f$ that grows too fast at infinity to define a tempered distribution, in which case the stronger condition fails. However, if a weak derivative $\partial_j f$ does define a tempered distribution—which is always the case for instance if $\partial_j f$ is of class L^p for some p, and notably if f belongs to a suitable Sobolev space—then it also satisfies the stronger condition, i.e. it is also a derivative of f in the sense of *tempered* distributions. The reason is that, by Exercise 10.64 below, $\mathscr{D}(\mathbb{R}^n)$ is dense in $\mathscr{S}(\mathbb{R}^n)$, so any two tempered distributions that evaluate the same on $\mathscr{D}(\mathbb{R}^n)$ are identical.

Exercise 10.64. Show that for any $\varphi \in \mathscr{S}(\mathbb{R}^n)$ and the family of compactly supported smooth cutoff functions $\psi_\epsilon : \mathbb{R}^n \to [0,1]$ in Exercise 10.47, $\psi_\epsilon \varphi \to \varphi$ in $\mathscr{S}(\mathbb{R}^n)$ as $\epsilon \to 0$. In particular, $\mathscr{D}(\mathbb{R}^n)$ is dense in $\mathscr{S}(\mathbb{R}^n)$.
Hint: For any multi-indices α and β, the condition $\varphi \in \mathscr{S}(\mathbb{R}^n)$ implies $\|x^\alpha \partial^\beta \varphi\|_{C^0(\mathbb{R}^n \setminus B_R)} \to 0$ as $R \to \infty$. (Why?)

The next set of exercises generalizes the convolution operator and its main properties from §10.3 to the context of tempered distributions.

Exercise 10.65. Recall the translation operator τ_v for functions f on $\mathbb{R}^n$ and $v \in \mathbb{R}^n$, defined by $(\tau_v f)(x) := f(x + v)$.

(a) Show that for every pair of multi-indices α and β, there exists a constant $C > 0$ and a finite set of pairs of multi-indices $\{(\alpha_i, \beta_i)\}_{i=1}^N$ such that

$$\|\tau_v \varphi\|_{\alpha,\beta} \leqslant C \left(\sum_{i=1}^N \|\varphi\|_{\alpha_i,\beta_i} \right) (1 + |v|^{|\alpha|}) \quad \text{for all} \quad \varphi \in \mathscr{S}(\mathbb{R}^n),\ v \in \mathbb{R}^n.$$

[27] The word "tempered" refers to conditions as in Exercise 10.61 and 10.62 that rule out functions like e^{x^2}, which grow too fast at infinity.

In particular, $\tau_v\varphi$ is also in $\mathscr{S}(\mathbb{R}^n)$ for every $\varphi \in \mathscr{S}(\mathbb{R}^n)$ and $v \in \mathbb{R}^n$.
Hint: By Exercise 8.2, you can assume $|\partial^\beta\varphi(x)| \leqslant \frac{c}{1+|x|^k}$ *for some* $k \in \mathbb{N}$ *arbitrarily large and a constant* $c > 0$ *determined by* k *and finitely many seminorms of* φ*. Estimate* $|x^\alpha\partial^\beta\varphi(x+v)|$ *by looking separately at the cases* $|x| \leqslant 2|v|$ *and* $|x| \geqslant 2|v|$*.*

(b) Show that Lemmas 10.24 and 10.25 remain valid with the space of test functions replaced by the Schwartz space (ignoring all conditions that involve compact support).

Exercise 10.66. Reread the proof of Corollary 10.26 and verify that, in light of Exercise 10.65, the function $F_\Lambda(x) := (\tau_x\Lambda)(\varphi)$ defined on $\mathbb{R}^n$ for any $\Lambda \in \mathscr{S}'(\mathbb{R}^n)$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$ is smooth and satisfies $\partial^\alpha F_\Lambda = F_{\partial^\alpha\Lambda}$ for all multi-indices α.

Exercise 10.67. Show that for any $\varphi \in \mathscr{S}(\mathbb{R}^n)$ and $\Lambda \in \mathscr{S}'(\mathbb{R}^n)$, the formula (10.3) defines a smooth function $\varphi * \Lambda$ on $\mathbb{R}^n$ satisfying $\partial^\alpha(\varphi * \Lambda) = (\partial^\alpha\varphi) * \Lambda = \varphi * (\partial^\alpha\Lambda)$ for all multi-indices α.

Exercise 10.68. Consider the convolution of two Schwartz functions $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$.

(a) Show that $\varphi * \psi$ is continuous and bounded on $\mathbb{R}^n$.
(b) Show that if ψ_j is a sequence converging in $\mathscr{S}(\mathbb{R}^n)$ to ψ, then $\varphi * \psi_j$ converges uniformly to $\varphi * \psi$.
(c) For $k = 1, \dots, n$ and a function f on $\mathbb{R}^n$, let $P_k f$ denote the function on $\mathbb{R}^n$ defined by $(P_k f)(x) := x_k f(x)$, so e.g. by Exercise 10.59, P_k defines a continuous linear operator $\mathscr{S}(\mathbb{R}^n) \to \mathscr{S}(\mathbb{R}^n)$. Show that $P_k(\varphi * \psi) = (P_k\varphi) * \psi + \varphi * P_k\psi$, and deduce that $P_k(\varphi * \psi)$ is continuous and bounded.
(d) Deduce that $\varphi * \psi \in \mathscr{S}(\mathbb{R}^n)$, and for any sequence $\psi_j \to \psi$ in $\mathscr{S}(\mathbb{R}^n)$, $\varphi * \psi_j \to \varphi * \psi$ in $\mathscr{S}(\mathbb{R}^n)$.

Proposition 10.69. *For* $\varphi \in \mathscr{S}(\mathbb{R}^n)$ *and* $\Lambda \in \mathscr{S}'(\mathbb{R}^n)$*,* $\varphi * \Lambda$ *is a polynomially bounded function and thus defines an element of* $\mathscr{S}'(\mathbb{R}^n)$*. Moreover, if* $\varphi_j \in \mathscr{S}(\mathbb{R}^n)$ *converges to* φ *in* $\mathscr{S}(\mathbb{R}^n)$*, then the tempered distributions* $\varphi_j * \Lambda$ *converge to* $\varphi * \Lambda$ *in* $\mathscr{S}'(\mathbb{R}^n)$*.*

Proof. We start by proving that $\varphi * \Lambda$ has polynomial growth. By one of the standard characterizations of continuity for linear operators on locally convex spaces (see Lemmas 10.93 and 10.94 in §10.8, or [RS80, §V.1]), the continuity of $\Lambda : \mathscr{S}(\mathbb{R}^n) \to V$ means that there exists a finite set of pairs of multi-indices $\{(\alpha_i, \beta_i)\}_{i=1}^N$ and a constant $C > 0$ such that

$$|\Lambda(\varphi)| \leqslant C\sum_{i=1}^N \|\varphi\|_{\alpha_i,\beta_i} \quad \text{for all} \quad \varphi \in \mathscr{S}(\mathbb{R}^n).$$

Using Exercise 10.65, the convolution $\varphi * \Lambda$ thus satisfies

$$|(\varphi * \Lambda)(x)| = |\Lambda(\tau_{-x}\sigma\varphi)| \leqslant C\sum_{i=1}^N \|\tau_{-x}(\sigma\varphi)\|_{\alpha_i,\beta_i} \leqslant C\sum_{i=1}^N c_i(1 + |x|^{|\alpha_i|}) \leqslant C'(1 + |x|^k)$$

for suitable constants $c_i > 0$, $C' > 0$ and $k \in \mathbb{N}$ sufficiently large. In this expression, the constant $C > 0$ is determined entirely by Λ, while only $c_1, \dots, c_N$ (and therefore also C') depend on φ; looking more closely at Exercise 10.65, we see moreover that they can be bounded linearly in terms of finitely many of the seminorms $\|\varphi\|_{\alpha,\beta}$. For this reason, if $\varphi_j \to \varphi$ is a convergent sequence in $\mathscr{S}(\mathbb{R}^n)$, the same argument gives

$$|(\varphi * \Lambda)(x) - (\varphi_j * \Lambda)(x)| = |\Lambda(\tau_{-x}\sigma(\varphi - \varphi_j))| \leqslant C_j(1 + |x|^k)$$

for constants $C_j > 0$ that converge to 0 as $j \to \infty$, thus for any $\psi \in \mathscr{S}(\mathbb{R}^n)$,

$$|(\varphi * \Lambda, \psi) - (\varphi_j * \Lambda, \psi)| \leqslant \int_{\mathbb{R}^n} |\psi| \cdot |\varphi * \Lambda - \varphi_j * \Lambda| \, dm \leqslant C_j \int_{\mathbb{R}^n} |\psi(x)|(1 + |x|^k)\, dx \to 0.$$

□

Exercise 10.70. Use Proposition 10.29 and the density of $\mathscr{D}(\mathbb{R}^n)$ in $\mathscr{S}(\mathbb{R}^n)$ to deduce that the relation $(\varphi * \Lambda, \psi) = \Lambda(\sigma\varphi * \psi)$ also holds for all $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$ and $\Lambda \in \mathscr{S}'(\mathbb{R}^n)$.

Exercise 10.71. Suppose $\rho_j : \mathbb{R}^n \to [0,\infty)$ is an approximate identity with shrinking support. Prove:

(a) For any $\varphi \in \mathscr{S}(\mathbb{R}^n)$, $\rho_j * \varphi \to \varphi$ in $\mathscr{S}(\mathbb{R}^n)$ as $j \to \infty$.
(b) For any $\Lambda \in \mathscr{S}'(\mathbb{R}^n)$, $\rho_j * \Lambda \to \Lambda$ in $\mathscr{S}'(\mathbb{R}^n)$ as $j \to \infty$. (This proves that $C^\infty(\mathbb{R}^n) \cap \mathscr{S}'(\mathbb{R}^n)$ is dense in $\mathscr{S}'(\mathbb{R}^n)$.)
(c) For any $f \in C^\infty(\mathbb{R}^n) \cap \mathscr{S}'(\mathbb{R}^n)$ and the family of compactly supported smooth cutoff functions $\psi_\epsilon : \mathbb{R}^n \to [0,1]$ in Exercise 10.47, $\psi_\epsilon f \to f$ in $\mathscr{S}'(\mathbb{R}^n)$ as $\epsilon \to 0$. (This proves that $C_0^\infty(\mathbb{R}^n)$ is dense in $\mathscr{S}'(\mathbb{R}^n)$.

While distributions are easier to work with than tempered distributions for many purposes, the major advantage of the latter is that they admit natural definitions of the Fourier transform and Fourier inverse operators.

Definition 10.72. We define $\mathscr{F}, \mathscr{F}^* : \mathscr{S}'(\mathbb{R}^n) \to \mathscr{S}'(\mathbb{R}^n)$ by

$$(\mathscr{F}\Lambda)(\varphi) := \widehat{\Lambda}(\varphi) := \Lambda(\widehat{\varphi}) \qquad \text{and} \qquad (\mathscr{F}^*\Lambda)(\varphi) := \check{\Lambda}(\varphi) := \Lambda(\check{\varphi}).$$

The continuity of $\mathscr{F}$ and $\mathscr{F}^*$ on $\mathscr{S}(\mathbb{R}^n)$ (Exercise 10.59) implies that they are also continuous on $\mathscr{S}'(\mathbb{R}^n)$, and the relations $\mathscr{F}\mathscr{F}^* = \mathscr{F}^*\mathscr{F} = \mathbb{1}$ extend immediately from $\mathscr{S}(\mathbb{R}^n)$ to $\mathscr{S}'(\mathbb{R}^n)$. The calculation via Fubini's theorem at the beginning of this subsection shows that our definition of $\mathscr{F}\Lambda$ and $\mathscr{F}^*\Lambda$ for any $\Lambda = \Lambda_f$ with $f \in L^1(\mathbb{R}^n)$ matches the result of the usual integral formula.

Exercise 10.73. For $f \in L^2(\mathbb{R}^n)$, use approximation by L^1-functions to show that $\mathscr{F}\Lambda_f = \Lambda_{\mathscr{F}f}$ and $\mathscr{F}^*\Lambda_f = \Lambda_{\mathscr{F}^*f}$, where $\mathscr{F}f$ and $\mathscr{F}^*f$ are defined as in §8.

Remark 10.74. Recall from Lemma 10.4 that two locally integrable functions are equal almost everywhere if and only if they define the same distribution. The same is true for tempered distributions since $\mathscr{D}(\mathbb{R}^n) \subset \mathscr{S}(\mathbb{R}^n)$. Exercise 10.73 thus shows that the most general possible definition of the Fourier transform, given by Definition 10.72, matches the definition we previously had for functions in $L^2(\mathbb{R}^n)$.

We can now make rigorous sense of formal relations such as $\int_{\mathbb{R}^n} e^{-2\pi i p\cdot x}\, dx = \delta(x)$ that appeared in §8, for instance:

Exercise 10.75. Regarding the Dirac δ-function and the constant function 1 as tempered distributions on $\mathbb{R}^n$, show that $\mathscr{F}(\delta) = \mathscr{F}^*\delta = 1$, hence $\mathscr{F}^*(1) = \mathscr{F}(1) = \delta$.

Exercise 10.76. Show that the relations in (8.4) between the operators $\mathscr{F}$, $\mathscr{F}^*$ and ∂^α remain valid when $f \in \mathscr{S}(\mathbb{R}^n)$ is replaced by a tempered distribution $\Lambda \in \mathscr{S}'(\mathbb{R}^n)$.

Exercise 10.77. Show that the relations $\mathscr{F}(\varphi * \Lambda) = \widehat{\varphi}\widehat{\Lambda}$ and $\mathscr{F}^*(\varphi * \Lambda) = \check{\varphi}\check{\Lambda}$ hold for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$ and $\Lambda \in \mathscr{S}'(\mathbb{R}^n)$.

10.7. **Distributions with compact support.** We saw in §8 that the Fourier transform exchanges regularity properties of a function with decay conditions at infinity, e.g. one can see this in the relations (8.4) that transform differentiation into multiplication by polynomials, and the fact that Lebesgue-integrable functions have continuous Fourier transforms. We would now like to explain a beautiful extension of this phenomenon into the realm of distributions.

Definition 10.78. The **support** $\operatorname{supp}(\Lambda) \subset \Omega$ of a distribution $\Lambda \in \mathscr{D}'(\Omega)$ is the complement of the union of all open subsets $\mathcal{U} \subset \Omega$ such that $\Lambda(\varphi) = 0$ for all $\varphi \in \mathscr{D}(\Omega)$ with $\operatorname{supp}(\varphi) \subset \mathcal{U}$. Equivalently, $\operatorname{supp}(\Lambda)$ is the intersection of all closed subsets $\mathcal{V} \subset \Omega$ such that $\Lambda(\varphi) = 0$ for all $\varphi \in \mathscr{D}(\Omega)$ with $\operatorname{supp}(\varphi) \cap \mathcal{V} = \varnothing$.

Remark 10.79. The support of $\Lambda \in \mathscr{D}'(\Omega)$ is in fact the smallest closed subset such that Λ vanishes on all test functions with support disjoint from $\operatorname{supp}(\Lambda)$, or equivalently, its complement is the largest open subset $\mathcal{U} \subset \Omega$ such that $\Lambda(\varphi)$ vanishes whenever $\operatorname{supp}(\varphi) \subset \mathcal{U}$. To see that $\Omega \backslash \operatorname{supp}(\Lambda)$ has the latter property, observe that for any $\varphi \in \mathscr{D}(\Omega)$ with $\operatorname{supp}(\varphi) \cap \operatorname{supp}(\Lambda) = \varnothing$, the compactness of $\operatorname{supp}(\varphi)$ implies that it is contained in the union of a finite collection of open subsets $\mathcal{U}_1, \dots, \mathcal{U}_N$ such that $\operatorname{supp}(\psi) \subset \mathcal{U}_i$ implies $\Lambda(\psi) = 0$ for any $\psi \in \mathscr{D}(\Omega)$. One can then use a partition of unity to write φ as $\sum_{i=1}^N \varphi_i$ for some $\varphi_i \in \mathscr{D}(\Omega)$ with $\operatorname{supp}(\varphi_i) \subset \mathcal{U}_i$, implying $\Lambda(\varphi_i) = 0$ for all i and thus $\Lambda(\varphi) = 0$.

Example 10.80. If $f \in L^1_{\mathrm{loc}}(\Omega)$ vanishes outside of some closed subset $\mathcal{V} \subset \Omega$, then $\operatorname{supp}(\Lambda_f) \subset \mathcal{V}$.

Example 10.81. For any $\Lambda \in \mathscr{D}'(\Omega)$ and $f \in C^\infty(\Omega)$, $f\Lambda \in \mathscr{D}'(\Omega)$ has $\operatorname{supp}(f\Lambda) \subset \operatorname{supp}(f)$ since $f\varphi \equiv 0$ whenever $\varphi \in \mathscr{D}(\Omega)$ has support disjoint from that of f.

Lemma 10.82. *A distribution $\Lambda \in \mathscr{D}'(\Omega)$ has compact support if and only if there exists a distribution $\Lambda' \in \mathscr{D}'(\Omega)$ and a smooth compactly supported function $f : \Omega \to \mathbb{K}$ such that $\Lambda = f\Lambda'$.*

Proof. The statement is obvious in one direction since $\operatorname{supp}(f\Lambda') \subset \operatorname{supp}(f)$. Conversely, suppose there exists a compact subset $K \subset \Omega$ such that $\Lambda(\varphi) = 0$ whenever $\operatorname{supp}(\varphi) \cap K = \varnothing$. Choose an open neighborhood $\mathcal{U} \subset \Omega$ of K with compact closure and a compactly supported function $f : \Omega \to [0, 1]$ such that $f|_{\mathcal{U}} \equiv 1$. We claim that $f\Lambda = \Lambda$. Indeed, for any $\varphi \in \mathscr{D}(\Omega)$, we can write $\varphi = f\varphi + (1 - f)\varphi$, where $(1 - f)\varphi$ vanishes on $\mathcal{U}$, thus its support is disjoint from K, implying $\Lambda(\varphi) = \Lambda(f\varphi) = (f\Lambda)(\varphi)$. ☐

Proposition 10.83. *If $\Lambda \in \mathscr{D}'(\Omega)$ has compact support, then Λ extends to a continuous linear map on the space $C^\infty(\Omega)$ of all scalar-valued smooth functions with the C^∞_{loc}-topology.*

Proof. Suppose $\varphi_j \in \mathscr{D}(\Omega)$ is a sequence converging in the C^∞_{loc}-topology to $\varphi_\infty \in \mathscr{D}(\Omega)$. By Lemma 10.82, we can write $\Lambda = f\Lambda'$ for some $\Lambda' \in \mathscr{D}'(\Omega)$ and a smooth function $f : \Omega \to \mathbb{K}$ with support in a compact set $K \subset \Omega$. Since $\varphi_j \to \varphi_\infty$ in the C^∞-topology over K, it follows that $f\varphi_j$ is C^∞-convergent to $f\varphi_\infty$, thus $f\varphi_j \to f\varphi_\infty$ in $\mathscr{D}(\Omega)$, so that the continuity of Λ' implies

$$\Lambda(\varphi_j) = \Lambda'(f\varphi_j) \to \Lambda'(f\varphi_\infty) = \Lambda(\varphi_j).$$

This proves that $\Lambda : \mathscr{D}(\Omega) \to V$ is continuous with respect to C^∞_{loc}-convergence. Since $\mathscr{D}(\Omega)$ is dense in $C^\infty(\Omega)$ with respect to this topology, it follows that Λ has a unique continuous extension to the larger space. ☐

Remark 10.84. Proposition 10.83 also has a converse; see Proposition 10.104.

In light of the obvious continuous inclusion $\mathscr{S}(\mathbb{R}^n) \hookrightarrow C^\infty(\mathbb{R}^n)$, in which $C^\infty(\mathbb{R}^n)$ carries the C^∞_{loc}-topology, we also have:

Corollary 10.85. *Every compactly supported distribution $\Lambda \in \mathscr{D}'(\mathbb{R}^n)$ is also a tempered distribution, i.e. it has a unique extension to a continuous linear map on $\mathscr{S}(\mathbb{R}^n)$.* ☐

If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ has compact support, then f also belongs to $L^1(\mathbb{R}^n)$, so its Fourier transform is given by

$$\widehat{f}(p) = \int_{\mathbb{R}^n} e^{-2\pi i p\cdot x} f(x)\, dx = \Lambda_f(e^{-2\pi i p\cdot x}),$$

where we have used Proposition 10.83 to extend the domain of Λ_f to smooth functions such as $x \mapsto e^{-2\pi i p\cdot x}$ that need not have compact support. As we saw in §8, the fact that f is of class L^1 implies that $\widehat{f}$ is continuous, but we can now say more: since the product of f with any polynomial is also a compactly supported L^1_{loc}-function and therefore belongs to $L^1(\mathbb{R}^n)$, $\widehat{f}$ also has continuous derivatives of all orders, i.e. it is smooth. The remarkable fact is that this result still holds when f is replaced by an arbitrary compactly supported distribution, which may have very badly behaved local singularities but still satisfies the best possible "decay" condition at infinity:

Theorem 10.86. *For any compactly supported distribution Λ on $\mathbb{R}^n$, $\mathscr{F}f$ and $\mathscr{F}^*f$ are smooth functions on $\mathbb{R}^n$ given by*

$$\mathscr{F}\Lambda(p) = \Lambda(e^{-2\pi i p\cdot x}), \qquad \mathscr{F}^*\Lambda(p) = \Lambda(e^{2\pi i p\cdot x}),$$

where Proposition 10.83 is used for evaluating Λ on smooth functions with noncompact support.

Proof. By Exercise 10.76, smoothness will follow immediately once we have proved that the stated formulas for $\mathscr{F}\Lambda$ and $\mathscr{F}^*\Lambda$ are correct, as multiplying Λ by any polynomial preserves the condition of compact support. We shall focus on the formula for $\mathscr{F}\Lambda$, since the parallel statement for $\mathscr{F}^*\Lambda$ has an almost identical proof. By Lemma 10.82, it would be equivalent to prove that for every $\Lambda \in \mathscr{D}'(\mathbb{R}^n)$ and every compactly supported smooth function $\psi : \mathbb{R}^n \to \mathbb{K}$, $\widehat{\psi\Lambda} \in \mathscr{S}'(\mathbb{R}^n)$ is given by the function $p \mapsto \Lambda(\psi e^{-2\pi i p\cdot x})$ on $\mathbb{R}^n$. The latter encapsulates two claims:

(1) The function $g(p) := \Lambda(\psi e^{-2\pi i p\cdot x})$ has sufficiently tame behavior at infinity to define a tempered distribution;
(2) For all $\varphi \in \mathscr{S}(\mathbb{R}^n)$,

$$(\psi\Lambda, \hat{\varphi}) = \int_{\mathbb{R}^n} \varphi(p) g(p)\, dp. \tag{10.8}$$

For the first claim, let us show that g has polynomial growth. Indeed, a straightforward change-of-variable calculation gives

$$\psi(x) e^{-2\pi i p\cdot x} = \mathscr{F}^*(\tau_p \hat{\psi})(x),$$

thus

$$g(p) = \Lambda(\mathscr{F}^* \tau_p \hat{\psi}) = (\mathscr{F}^*\Lambda)(\tau_p \hat{\psi}) = \tau_{-p}\check{\Lambda}(\hat{\psi}) = \tau_{-p}\check{\Lambda}(\sigma(\sigma\hat{\psi})) = (\sigma\hat{\psi} * \check{\Lambda})(-p),$$

and the claim follows from Proposition 10.69 since $\sigma\hat{\psi} \in \mathscr{S}(\mathbb{R}^n)$ and $\check{\Lambda} \in \mathscr{S}'(\mathbb{R}^n)$.

In light of this result, both sides of (10.8) now clearly define continuous linear functions of $\varphi \in \mathscr{S}(\mathbb{R}^n)$, so to prove that they are identical, it will suffice to show this for all φ in the dense subspace $\mathscr{D}(\mathbb{R}^n)$. The goal is thus to prove that

$$\Lambda\left(x \mapsto \psi(x) \int_{\mathbb{R}^n} e^{-2\pi i p\cdot x} \varphi(p)\, dp \right) = \int_{\mathbb{R}^n} \varphi(p) \Lambda(\psi e^{-2\pi i p\cdot x})\, dp$$

holds for all $\Lambda \in \mathscr{D}'(\mathbb{R}^n)$ and $\varphi, \psi \in \mathscr{D}(\mathbb{R}^n)$. Writing $\mathbb{1} \in \mathscr{D}'(\mathbb{R}^n)$ for the scalar-valued distribution $\mathbb{1}(\varphi) := \int_{\mathbb{R}^n} \varphi\, dm$, Theorem 10.41 identifies both sides of this equation with $(\Lambda \otimes \mathbb{1})(F)$ for the test function $F \in \mathscr{D}(\mathbb{R}^n \times \mathbb{R}^n)$ given by $F(x,p) := \psi(x)\varphi(p)e^{-2\pi i p\cdot x}$. ☐

10.8. **Appendix: The topology of the space of test functions.** For a working knowledge of the theory of distributions, it is usually not necessary to understand the topology of the space $\mathscr{D}(\Omega)$ beyond the notions described in Definition 10.7 of convergent sequences and continuity of linear maps on $\mathscr{D}(\Omega)$. Nonetheless, the further development of the theory requires knowing that $\mathscr{D}(\Omega)$ can also be viewed as a topological vector space, in which convergence and continuity are determined by the topology. You may have noticed in Definition 10.7 that the notion of convergence in $\mathscr{D}(\Omega)$ is extremely strict, i.e. it is very hard for a sequence of test functions to converge. This strictness is an advantage, because it means that it is that much easier for a linear functional on $\mathscr{D}(\Omega)$ to be continuous; in other words, having fewer convergent sequences in $\mathscr{D}(\Omega)$ makes the space of distributions $\mathscr{D}'(\Omega)$ larger. This will mean that the topology of $\mathscr{D}(\Omega)$ needs to be quite strong,[28] e.g. it needs to involve conditions on derivatives of arbitrarily high orders, and therefore cannot be described merely in terms of a norm, so $\mathscr{D}(\Omega)$ will not be a Banach space. One might reasonably hope for it to be a Fréchet space, like the Schwartz space $\mathscr{S}(\mathbb{R}^n)$ (see §10.6), but this will also turn out to be too ambitious (see Remark 10.89). The next best thing would be a locally convex space, and this is not hard to achieve.

[28]Given two topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ on the same set X, one says that $\mathcal{T}_1$ is **stronger** (or **finer**, or **larger**) than $\mathcal{T}_2$ if every set in $\mathcal{T}_2$ also belongs to $\mathcal{T}_1$. One also says in this case that $\mathcal{T}_2$ is **weaker** (or **coarser** or **smaller**) than $\mathcal{T}_1$. Making a topology on X stronger makes it harder for sequences in X to converge and harder for maps from other spaces into X to be continuous, but easier for maps from X to other spaces to be continuous.

10.8.1. *Definition and properties of the topology.* For a compact subset $K \subset \Omega$, consider the linear subspace

$$\mathscr{D}_K(\Omega) := \left\{\varphi \in \mathscr{D}(\Omega) \;\middle|\; \operatorname{supp}(\varphi) \subset K\right\}.$$

The countable family of norms $\|\cdot\|_{C^m}$ for integers $m \geqslant 0$ endows $\mathscr{D}_K(\Omega)$ with the structure of a Fréchet space such that convergence of a sequence $\varphi_j \to \varphi_\infty$ in $\mathscr{D}_K(\Omega)$ means uniform convergence $\partial^\alpha \varphi_j \to \partial^\alpha \varphi_\infty$ for every multi-index α. We shall assume $\mathscr{D}_K(\Omega)$ to be endowed with this Fréchet space topology from now on. It will frequently be useful to observe that since $\mathscr{D}_K(\Omega)$ is metrizable, a function defined on $\mathscr{D}_K(\Omega)$ is continuous if and only if it is sequentially continuous.

According to Definition 10.7, a convergent sequence in $\mathscr{D}_K(\Omega)$ is also convergent in $\mathscr{D}(\Omega)$, so the topology we define on $\mathscr{D}(\Omega)$ should have the property that the obvious inclusion

$$\mathscr{D}_K(\Omega) \hookrightarrow \mathscr{D}(\Omega) \tag{10.9}$$

is sequentially continuous for every compact set $K \subset \Omega$. If these inclusions are continuous, and $\|\cdot\|$ is a seminorm on $\mathscr{D}(\Omega)$ that is continuous with respect to its topology, then $\|\cdot\|$ will also restrict to a continuous seminorm on $\mathscr{D}_K(\Omega)$. The following definition therefore produces the strongest locally convex topology on $\mathscr{D}(\Omega)$ for which the inclusions (10.9) are all continuous.

Definition 10.87. A **good seminorm** on $\mathscr{D}(\Omega)$ is a seminorm whose restriction to the subspace $\mathscr{D}_K(\Omega) \subset \mathscr{D}(\Omega)$ is continuous for every $K \subset \Omega$ compact. We endow $\mathscr{D}(\Omega)$ with the locally convex topology generated by the family of all good seminorms, i.e. a set $\mathcal{U} \subset \mathscr{D}(\Omega)$ is open if and only if for every $\varphi \in \mathcal{U}$, there exists a seminorm $\|\cdot\|$ on $\mathscr{D}(\Omega)$ such that

$$\left\{\psi \in \mathscr{D}(\Omega) \;\middle|\; \|\psi - \varphi\| < 1\right\} \subset \mathcal{U}$$

and $\|\cdot\|$ is continuous on $\mathscr{D}_K(\Omega)$ for all $K \subset \Omega$ compact.[29]

The next exercise shows that good seminorms on $\mathscr{D}(\Omega)$ exist in abundance, thus the topology we have defined on $\mathscr{D}(\Omega)$ is quite large.

Exercise 10.88. Show that each of the following defines a good seminorm on $\mathscr{D}(\Omega)$:

(a) $\|\varphi\|_\alpha := \max_{x\in\Omega} |\partial^\alpha \varphi(x)|$ for any multi-index α. (The C^m-norm for any $m \geqslant 0$ is a finite sum of seminorms of this type, thus it is also a good seminorm.)

(b) $\|\varphi\|_f := \|f\varphi\|$ where $\|\cdot\|$ is any good seminorm and $f : \Omega \to \mathbb{R}$ is any smooth function.

(c) $\|\varphi\|_{f,\alpha} := \|f\varphi\|_\alpha$ for any multi-index α and continuous function $f : \Omega \to \mathbb{R}$. For this example, the open set $\{\|\varphi\|_{f,\alpha} < 1\}$ describes all $\varphi \in \mathscr{D}(\Omega)$ that satisfy $|\partial^\alpha \varphi| < 1/|f|$ everywhere on Ω, where we adopt the convention $1/0 := \infty$ so that the condition is vacuous wherever $f = 0$.

Remark 10.89. The following observations show that the topology we've defined on $\mathscr{D}(\Omega)$ cannot be metrizable, so $\mathscr{D}(\Omega)$ is not a Fréchet space. If d were a metric defining the topology of $\mathscr{D}(\Omega)$, then for every $\varphi \in \mathscr{D}(\Omega)$, the sets $\mathcal{U}_j := \left\{\psi \in \mathscr{D}(\Omega) \;\middle|\; d(\varphi,\psi) < 1/j\right\}$ for $j \in \mathbb{N}$ would define a countable sequence of neighborhoods of φ with the property that *every* neighborhood of φ contains $\mathcal{U}_k$ for some $k \in \mathbb{N}$.[30] Since the topology is determined by good seminorms, this would equivalently mean that there exists a sequence of good seminorms $\|\cdot\|_j$ for $j \in \mathbb{N}$ such that for every good seminorm $\|\cdot\|$, the set $\left\{\psi \in \mathscr{D}(\Omega) \;\middle|\; \|\psi\| < 1\right\}$ contains $\left\{\psi \in \mathscr{D}(\Omega) \;\middle|\; \|\psi\|_k < 1\right\}$ for some $k \in \mathbb{N}$; in other words,

$$\|\cdot\| \leqslant \|\cdot\|_k \quad \text{for some} \quad k \in \mathbb{N}.$$

[29] In describing the topology of $\mathscr{D}(\Omega)$ in this way, we are using the easily verifiable fact that the maximum of any finite collection of good seminorms is also a good seminorm, and so is any positive multiple of a good seminorm. This implies that for any collection of good seminorms $\|\cdot\|_i$ and any $\epsilon_i > 0$ with $i = 1, \dots, N$, the finite intersection of the open neighborhoods $\{\psi \in \mathscr{D}(\Omega) \mid \|\psi - \varphi\|_i < \epsilon_i\}$ for $i = 1, \dots, N$ can equally well be described as $\{\psi \in \mathscr{D}(\Omega) \mid \|\psi - \varphi\| < 1\}$ where $\|f\| := \max\left\{\frac{\|f\|_1}{\epsilon_1}, \dots, \frac{\|f\|_N}{\epsilon_N}\right\}$ defines another good seminorm.

[30] A collection of neighborhoods with this property is called a **countable neighborhood base** of φ. A topological space in which every point admits a countable neighborhood base is called **first countable**. What Remark 10.89 shows in effect is that every metrizable space is first countable, but $\mathscr{D}(\Omega)$ is not.

By Exercise 10.88, it would follow that for every continuous function $f : \Omega \to \mathbb{R}$, there exists $k \in \mathbb{N}$ such that

$$\|f\varphi\|_{C^0} \leqslant \|\varphi\|_k \quad \text{for all} \quad \varphi \in (\Omega).$$

To see that this is impossible, pick a sequence of nontrivial functions $\varphi_1, \varphi_2, \varphi_3, \ldots \in \mathscr{D}(\Omega)$ whose supports are pairwise disjoint compact sets $K_1, K_2, K_3, \ldots \subset \Omega$, and choose $f : \Omega \to \mathbb{R}$ to be a continuous function that satisfies

$$f > \frac{\|\varphi_j\|_j}{\|\varphi_j\|_{C^0}} \quad \text{on } K_j \quad \text{for all } j \in \mathbb{N}.$$

Then $\|f\varphi_j\|_{C^0} > \|\varphi_j\|_j$ for every $j \in \mathbb{N}$, giving a contradiction.

Lemma 10.90. *If $\varphi_j \to \varphi_\infty$ in the topology of $\mathscr{D}(\Omega)$, then there exists a compact subset $K \subset \Omega$ such that $\varphi_j \in \mathscr{D}_K(\Omega)$ for every $j \in \mathbb{N} \cup \{\infty\}$.*

Proof. If not, then after replacing φ_j with a subsequence, we can find a sequence of points $x_j \in \Omega$ that lie outside the support of φ_∞, have no accumulation point, and satisfy $\varphi_j(x_j) \neq 0$ for every j. Choose a continuous function $f : \Omega \to (0, \infty)$ such that $f(x_j) \leqslant |\varphi_j(x_j)|$ for every j. Then by Exercise 10.88, $\mathcal{U} := \left\{\varphi \in \mathscr{D}(\Omega) \;\middle|\; |\varphi - \varphi_\infty| < f\right\}$ is an open neighborhood of φ_∞ in $\mathscr{D}(\Omega)$, but $\varphi_j \notin \mathcal{U}$ for every j, so φ_j cannot converge to φ_∞. ☐

Corollary 10.91. *A sequence $\varphi_j \in \mathscr{D}(\Omega)$ converges to $\varphi_\infty \in \mathscr{D}(\Omega)$ if and only if there exists a compact set $K \subset \Omega$ such that $\varphi_j \in \mathscr{D}_K(\Omega)$ for all $j \in \mathbb{N} \cup \{\infty\}$ and $\varphi_j \to \varphi_\infty$ in the topology of $\mathscr{D}_K(\Omega)$.* ☐

As preparation for the next result, we need some general facts about continuity for linear maps between locally convex spaces. A preliminary remark about locally convex topologies is in order. If X carries the locally convex topology generated by a given family of seminorms $\{\|x\|_\alpha\}_{\alpha \in I}$,then by definition, every open set in X is a union of finite intersections of sets of the form $\left\{x \in X \;\middle|\; \|x - x_0\|_\alpha < \epsilon\right\}$ for arbitrary $x_0 \in X$, $\alpha \in I$ and $\epsilon > 0$. Equivalently, a set $\mathcal{U} \subset Y$ is open if and only if for every $x_0 \in \mathcal{U}$, there exists a nonempty finite subset $I_0 \subset I$ and numbers $\epsilon_\alpha > 0$ for $\alpha \in I_0$ such that

$$x \in X \text{ with } \|x - x_0\|_\alpha < \epsilon_\alpha \text{ for all } \alpha \in I_0 \quad \Rightarrow \quad x \in \mathcal{U}.$$

The seminorms $\|\cdot\|_\alpha : X \to [0, \infty)$ are each continuous functions, and in the situation above, $\|x\| := \sum_{\alpha \in I_0} \frac{\|x\|_\alpha}{\epsilon_\alpha}$ also defines a continuous seminorm; the aforementioned condition can then equally well be described as

$$x \in X \text{ with } \|x - x_0\| < 1 \quad \Rightarrow \quad x \in \mathcal{U}.$$

This provides a briefer way of characterizing open sets: $\mathcal{U} \subset X$ is open if and only if for every $x_0 \in \mathcal{U}$, there exists a continuous seminorm $\|\cdot\|$ such that every $x \in X$ with $\|x - x_0\| < 1$ belongs to $\mathcal{U}$. The sufficiency of this condition is clear since continuity of $\|\cdot\|$ implies that every set of the form $\left\{x \in X \;\middle|\; \|x - x_0\| < 1\right\}$ is open.

Lemma 10.92. *On a topological vector space X, a seminorm $\|\cdot\| : X \to \mathbb{R}$ is continuous if and only if the set $\left\{x \in X \;\middle|\; \|x\| < 1\right\} \subset X$ is open.*

Proof. In one direction, the implication is an immediate consequence of the definition of continuity and the fact that $(-1, 1) \subset \mathbb{R}$ is open. For the converse, we use the fact that for every $x_0 \in X$ and $\epsilon > 0$, the invertible affine map $\Phi : X \to X : x \mapsto x_0 + \epsilon x$ is a homeomorphism, thus if $B := \left\{x \in X \;\middle|\; \|x\| < 1\right\}$ is open, then so is $\Phi(B) = \left\{x \in X \;\middle|\; \|x - x_0\| < \epsilon\right\}$. Given this, if $\mathcal{V} \subset [0, \infty)$ is any open subset and $x_0 \in X$ satisfies $\|x_0\| \in \mathcal{V}$, then choosing any $\epsilon > 0$ such that $(\|x_0\| - \epsilon, \|x_0\| + \epsilon) \subset \mathcal{V}$, the triangle inequality implies that every $x \in X$ in the open set $\{x \in X \mid \|x - x_0\| < \epsilon\}$ satisfies $\|x\| \leqslant \|x_0\| + \|x - x_0\| < \|x_0\| + \epsilon$ and $\|x\| \geqslant \|x_0\| - \|x_0 - x\| > \|x_0\| - \epsilon$, so this open subset belongs to the preimage of $\mathcal{V}$ under $\|\cdot\|$, proving that this preimage is open. ☐

Lemma 10.93. *Suppose X is a locally convex space whose topology is determined by the family of seminorms $\{\|\cdot\|_\alpha\}_{\alpha\in I}$. Then a seminorm $\|\cdot\|$ on X is continuous if and only if there exists a nonempty finite subset $I_0 \subset I$ and a constant $C > 0$ such that*

$$\|x\| \leqslant C \sum_{\alpha\in I_0} \|x\|_\alpha \quad \text{for all} \quad x \in X.$$

Proof. We claim first that if $\|\cdot\|_1$ is a continuous seminorm and $\|\cdot\| \leqslant \|\cdot\|_1$, then $\|\cdot\|$ is also continuous. Indeed, consider $B := \{x \in X \mid \|x\| < 1\}$, and for any $x_0 \in B$, choose $\epsilon > 0$ such that $\|x_0\| + \epsilon < 1$. If $\|\cdot\|_1$ is continuous, then the set $\mathcal{U} := \{x \in X \mid \|x - x_0\|_1 < \epsilon\}$ is an open neighborhood of x_0, and if $\|\cdot\| \leqslant \|\cdot\|_1$, then every $x \in \mathcal{U}$ satisfies

$$\|x\| \leqslant \|x_0\| + \|x - x_0\| \leqslant \|x_0\| + \|x - x_0\|_1 < \|x_0\| + \epsilon < 1,$$

implying $\mathcal{U} \subset B$. This proves that $B \subset X$ is open, so by Lemma 10.92, $\|\cdot\|$ is continuous.

By the assumptions of the lemma, the seminorms $\|\cdot\|_\alpha$ are continuous for all $\alpha \in I$, thus $C\sum_{\alpha\in I_0}\|\cdot\|_\alpha$ is also a continuous seminorm for any $C > 0$ and any finite set $I_0 \subset I$. The claim in the previous paragraph thus implies one direction of the lemma.

For the other direction, assume $\|\cdot\|$ is continuous, so $B := \{x \in X \mid \|x\| < 1\}$ is an open set. Since the family of seminorms $\{\|\cdot\|_\alpha\}_{\alpha\in I}$ generates the topology of X, it follows that B contains a neighborhood of $0 \in X$ in the form

$$\mathcal{U} := \{x \in X \mid \|x\|_\alpha < \epsilon_\alpha \text{ for every } \alpha \in I_0\}$$

for some nonempty finite subset $I_0 \subset I$ and real numbers $\{\epsilon_\alpha > 0\}_{\alpha\in I_0}$. In other words,

$$\|x\|_\alpha < \epsilon_\alpha \text{ for every } \alpha \in I_0 \quad \Rightarrow \quad \|x\| < 1. \tag{10.10}$$

We claim that $\|x\| \leqslant C\sum_{\alpha\in I_0}\|x\|_\alpha$ holds for every $x \in X$, where $C > 0$ is a constant independent of x. There is nothing to prove if $\|x\| = 0$, so consider $x \in X$ with $\|x\| > 0$. At least one of the $\|x\|_\alpha$ for $\alpha \in I_0$ must then also be positive, as otherwise multiplying x by a sufficiently large positive scalar would produce a contradiction to (10.10). The quotient

$$Q(x) := \frac{\|x\|}{\sum_{\alpha\in I_0}\|x\|_\alpha}$$

is therefore well defined whenever $\|x\| > 0$, and we claim that on this subset of X, it is bounded. If not, then there exists a sequence $x_j \in X$ with $\|x_j\| > 0$ and $Q(x_j) \to \infty$. But each x_j can be multiplied by a positive scalar without changing the value of $Q(x_j)$, thus we are free to assume without loss of generality that the denominator in the definition of $Q(x_j)$ some fixed constant less than $\min_{\alpha\in I_0}\epsilon_\alpha$ for every j. In this case, (10.10) implies that the numerator is less than 1 and thus gives a bound on $Q(x_j)$, which is a contradiction. ☐

Lemma 10.94. *For two locally convex spaces X and Y, a linear map $\Lambda : X \to Y$ is continuous if and only if for every continuous seminorm $\|\cdot\|_Y$ on Y, there exists a continuous seminorm $\|\cdot\|_X$ on X such that $\|\Lambda(x)\|_Y \leqslant \|x\|_X$.*

Proof. Assume the second condition holds, $\mathcal{V} \subset Y$ is an open set, and $x_0 \in X$ is a point with $y_0 := \Lambda(x_0) \in \mathcal{V}$. The openness of $\mathcal{V}$ implies that for some continuous seminorm which we will denote by $\|\cdot\|_Y$, $\{y \in Y \mid \|y - y_0\|_Y < 1\}$ defines an open neighborhood of y that is contained in $\mathcal{V}$. If $\|\cdot\|_X$ is a continuous seminorm on X satisfying $\|\Lambda(x)\|_Y \leqslant \|x\|_X$ for all X, it follows that $\{x \in X \mid \|x - x_0\|_X < 1\}$ is an open neighborhood of x_0 in X such that for all x in this neighborhood, $\|\Lambda(x) - y_0\|_Y = \|\Lambda(x - x_0)\|_Y \leqslant \|x - x_0\|_X < 1$, implying $x \in \Lambda^{-1}(\mathcal{V})$. This proves that $\Lambda^{-1}(\mathcal{V}) \subset X$ is open and thus that Λ is continuous.

Conversely, suppose Λ is continuous and $\|\cdot\|_Y$ is an arbitrary continuous seminorm on Y. Then $B := \{y \in Y \mid \|y\|_Y < 1\}$ is open, hence $\Lambda^{-1}(B) \subset X$ is an open neighborhood of 0 and therefore contains $\mathcal{U} := \{x \in X \mid \|x\|_X < 1\}$ for some continuous seminorm $\|\cdot\|_X$ on X. In other words, we have

$$\|x\|_X < 1 \qquad \Rightarrow \qquad \|\Lambda(x)\|_Y < 1 \tag{10.11}$$

for all $x \in X$. We claim that $\|\Lambda(x)\|_Y \leqslant \|x\|_X$ holds for all $x \in X$. If $\|\Lambda(x)\|_Y = 0$ there is nothing to prove, so assume $\|\Lambda(x)\| > 0$. Then $\|x\|_X$ must also be positive, as otherwise multiplying x by a sufficiently large positive scalar produces a contradiction to (10.11). It follows that the quotient $Q(x) := \|\Lambda(x)\|_Y/\|x\|_X > 0$ is well defined whenever its numerator is nonzero, and clearly it does not change if x is multiplied by any positive scalar, thus we are free to assume $\|x\|_X = 1-\epsilon$ for any $\epsilon > 0$ arbitrarily small. Under this assumption, (10.11) implies $|\Lambda(x)\|_Y < 1$ and thus $Q(x) < 1/(1-\epsilon)$; since $\epsilon > 0$ was arbitrary, it follows that $Q(x) \leqslant 1$. ☐

Proposition 10.95. *For any locally convex space X, a linear map $\Lambda : \mathscr{D}(\Omega) \to X$ is continuous if and only if its restrictions $\Lambda|_{\mathscr{D}_K(\Omega)} : \mathscr{D}_K(\Omega) \to X$ are continuous for all compact $K \subset \Omega$.*

Proof. Since the inclusions $\mathscr{D}_K(\Omega) \hookrightarrow \mathscr{D}(\Omega)$ are continuous, the statement is obvious in one direction. We need to show that if Λ has a continuous restriction to every $\mathscr{D}_K(\Omega)$, then it is continuous on $\mathscr{D}(\Omega)$. For this, it will be convenient to choose an open covering of Ω by countably many subsets $\{\Omega_j\}_{j\in\mathbb{N}}$ with the following properties:

(1) The covering is **locally finite**, i.e. every point in Ω has a neighborhood that intersects at most finitely many of the Ω_j;
(2) $K_j := \bar{\Omega}_j$ is compact for every j.

For a concrete construction of $\{\Omega_j\}_{j\in\mathbb{N}}$, choose a strictly increasing sequence $r_j > 0$ with $\lim_{j\to\infty} r_j = \sup\{|x| \mid x \in \Omega\}$, another sequence $\epsilon_j > 0$ such that $r_1 - \epsilon_1 > 0$ and $r_j - \epsilon_j > r_{j-1}$ for every $j \geqslant 2$, and define

$$\Omega_j := \left\{x \in \Omega \;\middle|\; r_{j-1} - \epsilon_{j-1} < |x| < r_j\right\}$$

where for $j = 1$ we interpret the lower bound on $|x|$ as a vacuous condition. With this construction, it is clear that one can also find a sequence of smooth functions $\rho_j : \Omega \to [0,1]$ such that each ρ_j is supported in Ω_j and $\sum_{j=1}^\infty \rho_j \equiv 1$, where the sum is finite at every point due to the local finiteness of the open covering.[31] Any $\varphi \in \mathscr{D}(\Omega)$ can now be decomposed as

$$\varphi = \sum_{j=1}^\infty \varphi_j, \qquad \text{where} \qquad \varphi_j := \rho_j\varphi \text{ has support in } K_j.$$

Observe that for every $\varphi \in \mathscr{D}(\Omega)$, only finitely many of the functions φ_j can be nonzero: indeed, the local finiteness of the covering $\{\Omega_j\}$ implies that at most finitely many of the sets Ω_j can intersect the compact set $\operatorname{supp}(\varphi)$.

To show that $\Lambda : \mathscr{D}(\Omega) \to X$ is continuous, it suffices by Lemma 10.94 to show that for any continuous seminorm $\|\cdot\|_X$ on X, there exists a good seminorm $\|\cdot\|$ on $\mathscr{D}(\Omega)$ such that $\|\Lambda(\varphi)\|_X \leqslant \|\varphi\|$ for all $\varphi \in \mathscr{D}(\Omega)$. The topology of $\mathscr{D}_{K_j}(\Omega)$ for each $j \in \mathbb{N}$ is generated by the monotone sequence of norms $\|\cdot\|_{C^m}$ for $m = 0,1,2,\ldots$, thus continuity of Λ on $\mathscr{D}_{K_j}(\Omega)$ implies that there exists an integer $m_j \geqslant 0$ and a positive number c_j such that

$$\|\Lambda(\psi)\|_X \leqslant c_j\|\psi\|_{C^{m_j}} \quad \text{for all} \quad \psi \in \mathscr{D}_{K_j}(\Omega).$$

Since the sum $\varphi = \sum_j \varphi_j$ is finite for each $\varphi \in \mathscr{D}(\Omega)$ and $\varphi_j \in \mathscr{D}_{K_j}(\Omega)$ for $j = 1,2,3,\ldots$, we can apply the triangle inequality and write

$$\|\Lambda(\varphi)\|_X \leqslant \sum_j \|\Lambda(\varphi_j)\|_X \leqslant \sum_j c_j\|\varphi_j\|_{C^{m_j}} \leqslant c_j'\|\varphi\|_{C^{m_j}},$$

where each of the modified constants $c_j' > 0$ depends on the C^{m_j}-norm of ρ_j but not on φ. With these constants fixed, it is easy to check that

$$\|\varphi\| := \sum_{j=1}^\infty c_j'\|\varphi\|_{C^{m_j}(\Omega_j)}$$

defines a good seminorm on $\mathscr{D}(\Omega)$, as for any compact $K \subset \Omega$, the restriction of this seminorm to $\mathscr{D}_K(\Omega)$ has only finitely many nonzero terms, and C^∞-convergence in $\mathscr{D}_K(\Omega)$ implies that

[31] A collection of functions ρ_j with these properties is called a **partition of unity** on Ω.

each individual term converges. Since $\|\Lambda\varphi\|_X \leqslant \|\varphi\|$ by construction, this establishes that $\Lambda : \mathscr{D}(\Omega) \to Y$ is continuous. □

The following easy consequence completes the proof of Proposition 10.8:

Corollary 10.96. *For any locally convex space X, a linear map $\Lambda : \mathscr{D}(\Omega) \to X$ is continuous if and only if it is sequentially continuous, i.e. for every convergent sequence $\varphi_j \to \varphi_\infty$ in $\mathscr{D}(\Omega)$, $\Lambda(\varphi_j) \to \Lambda(\varphi_\infty)$.*

Proof. By a standard result in point-set topology, continuous maps are always sequentially continuous. Conversely, if $\Lambda : \mathscr{D}(\Omega) \to X$ is sequentially continuous, then its restriction to $\mathscr{D}_K(\Omega)$ for each compact set $K \subset \Omega$ is sequentially continuous, and since $\mathscr{D}_K(\Omega)$ is metrizable, it follows that the restriction of Λ to $\mathscr{D}_K(\Omega)$ is also continuous. By Proposition 10.95, Λ itself is therefore continuous. □

Remark 10.97. By another standard result in point-set topology (see e.g. [Wen18, §4]), a sequentially continuous map $f : X \to Y$ between two topological spaces is continuous whenever X is first countable. We did not claim this to be true for $\mathscr{D}(\Omega)$, which is *not* first countable according to Remark 10.89, but Corollary 10.96 says that it is nonetheless true specifically for *linear* maps to other locally convex spaces. Philosophically, the reason this works is that $\mathscr{D}(\Omega)$ can be viewed—in a sense to be made precise in §10.8.2 below—as a *limit* of a family of spaces in which sequential continuity does imply continuity, namely the metrizable spaces $\mathscr{D}_K(\Omega)$ for $K \subset \Omega$ compact.

10.8.2. *Inductive limits.* The topology we've defined on $\mathscr{D}(\Omega)$ is often referred to as an *inductive limit* topology. While one can understand all of its properties without knowing what this term means, let us take a moment to discuss the wider context in which it arises.

We need to introduce a few notions from abstract category theory. For the particular application relevant here, the "category" we have in mind is the class of locally convex spaces (these are the *objects* of the category), and the natural class of maps between two such spaces consists of all continuous linear maps (these are the *morphisms* of the category). We shall formulate the definitions below in terms of this particular category just for concreteness, but they would still make sense in any other category, e.g. topological spaces and continuous maps, vector spaces and linear maps, groups and group homomorphisms, and so forth.

Suppose I is a set with a pre-order $\prec$, i.e. $\prec$ is reflexive ($\alpha \prec \alpha$) and transitive ($\alpha \prec \beta$ and $\beta \prec \gamma$ implies $\alpha \prec \gamma$), but the relations $\alpha \prec \beta$ and $\beta \prec \alpha$ need not imply $\alpha = \beta$, so $\prec$ need not be a partial order. The pair $(I, \prec)$ is called a **directed set** if for every pair $\alpha, \beta \in I$, there exists $\gamma \in I$ with $\gamma \succ \alpha$ and $\gamma \succ \beta$. An obvious example is $\mathbb{N}$ with its usual total order $\prec := \leqslant$. A more interesting and relevant example for our purposes is to define I as the set of all compact subsets in a fixed open set $\Omega \subset \mathbb{R}^n$, with $K \prec K'$ defined to mean $K \subset K'$. Notice that the ordering relation in this example is a partial order, but not a total order since for any two compact subsets, it need not be true that either is contained in the other. It forms a directed set because whenever $K, K' \subset \Omega$ are both compact, $K \cup K'$ is another compact subset of Ω that contains both of them.

Definition 10.98. A **direct system** (or **inductive system**) of locally convex spaces consists of a directed set $(I, \prec)$ and a family of locally convex spaces $\{X_\alpha\}_{\alpha \in I}$ together with continuous linear maps $\varphi_{\beta\alpha} : X_\alpha \to X_\beta$ defined for each $\alpha, \beta \in I$ with $\alpha \prec \beta$, such that

$$\varphi_{\alpha\alpha} = \mathrm{Id}_{X_\alpha}$$

and the diagram

$$\begin{tikzcd} X_\alpha \arrow[r, "\varphi_{\beta\alpha}"] \arrow[rr, bend right, "\varphi_{\gamma\alpha}"'] & X_\beta \arrow[r, "\varphi_{\gamma\beta}"] & X_\gamma \end{tikzcd}$$

commutes for every triple $\alpha, \beta, \gamma \in I$ with $\alpha \prec \beta \prec \gamma$.

The notion of "convergence" for a direct system must necessarily look somewhat different from what we've seen before for sequences, as there is no meaningful topology to be defined on the "set" of all locally convex spaces.[32] The idea is instead to measure the convergence of a direct system $\{X_\alpha, \varphi_{\beta\alpha}\}$ in terms of the continuous linear maps from each X_α to other fixed spaces.

Definition 10.99. For a direct system $\{X_\alpha, \varphi_{\beta\alpha}\}$ of locally convex spaces over the directed set $(I, \prec)$, a **target** $\{Y, f_\alpha\}$ of the system consists of a locally convex space Y together with associated continuous linear maps $f_\alpha : X_\alpha \to Y$ for each $\alpha \in I$ such that the diagram

$$\begin{tikzcd} X_\alpha \arrow[rr, "\varphi_{\beta\alpha}"] \arrow[dr, "f_\alpha"] & & X_\beta \arrow[dl, "f_\beta"] \\ & Y & \end{tikzcd}$$

commutes for every pair $\alpha, \beta \in I$ with $\alpha \prec \beta$.

Definition 10.100. A target $\{X_\infty, \varphi_\alpha\}$ of the direct system $\{X_\alpha, \varphi_{\beta\alpha}\}$ is called a **direct limit** (or **inductive limit** or **colimit**) of the system and written as

$$X_\infty = \varinjlim\{X_\alpha\}$$

if it satisfies the following "universal" property: for all targets $\{Y, f_\alpha\}$ of $\{X_\alpha, \varphi_{\beta\alpha}\}$, there exists a unique continuous linear map $f_\infty : X_\infty \to Y$ such that the diagram

$$\begin{tikzcd} X_\alpha \arrow[r, "\varphi_\alpha"] \arrow[dr, "f_\alpha"'] & X_\infty \arrow[d, "f_\infty"] \\ & Y \end{tikzcd}$$

commutes for every $\alpha \in I$.

The essential meaning of a direct limit can be encoded in the diagram

$$\begin{tikzcd} X_\alpha \arrow[r, "\varphi_{\beta\alpha}"] \arrow[drrrr] & X_\beta \arrow[r, "\varphi_{\gamma\beta}"] \arrow[drrr] & X_\gamma \arrow[r] \arrow[drr] & \ldots \arrow[r] & \varinjlim\{X_\alpha\} \arrow[d, dashed] \\ & & & & Y \end{tikzcd}$$

where we assume $\alpha \prec \beta \prec \gamma \prec \ldots \in I$. The key feature of the space $\varinjlim\{X_\alpha\}$ is that whenever a space Y and continuous linear maps $X_\alpha \to Y$ in a commuting diagram of this type are given, the "limit" map from $\varinjlim\{X_\alpha\}$ to Y indicated by the dashed arrow must also exist (as a continuous linear map) and be unique.

Note that these definitions on their own give no guarantee for any given direct system that a direct limit must exist, and if it exists, then it is generally not unique. Indeed:

Exercise 10.101. If $\{X, f_\alpha\}$ is a direct limit of $\{X_\alpha, \varphi_{\beta\alpha}\}$ and Y is another locally convex space such that there exists a continuous linear isomorphism $\psi : X \to Y$ with a continuous inverse, show that $\{Y, \psi \circ f_\alpha\}$ is also a direct limit of $\{X_\alpha, \varphi_{\beta\alpha}\}$.
Remark: The invertibility of ψ is needed only for showing that $\{Y, \psi \circ f_\alpha\}$ satisfies the universal property; it is already a target without this.

The non-uniqueness exhibited by the exercise above is however the worst thing that can happen: if $\{X, f_\alpha\}$ and $\{Y, g_\alpha\}$ are any two direct limits of the same system $\{X_\alpha, \varphi_{\beta\alpha}\}$, then the universal property provides unique continuous linear maps $g_\infty : X \to Y$ and $f_\infty : Y \to X$ satisfying $g_\infty \circ f_\alpha = g_\alpha$ and $f_\infty \circ g_\alpha = f_\alpha$ for every $\alpha \in I$. It follows that $f_\infty \circ g_\infty$ is the unique continuous linear map $X \to X$ satisfying $(f_\infty \circ g_\infty) \circ f_\alpha = f_\alpha$ for every $\alpha \in I$, which implies

[32] And strictly speaking, the collection of all locally convex spaces is far too large to be called a *set*; it is instead a *proper class*. This remark is included only for the sake of readers who truly care about abstract set theory.

$f_\infty \circ g_\infty = \mathrm{Id}_X$. A similar argument shows $g_\infty \circ f_\infty = \mathrm{Id}_Y$, thus X and Y are isomorphic, and there is a distinguished isomorphism relating them. For this reason, we typically refer to "the" (rather than "a") direct limit of any system for which a limit exists.

Example 10.102. Given an open set $\Omega \subset \mathbb{R}^n$, take $(I, \prec)$ to be the set of all compact subsets $K \subset \Omega$ with $K \prec K'$ defined to mean $K \subset K'$. There is then a direct system $\{X_K, \varphi_{K',K}\}$ over $(I, \prec)$ such that $X_K = \mathscr{D}_K(\Omega)$ and $\varphi_{K',K}$ is the obvious inclusion map $\mathscr{D}_K(\Omega) \hookrightarrow \mathscr{D}_{K'}(\Omega)$, defined whenever $K \subset K'$. Define $\varphi_K : \mathscr{D}_K(\Omega) \hookrightarrow \mathscr{D}(\Omega)$ also as the natural inclusion for each $K \in I$. Proposition 10.95 can then be reinterpreted as the statement that $\{\mathscr{D}(\Omega), \varphi_K\}$ is a universal target for the direct system $\{\mathscr{D}_K(\Omega), \varphi_{K',K}\}$, in other words,

$$\mathscr{D}(\Omega) = \varinjlim \{\mathscr{D}_K(\Omega)\}.$$

This is why the topology of $\mathscr{D}(\Omega)$ is often called the **inductive limit topology** determined by the natural Fréchet space topologies of $\mathscr{D}_K(\Omega)$ for all compact $K \subset \Omega$.

Remark 10.103. One really should call the topology on $\mathscr{D}(\Omega)$ a **locally convex inductive limit topology**, as omitting the words "locally convex" can potentially cause confusion. A topologist would interpret the words "inductive limit topology" to mean a universal target in the sense of Definition 10.100, but with X_∞ and Y allowed in general to be arbitrary topological spaces (not necessarily topological vector spaces), and all maps required to be continuous but not necessarily linear. It is not hard to show that the direct limit in this sense of the system $\{\mathscr{D}_K(\Omega), \varphi_{K',K}\}$ can be identified again with the vector space $\mathscr{D}(\Omega)$, but endowed with an even stronger topology, for which a set $\mathcal{U} \subset \mathscr{D}(\Omega)$ is open if and only if $\mathcal{U} \cap \mathscr{D}_K(\Omega) \subset \mathscr{D}_K(\Omega)$ is open for every compact $K \subset \Omega$. This topology has the same notion of convergent sequences as the locally convex topology we defined, and it has the nice property that for any topological space X, a (not necessarily linear) map $f : \mathscr{D}(\Omega) \to X$ is continuous if and only if its restriction to $\mathscr{D}_K(\Omega)$ is continuous for every compact $K \subset \Omega$. However, since this topology contains sets that are not open in the locally convex inductive limit topology, it cannot be locally convex—in fact there is no good reason to expect $\mathscr{D}(\Omega)$ with this topology to be a topological vector space.

10.8.3. *Comparison with other topologies.* There are other natural topologies one could imagine defining on the space of smooth functions with compact support, and it is natural to wonder why the inductive limit topology defined in §10.8.1 is a better choice. The obvious answer is that since we defined the topology on $\mathscr{D}(\Omega)$ to be as strong as possible while still being locally convex, this makes the space of distributions $\mathscr{D}'(\Omega)$ as large as possible. But let us briefly discuss some alternatives. In order to avoid confusion, we will refer to the space of smooth compactly supported functions $\Omega \to \mathbb{R}$ in this subsection as

$$C_0^\infty(\Omega),$$

reserving the notation $\mathscr{D}(\Omega)$ for the case where this space is endowed with the specific topology from §10.8.1.

Alternative 1: *The C^∞_{loc}-topology.*
The space $C^\infty(\Omega)$ of all (not necessarily compactly supported) smooth functions $\Omega \to \mathbb{R}$ admits a natural Fréchet space topology for which convergence means uniform convergence of derivatives of all orders on compact subsets. This is often called C^∞_{loc}-convergence. A countable family of seminorms for the C^∞_{loc}-topology, also sometimes called the **weak** or **compact-open C^∞-topology**, is given by

$$\|\varphi\|_{m,j} := \|\varphi\|_{C^m(K_j)} \qquad \text{for integers } m \geqslant 0,\ j \geqslant 1,$$

where $K_1 \subset K_2 \subset K_3 \subset \ldots \bigcup_{j\in\mathbb{N}} K_j = \Omega$ is any exhausting sequence of compact subsets such that K_j is contained in the interior of K_{j+1} for every j. This gives the right notion of convergence because every compact set is contained in K_j for j sufficiently large, and it defines a metrizable topology since the family of seminorms is countable (see e.g. [RS80, Theorem V.5]). Continuity on $C^\infty(\Omega)$ is thus equivalent to sequential continuity, and since the notion of C^∞_{loc}-convergence can be expressed without referring to the specific choice of exhaustion $K_1 \subset K_2 \subset K_3 \subset \ldots$, the C^∞_{loc}-topology is also independent of that choice.

As a subspace of $C^\infty(\Omega)$, $C^\infty_0(\Omega)$ inherits a metrizable topology for which convergence means C^∞_{loc}-convergence. The first thing to notice, however, is that $C^\infty_0(\Omega)$ is not a Fréchet space with this topology, i.e. it is not complete, because it is not a *closed* subspace of $C^\infty(\Omega)$. In fact, by choosing any sequence of smooth cutoff functions $\rho_j \in C^\infty_0(\Omega)$ with $\rho_j|_{K_j} \equiv 1$ for every j, it is easy to check that for every $\varphi \in C^\infty(\Omega)$, $\varphi_j := \rho_j \varphi \to \varphi$ in C^∞_{loc}, thus $C^\infty_0(\Omega)$ is dense in $C^\infty(\Omega)$ with respect to the C^∞_{loc}-topology. This has an immediate consequence for the space of continuous linear functionals on $C^\infty_0(\Omega)$: any linear functional $\Lambda : C^\infty_0(\Omega) \to \mathbb{R}$ that is continuous in the C^∞_{loc}-topology must admit a continuous extension to a linear functional on $C^\infty(\Omega)$. Most distributions clearly do not have this property; since functions $\varphi \in C^\infty(\Omega)$ can grow arbitrarily large near infinity or near the boundary of Ω, even globally integrable functions $f : \Omega \to \mathbb{R}$ do not generally define continuous functionals of $\varphi \in C^\infty(\Omega)$ under the pairing $\Lambda_f(\varphi) := \int_\Omega \varphi f \, dm$. On the other hand, it is possible to give a precise characterization of the distributions for which this works.

Proposition 10.104. *A distribution $\Lambda \in \mathscr{D}'(\Omega)$ is continuous with respect to the C^∞_{loc}-topology on $C^\infty_0(\Omega)$ if and only if it has compact support.*

Proof. In one direction, this statement follows from Proposition 10.83. For the converse, continuity of $\Lambda \in \mathscr{D}'(\Omega)$ with respect to C^∞_{loc}-convergence implies since $C^\infty_0(\Omega)$ is dense in $C^\infty(\Omega)$ that Λ extends to a C^∞_{loc}-continuous linear functional on $C^\infty(\Omega)$. If $\mathrm{supp}(\Lambda)$ is not compact, then for every compact set $K \subset \Omega$, there exists a test function $\varphi \in \mathscr{D}(\Omega)$ with $\mathrm{supp}(\varphi) \cap K = \emptyset$ and $\Lambda(\varphi) \neq 0$. We can therefore find an exhausting sequence of compact subsets $K_1 \subset K_2 \subset K_3 \subset \ldots \subset \bigcup_{j \in \mathbb{N}} K_j = \Omega$ and associated test functions $\varphi_1, \varphi_2, \varphi_3, \ldots \in \mathscr{D}(\Omega)$ such that $\mathrm{supp}(\varphi_j) \subset K_j \backslash K_{j-1}$ and $\Lambda(\varphi_j) \neq 0$ for all j. For any choice of constnts $c_j \in \mathbb{R}$, the sequence $\psi_k := \sum_{j=1}^k \varphi_j \in \mathscr{D}(\Omega)$ is then C^∞_{loc}-convergent to a smooth function $\psi_\infty \in C^\infty(\Omega)$, but the constants c_j can easily be chosen to make sure that $\Lambda(\psi_k) = \sum_{j=1}^k c_j \Lambda(\varphi_j)$ diverges as $k \to \infty$, giving a contradiction. ☐

Alternative 2: *The C^∞-topology.*
The countable family of norms $\| \cdot \|_{C^m}$ for integers $m \geqslant 0$ determines a Fréchet space topology on the subspace

$$C^\infty_b(\Omega) := \left\{ \varphi \in C^\infty(\Omega) \;\middle|\; \partial^\alpha \varphi \text{ is bounded for every multi-index } \alpha \right\} \subset C^\infty(\Omega).$$

The associated notion of *C^∞-convergence* means uniform convergence for derivatives of all orders, not just on compact subsets but globally on Ω, thus C^∞-convergence implies (but is not implied by) C^∞_{loc}-convergence, and the C^∞-topology on the subspace $C^\infty_0(\Omega)$ is strictly stronger than the C^∞_{loc}-topology. This sounds like good news for the theory of distributions, as it means that the space of C^∞-continuous linear functionals is larger than the space of compactly supported distributions considered in Proposition 10.104. But the next exercise shows that it is still not large enough to contain all locally integrable functions.

Exercise 10.105. Find a sequence $\varphi_j \in C^\infty_0(\mathbb{R})$ such that $\varphi_j \to 0$ in the C^∞-topology, but $\int_{\mathbb{R}} \varphi_j \, dm = 1$ for all j. This implies that the distribution $\Lambda_f : \mathscr{D}(\Omega) \to \mathbb{R}$ defined via the locally integrable function $f := 1$ on $\mathbb{R}$ is not continuous with respect to the C^∞-topology.

A further hint that the C^∞-topology is not an ideal choice for $\mathscr{D}(\Omega)$ arises from the observation that $C^\infty_0(\Omega)$ is not a C^∞-closed subspace of $C^\infty_b(\Omega)$; one can easily find C^∞-convergent sequences of compactly supported functions whose limits do not have compact support. It follows that every C^∞-continuous linear functional on $C^\infty_0(\Omega)$ must admit a continuous extension to a subspace of $C^\infty_b(\Omega)$ that is strictly larger than $C^\infty_0(\Omega)$. Exercise 10.105 shows that even relatively tame functions like $f \equiv 1$ on $\mathbb{R}$ need not define distributions that are extendable in this sense.

Remark 10.106. One other theoretical drawback of the C^∞-topology is worth mentioning. All the other topologies discussed in this section can be defined in more general contexts, e.g. for the space of compactly supported smooth functions on a finite-dimensional manifold, and the weak and strong C^∞-topologies can even be defined for spaces of smooth maps from one finite-dimensional manifold to another. Generalizations of this type are essential for certain fundamental perturbation results in differential topology (see e.g. [Hir94, Chapter 2]). The definitions in this setting become more complicated, as they necessarily involve choices of local coordinate charts, and one must then verify that the topologies defined in this way are independent of choices. For the weak C^∞-topology and its strong variants to be discussed below, this is not difficult, because while the C^m-norm of a function can certainly change if one composes the function with a smooth coordinate transformation, this change can be bounded as long as it is only being considered over a compact subset. The ordinary C^∞-topology for functions $\Omega \to \mathbb{R}$ is simpler to define, but since it involves C^m-norms over noncompact sets, it does not have such coordinate-invariant properties and thus cannot be defined in a meaningful way for functions on a noncompact manifold.

Alternative 3: *The (strong) Whitney C^∞-topology.*
To describe the Whitney C^∞-topology, one should first describe the Whitney C^m-topology for $0 \leqslant m < \infty$ on $C^\infty(\Omega)$. We give two definitions: first, it is the smallest topology containing all sets of the form

$$\mathcal{U}(\varphi, \alpha, f) := \left\{\psi \in C^\infty(\Omega) \;\middle|\; |\partial^\alpha(\psi - \varphi)| < f\right\}$$

for arbitrary choices of $\varphi \in C^\infty(\Omega)$, multi-indices α of order at most m and continuous functions $f : \Omega \to (0, \infty)$. Equivalently, one can generate this topology with sets of the form

$$\mathcal{V}(\varphi, \{\Omega_i\}, \{k_i\}, \{\epsilon_i\}) := \left\{\psi \in C^\infty(\Omega) \;\middle|\; \|\psi - \varphi\|_{C^{k_i}(\Omega_i)} < \epsilon_i\right\}, \tag{10.12}$$

for all positive choices of $\varphi \in C^\infty(\Omega)$, locally finite open coverings $\{\Omega_i\}_{i\in I}$ of Ω, and collections of numbers $\{k_i \in \{0, \dots, m\}\}_{i\in I}$ and $\{\epsilon_i > 0\}_{i\in I}$.

Exercise 10.107. Show that the two definitions of the Whitney C^m-topology given above are equivalent.

The **Whitney C^∞-topology** is now defined to be the smallest topology on $C^\infty(\Omega)$ that contains the Whitney C^m-topology for every $m \geqslant 0$, i.e. it is generated by the sets $\mathcal{U}(\varphi, \alpha, f)$ without any bound on the order of the multi-index α, or by $\mathcal{V}(\varphi, \{\Omega_i\}, \{k_i\}, \{\epsilon_i\})$, in which the set of integers $\{k_i \geqslant 0\}_{i\in I}$ is always required to be bounded, but no fixed bound is imposed.

It is straightforward to transform the definitions of $\mathcal{U}(\varphi, \alpha, f) \subset C^\infty(\Omega)$ and $\mathcal{V}(\varphi, \{\Omega_i\}, \{k_i\}, \{\epsilon_i\})$ into conditions of the form $\{\|\psi - \varphi\| < 1\}$ for suitable seminorms $\|\cdot\|$, thus the Whitney C^∞-topology is locally convex. One can however use the argument of Remark 10.89 to show that it is not first countable, and thus not metrizable. Here is a clear advantage of the Whitney topology in comparison with the C^∞ and C^∞_{loc}-topologies:

Proposition 10.108. *In the Whitney C^∞-topology, $C_0^\infty(\Omega)$ is a closed subspace of $C^\infty(\Omega)$.*

Proof. One needs to show that $C^\infty(\Omega) \backslash C_0^\infty(\Omega)$ is open. If $\varphi \in C^\infty(\Omega)$ does not have compact support, then there exists a continuous function $f : \Omega \to (0, \infty)$ and a sequence $x_j \in \Omega$ with no accumulation point such that $f(x_j) < |\varphi(x_j)|$ for all j. The set $\left\{\psi \in C^\infty(\Omega) \;\middle|\; |\psi - \varphi| < f\right\}$ is then a Whitney-open neighborhood of φ consisting of functions ψ that satisfy $\psi(x_j) \neq 0$ for all j and thus never have compact support. □

A similar argument to Lemma 10.90 and Corollary 10.91 also shows:

Proposition 10.109. *A sequence in $C_0^\infty(\Omega)$ converges in the Whitney C^∞-topology if and only if it converges in $\mathscr{D}(\Omega)$.* □

One can easily show that the seminorms one uses to define the open sets $\mathcal{U}(\varphi, \alpha, f)$ or $\mathcal{V}(\varphi, \{\Omega_i\}, \{k_i\}, \{\epsilon_i\})$ are also *good* seminorms in the sense of Definition 10.87, thus the topology of $\mathscr{D}(\Omega)$ contains the Whitney C^∞-topology, and Proposition 10.109 reveals that the two topologies are evidently quite similar. In fact, the Whitney C^0-topology is already strong enough to make all functionals of the form $\Lambda_f(\varphi) = \int_\Omega \varphi f \, dm$ for $f \in L^1_{\mathrm{loc}}(\Omega)$ continuous, thus the Whitney C^∞-topology also has this clearly desirable property. The next example shows however that the topology of $\mathscr{D}(\Omega)$ is *strictly* stronger, so that the space of distributions is still strictly larger than the space of linear functionals on $C_0^\infty(\Omega)$ that are continuous in the Whitney topology. In light of Corollary 10.96, this also reveals that for the Whitney topology on $C_0^\infty(\Omega)$, sequential continuity of a linear functional does not imply continuity.

Example 10.110. Consider the real-valued distribution $\Lambda : \mathscr{D}(\mathbb{R}) \to \mathbb{R}$ defined by

$$\Lambda(\varphi) := \sum_{k=0}^{\infty} \varphi^{(k)}(k).$$

This is well defined on any individual test function $\varphi \in \mathscr{D}(\mathbb{R})$ since only finitely many terms in the sum are nonzero, and the same is true for any convergent sequence of test functions, thus Λ is sequentially continuous and therefore continuous on $\mathscr{D}(\mathbb{R})$. But it is not continuous with respect to the Whitney C^∞-topology on $C_0^\infty(\mathbb{R})$. To see this, consider $\Lambda^{-1}((-1,1))$. If this were open in the Whitney topology, then there would need to exist a finite collection of multi-indices $\alpha_1, \dots, \alpha_N$ and continuous functions $f_1, \dots, f_N : \mathbb{R} \to (0, \infty)$ such that $\bigcap_{j=1}^N \mathcal{U}(0, \alpha_j, f_j) \subset \Lambda^{-1}((-1,1))$, meaning

$$|\partial^{\alpha_j} \varphi| < f_j \text{ for all } j = 1, \dots, N \qquad \Rightarrow \qquad |\Lambda(\varphi)| < 1.$$

But this condition constrains only finitely many derivatives of φ, thus one can always find a function that satisfies it but has $|\Lambda(\varphi)| \geqslant 1$ due to the behavior of some derivative of even higher order.

Alternative 4: *The (very) strong C^∞-topology.* A minor modification to the definition of the Whitney C^∞-topology gives rise to an even stronger topology which we shall refer to as the *strong C^∞-topology.*[33] It is generated by all sets of the form $\mathcal{V}(\varphi, \{\Omega_i\}, \{k_i\}, \{\epsilon_i\})$ as in (10.12), for arbitrary locally finite open coverings $\{\Omega_i\}_{i \in I}$, sets of nonnegative integers $\{k_i\}_{i \in I}$ and positive numbers $\{\epsilon_i\}_{i \in I}$. The crucial difference is that in our definition of the Whitney C^∞-topology, the set of integers $\{k_i\}_{i \in I}$ was always required to be bounded, and this is no longer required. Note that since the open covering $\{\Omega_i\}_{i \in I}$ is locally finite, only finitely many of the sets can intersect any given compact subset of Ω, but there still may be infinitely many sets in the covering. The result is that neighborhoods generating the strong topology are required to satisfy conditions on only finitely many derivatives over each individual compact subset, but globally on Ω, there may be conditions on derivatives of all orders.

If one only considers convergence of sequences, then there is no difference between the strong and Whitney C^∞-topologies: Proposition 10.109 admits the same proof for the strong topology and shows that it also has the same notion of convergence as $\mathscr{D}(\Omega)$. That it is nonetheless *strictly* stronger than the Whitney topology follows from Example 10.110 and the following:

Exercise 10.111. Show that the strong C^∞-topology on $C_0^\infty(\Omega)$ is equivalent to the topology of $\mathscr{D}(\Omega)$.

The strong C^∞-topology is thus merely a different perspective on the locally convex inductive limit topology, one that does not require talking about the Fréchet subspaces $\mathscr{D}_K(\Omega)$ with $K \subset \Omega$ compact. This approach to the topology of $\mathscr{D}(\Omega)$ is discussed in more detail in [Hor66, §2.12], which gives in particular an explicit family of good seminorms generating the topology.

[33]The literature is not unanimous on the terminology for these topologies: different sources may use the words "strong topology" or "Whitney topology" to refer to either of alternatives 3 and 4, and one occasionally even finds an authoritative source that fails to distinguish between them. (I am thinking especially of [Hir94], which defines the Whitney topology in §2.1 and the strong topology in §2.4 but states erroneously that they are equivalent.) Alternative 4 is occasionally also called the *very strong C^∞*-topology, e.g. in [Ill03].

References

[AF03] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, 2nd ed., Pure and Applied Mathematics (Amsterdam), vol. 140, Elsevier/Academic Press, Amsterdam, 2003.

[AE01] H. Amann and J. Escher, *Analysis. III*, Grundstudium Mathematik. [Basic Study of Mathematics], Birkhäuser Verlag, Basel, 2001 (German, with German summary).

[Ban31] S. Banach, *Über die Baire'sche Kategorie gewisser Funktionenmengen*, Studia Mathematica **3** (1931), 174–179.

[BS18] T. Bühler and D. A. Salamon, *Functional analysis*, Graduate Studies in Mathematics, vol. 191, American Mathematical Society, Providence, RI, 2018.

[Cla36] J. A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc. **40** (1936), no. 3, 396–414.

[DM72] H. Dym and H. P. McKean, *Fourier series and integrals*, Academic Press, New York-London, 1972. Probability and Mathematical Statistics, No. 14.

[FH15] P. M. Fitzpatrick and B. R. Hunt, *Absolute continuity of a function and uniform integrability of its divided differences*, Amer. Math. Monthly **122** (2015), no. 4, 362–366.

[Har16] G. H. Hardy, *Weierstrass's non-differentiable function*, Trans. Amer. Math. Soc. **17** (1916), no. 3, 301–325.

[Hei05] J. Heinonen, *Lectures on Lipschitz analysis*, Report. University of Jyväskylä Department of Mathematics and Statistics, vol. 100, University of Jyväskylä, Jyväskylä, 2005.

[Hir94] M. W. Hirsch, *Differential topology*, Springer-Verlag, New York, 1994.

[Hor66] J. Horváth, *Topological vector spaces and distributions. Vol. I*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1966.

[Ill03] S. Illman, *The very-strong C^∞ topology on $C^\infty(M,N)$ and K-equivariant maps*, Osaka J. Math. **40** (2003), no. 2, 409–428.

[Joh10] J. Johnsen, *Simple proofs of nowhere-differentiability for Weierstrass's function and cases of slow growth*, J. Fourier Anal. Appl. **16** (2010), no. 1, 17–33.

[Kat04] Y. Katznelson, *An introduction to harmonic analysis*, 3rd ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2004.

[Lan93] S. Lang, *Real and functional analysis*, 3rd ed., Springer-Verlag, New York, 1993.

[LL01] E. H. Lieb and M. Loss, *Analysis*, 2nd ed., Graduate Studies in Mathematics, vol. 14, American Mathematical Society, Providence, RI, 2001.

[Maz31] S. Mazurkiewicz, *Sur les fonctions non dérivables*, Studia Mathematica **3** (1931), 92–94.

[RS80] M. Reed and B. Simon, *Methods of modern mathematical physics. I*, 2nd ed., Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1980. Functional analysis.

[RSN90] F. Riesz and B. Sz.-Nagy, *Functional analysis*, Dover Books on Advanced Mathematics, Dover Publications, Inc., New York, 1990. Translated from the second French edition by Leo F. Boron; Reprint of the 1955 original.

[Roy88] H. L. Royden, *Real analysis*, 3rd ed., Macmillan Publishing Company, New York, 1988.

[RF10] H. L. Royden and P. M. Fitzpatrick, *Real analysis*, 4th ed., Pearson, 2010.

[Rud76] W. Rudin, *Principles of mathematical analysis*, 3rd ed., McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976. International Series in Pure and Applied Mathematics.

[Rud87] ———, *Real and complex analysis*, 3rd ed., McGraw-Hill Book Co., New York, 1987.

[Sal16] D. A. Salamon, *Measure and integration*, EMS Textbooks in Mathematics, European Mathematical Society (EMS), Zürich, 2016. MR3469972

[Sch99] E. Schechter, *Handbook of analysis and its foundations*, Academic Press, Inc., San Diego, CA, 1999.

[Shi18] N. Shioji, *Simple proofs of the uniform convexity of L^p and the Riesz representation theorem for L^p*, Amer. Math. Monthly **125** (2018), no. 8, 733–738.

[Tay96] M. E. Taylor, *Partial differential equations. I*, Applied Mathematical Sciences, vol. 115, Springer-Verlag, New York, 1996.

[Tes] G. Teschl, *Topics in Linear and Nonlinear Functional Analysis*, Graduate Studies in Mathematics, American Mathematical Society, Providence, RI. to appear; manuscript available at `https://www.mat.univie.ac.at/~gerald/ftp/book-fa/`.

[Wen18] C. Wendl, *Topology I and II* (2018). Notes from the course at HU Berlin, available at `https://www.mathematik.hu-berlin.de/~wendl/Winter2018/Topologie2/lecturenotes.pdf`.

[Wen19] ———, *Parameterabhängige Lebesgue-Integrale* (2019). Skript zur Vorlesung Analysis III an der HU Berlin, verfügbar unter `https://www.mathematik.hu-berlin.de/~wendl/Winter2019/Analysis3/Skript_parameter.pdf`

Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

Email address: `wendl@math.hu-berlin.de`