Analysis of Functions

Dr. Claude Warnick

May 1, 2021

Schedule

ANALYSIS OF FUNCTIONS (D) **24 lectures, Lent Term**

Part II Linear Analysis and Part II Probability and Measure are essential.

Lebesgue integration theory
Review of integration: simple functions, monotone and dominated convergence; existence of Lebesgue measure; definition of L^p spaces and their completeness. The Lebesgue differentiation theorem. Egorov's theorem, Lusin's theorem. Mollification by convolution, continuity of translation and separability of L^p when $p \neq \infty$. [5]

Banach and Hilbert space analysis
Strong, weak and weak-* topologies; reflexive spaces. Review of the Riesz representation theorem for Hilbert spaces; the Radon–Nikodym theorem; the dual of L^p. Compactness: review of the Ascoli–Arzelà theorem; weak-* compactnesss of the unit ball for separable Banach spaces. The Riesz representation theorem for spaces of continuous functions. The Hahn–Banach theorem and its consequences: separation theorems; Mazur's theorem. [7]

Fourier analysis
Definition of Fourier transform in L^1; the Riemann–Lebesgue lemma. Fourier inversion theorem. Extension to L^2 by density and Plancherel's isometry. Duality between regularity in real variable and decay in Fourier variable. [3]

Generalized derivatives and function spaces
Definition of generalized derivatives and of the basic spaces in the theory of distributions: $\mathcal{D}/\mathcal{D}'$ and $\mathcal{S}/\mathcal{S}'$. The Fourier transform on $\mathcal{S}'$. Periodic distributions; Fourier series; the Poisson summation formula. Definition of the Sobolev spaces H^s in $\mathbb{R}^d$. Sobolev embedding. The Rellich–Kondrashov theorem. The trace theorem. [5]

Applications
Construction and regularity of solutions for elliptic PDEs with constant coefficients on $\mathbb{R}^n$. Construction and regularity of solutions for the Dirichlet problem of Laplace's equation. The spectral theorem for the Laplacian on a bounded domain. *The direct method of the Calculus of Variations.* [4]

Appropriate books

H. Brézis *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* Universitext, Springer 2011

A.N. Kolmogorov, S.V. Fomin *Elements of the Theory of Functions and Functional Analysis.* Dover Books on Mathematics 1999

E.H. Lieb and M. Loss *Analysis.* Second edition, AMS 2001

Contents

Comments and Acknowledgements

These notes contain considerably more material than will be lectured or is contained in the schedules. The Appendices contain background material which may have been covered in the pre-requisite courses Linear Analysis and Measure Theory, but do not directly form part of the course. Some of the material on Distributions and the Fourier Transform formed part of a longer course I lectured on just these topics, and so goes into more detail than we will cover in this course. However, I have not drastically trimmed the notes, as some of the material not covered in lectures may nevertheless be interesting to students. There are many bonus exercises in these notes, marked by $(*)$, the example sheet exercises have their own numbers.

The textbooks I have found useful in preparing these notes include

- "Real Analysis", Stein and Shakarchi;
- "Measure Theory and Integration", Taylor;
- "Distributions, Theory and Applications", Duistermaat and Kolk;
- "Introduction to the theory of distributions", Friedlander and Joshi;
- "Fourier Analysis", Körner;
- "Functional Analysis", Rudin;
- "Topology", Munkres.

Please pass any corrections to `c.m.warnick@maths.cam.ac.uk`.

Chapter 1

Lebesgue Integration Theory

1.1 Introduction

This course can be thought of as "putting the *function* into functional analysis"[1]. The course *Linear Analysis* builds on earlier material in the Tripos to describe how a vector space structure (the linear part) interacts with a topological structure (the analysis part). This leads to a very beautiful abstract theory of Banach and Hilbert spaces, as well as other more general topological vector spaces. In this course, we shall see how many of these abstract results relate to more concrete spaces, in particular spaces of functions.

You are (hopefully) familiar from Part IB with the space $C^0([a,b])$ of continuous functions $f : [a,b] \to \mathbb{C}$, equipped with the norm:

$$\|f\|_{C^0} = \sup_{a \leqslant x \leqslant b} |f(x)| .$$

The completeness of this space follows from standard results concerning the uniform convergence of sequences of uniformly continuous functions, hence this is a Banach space. This space and its generalisations are important in many applications (for example in the proof of the Picard-Lindelöf Theorem, and the Schauder Theory for elliptic PDE).

Other spaces of functions naturally arise in many settings. For example, when studying Fourier series defined on $[a,b]$, it is natural to consider the space of continuous functions $f : [a,b] \to \mathbb{C}$ equipped with the norm:

$$\|f\|_{L^2} = \left(\int_a^b |f(x)|^2 \, dx \right)^{\frac{1}{2}} .$$

This norm comes from an inner product in a natural way. This space is not complete: we can construct a sequence of continuous functions f_i such that $(f_i)_{i \in \mathbb{N}}$ is Cauchy with respect to the L^2 norm, but for which there is no continuous function f such that $f_i \to f$ in L^2. One might hope to fix this by considering the space $\mathscr{R}([a,b])$ of Riemann integrable functions, however, we encounter two issues. Firstly, there are non-zero $f \in \mathscr{R}([a,b])$ such that $\|f\|_{L^2} = 0$, so $\|\cdot\|_{L^2}$ ceases to be a norm. In order to avoid this we can work instead

[1] More ambitiously, one could attempt to "put the *fun* into putting the *function* into functional analysis", but we do not aim so high.

with the space $R([a,b]) = \mathscr{R}([a,b])/\sim$ where we quotient by the equivalence relation $f \sim g$ if $\|f-g\|_{L^2} = 0$. The second issue is more serious: even working with the space $R([a,b])$, we do not find completeness with respect to the L^2 norm.

Exercise(∗). a) Find a sequence of continuous functions $f_i : [a,b] \to \mathbb{C}$ such that $(f_i)_{i\in\mathbb{N}}$ is Cauchy with respect to the L^2 norm, but for which there is no continuous function $f : [a,b] \to \mathbb{C}$ such that $f_i \to f$ in L^2.

b) Find a sequence $f_i \in R([a,b])$ such that $(f_i)_{i\in\mathbb{N}}$ is Cauchy with respect to the L^2 norm, but for which there is no $f \in R([a,b])$ such that $f_i \to f$ in L^2.

The solution to this problem is hopefully familiar to you. We should abandon the Riemann integral and work instead with the Lebesgue integral. This brings in our second pre-requisite course, *Probability and Measure*. The construction of the theory of measures and the Lebesgue integral is considerably more involved than that of the Riemann integral, however the pay-off is that the resulting theory of integration is much more powerful. In this course, we shall briefly review the theory of Lebesgue integration that you should have learned last term, before moving on to make use of measure theory, in combination with functional analysis, to understand various function spaces with importance in many branches of analysis.

1.2 Spaces of differentiable functions

Before reviewing integration, we briefly state some facts about the spaces of smooth functions. Let $\Omega \subset \mathbb{R}^n$ be an open set. We denote by $C^k(\Omega)$ the space of all k-times continuously differentiable complex valued functions on Ω, and by

$$C^\infty(\Omega) = \bigcap_{k=0}^{\infty} C^k(\Omega),$$

the set of smooth functions on Ω.

When dealing with partial derivatives of high orders, the notation can get rather messy. To mitigate this, it's convenient to introduce *multi-indices*. We define a multi-index α to be an element of $(\mathbb{Z}_{\geqslant 0})^n$, i.e. a $n-$vector of non-negative integers $\alpha = (\alpha_1, \ldots, \alpha_n)$. We define $|\alpha| = \alpha_1 + \ldots + \alpha_n$ and

$$\frac{\partial^{|\alpha|} f}{\partial x^\alpha} = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \left(\frac{\partial}{\partial x_2}\right)^{\alpha_2} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} f,$$

in other words, we differentiate α_1 times with respect to x_1, α_2 times with respect to x_2 and so on. When it's unambiguous on which variables the derivative acts, we will also use the more compact notation:

$$D_i := \frac{\partial}{\partial x_i},$$

and

$$D^\alpha := \frac{\partial^{|\alpha|}}{\partial x^\alpha}.$$

For a vector $x \in \mathbb{R}^n$, we will also use the notation:

$$x^\alpha := (x_1)^{\alpha_1}(x_2)^{\alpha_2}\cdots(x_n)^{\alpha_n},$$

finally, we define

$$\alpha! = \alpha_1!\alpha_2!\cdots\alpha_n!.$$

The spaces $C^k(\Omega)$ and $C^\infty(\Omega)$ are vector spaces over $\mathbb{C}$, where addition and scalar multiplication are defined pointwise. If $\phi_1, \phi_2 \in C^k(\Omega)$ and $\lambda \in \mathbb{C}$, we define the maps $\phi_1 + \phi_2$, $\lambda\phi_1$ by

$$\begin{aligned} \phi_1 + \phi_2 \ &: \ \Omega \to \mathbb{C}, & \lambda\phi_1 \ &: \ \Omega \to \mathbb{C}, \\ & \quad x \mapsto \phi_1(x) + \phi_2(x), & & \quad x \mapsto \lambda\phi_1(x). \end{aligned} \tag{1.1}$$

> **Exercise($*$).** Show that with the definitions (1.1) the space $C^k(\Omega)$ is a vector space over $\mathbb{C}$, and that $C^l(\Omega)$ is a vector subspace of $C^k(\Omega)$ provided $k \leqslant l \leqslant \infty$.

Definition 1.1. *If $\phi \in C^0(\Omega)$, the* support *of ϕ is the set:*

$$\operatorname{supp}\phi = \overline{\{x \in \Omega : \phi(x) \neq 0\}},$$

where the closure is understood to be relative[2] *to Ω. That is $\operatorname{supp}\phi$ is the closure of the set on which ϕ is not zero. We say that ϕ has compact support if $\operatorname{supp}\phi$ is compact.*

For $0 \leqslant k \leqslant \infty$, we define $C^k_c(\Omega)$ to be the subset of $C^k(\Omega)$ consisting of functions with compact support. $C^k_c(\Omega)$ is a vector subspace of $C^k(\Omega)$.

Theorem 1.1. *There exists a function $\psi \in C^\infty_c(\mathbb{R}^n)$ such that*

i) $\psi \geqslant 0$

ii) $\psi(0) \neq 0$

iii) $\operatorname{supp}\psi \subset B_1(0) := \{x \in \mathbb{R}^n : |x| < 1\}$

iv) We have:

$$\int_{\mathbb{R}^n} \psi(x)dx = 1.$$

Proof. First, we note that the function:

$$\chi(t) = \begin{cases} 0 & t \leqslant 0 \\ e^{-\frac{1}{t}} & t > 0 \end{cases}$$

is smooth, i.e. $\chi \in C^\infty(\mathbb{R})$. Moreover, $\chi \geqslant 0$ and $\chi(1) \neq 0$. We define $\psi_0(x) = \chi\left(1 - 2|x|^2\right)$. Since the map $x \mapsto |x|^2$ is smooth, $\psi_0 \in C^\infty(\mathbb{R}^n)$. We set:

$$\psi(x) = \frac{\psi_0(x)}{\int_{\mathbb{R}^n}\psi_0(x)dx}.$$

It is easy to verify that ψ satisfies conditions $i) - iv)$. □

[2] If $\Omega \subset \mathbb{R}^n$ is open, and $A \subset \Omega$, then the closure of A relative to Ω is the intersection of Ω with the closure of A as a subset of $\mathbb{R}^n$. Note that the closure of A relative to Ω may not be closed as a subset of $\mathbb{R}^n$.

Corollary 1.2. *For $\Omega \subset \mathbb{R}^n$ open, $C_c^\infty(\Omega)$ is not the trivial subspace $\{0\} \subset C^\infty(\Omega)$.*

Proof. Since Ω is open, there exists ϵ, x such that the ball $B_\epsilon(x) = \{y \in \mathbb{R}^n : |y - x| < \epsilon\}$ is contained in Ω. The function $y \mapsto \psi\left[\epsilon^{-1}(y - x)\right]$ is easily seen to belong to $C_c^\infty(\Omega)$. □

> **Exercise(∗).** Construct explicitly a function $\psi \in C_c^\infty(\mathbb{R}^n)$ such that
>
> i) $0 \leqslant \psi \leqslant 1$
>
> ii) $\operatorname{supp} \psi \subset B_2(0)$
>
> iii) $\psi(x) = 1$ for $|x| \leqslant 1$.

Suppose $\Omega \subset \Omega'$, where both are open subsets of $\mathbb{R}^n$. If $\phi \in C_c^k(\Omega)$, then we can extend ϕ to a function on Ω' by setting $\phi = 0$ on $\Omega' \setminus \Omega$. This extended function will be smooth in Ω' and we do not alter the support, so in this way we see that $C_c^k(\Omega)$ is a vector subspace of $C_c^k(\Omega')$.

The following result is useful:

Lemma 1.3. *Suppose $\Omega \subset \mathbb{R}^n$ is open and $K \subset \Omega$ is compact. Then $d(K, \partial\Omega) > 0$, where:*

$$d(K, \partial\Omega) := \inf_{x \in K, y \in \partial\Omega} |x - y| .$$

Proof. K is compact, so $K \subset B_R(0)$ for some $R > 0$. Let $\Omega_R = \Omega \cap B_R(0)$. Ω_R is open and bounded, with $K \subset \Omega_R$. It suffices to show that $d(K, \partial\Omega_R) > 0$. Since Ω_R is bounded, $\partial\Omega_R$ is compact. Therefore the map:

$$\begin{aligned} f \quad : \quad & K \times \partial\Omega_R \to \mathbb{R}_{\geq 0}, \\ & (x, y) \mapsto |x - y| , \end{aligned}$$

is a continuous map on a compact set, hence it achieves its minimum d at (x_0, y_0). Suppose that $d = 0$, then $x_0 = y_0$, but $x_0 \in K \subset \Omega_R$ and $y_0 \in \partial\Omega_R \subset \Omega_R^c$ a contradiction. Thus $d > 0$ and we're done. □

Corollary 1.4. *If $\phi \in C_c^k(\Omega)$, extend ϕ to $\mathbb{R}^n$ by $\phi = 0$ on Ω^c. Define $\tau_x\phi$ by:*

$$\begin{aligned} \tau_x\phi \quad : \quad & \Omega \to \mathbb{C}, \\ & y \mapsto \phi(y - x). \end{aligned} \tag{1.2}$$

Then there exists $\epsilon > 0$ such that $\tau_x\phi \in C_c^k(\Omega)$ for all $x \in B_\epsilon(0)$.

Proof. We have

$$\operatorname{supp} \tau_x\phi = \operatorname{supp} \phi + x$$

Since $\operatorname{supp} \phi$ is compact, $\operatorname{supp} \tau_x\phi$ is just a translate of a compact set, so is compact as a subset of $\mathbb{R}^n$. We need to check that $\operatorname{supp} \tau_x\phi \subset \Omega$. We have $d(\operatorname{supp} \phi, \partial\Omega) = \delta > 0$. Set $\epsilon = \delta/2$. Then we have, by Lemma 1.3

$$\operatorname{supp} \phi + B_\epsilon(0) \subset \Omega$$

but if $x \in B_\epsilon(0)$, then $\operatorname{supp} \tau_x\phi \subset \operatorname{supp} \phi + B_\epsilon(0)$ and we're done. □

1.3 Review of integration

In this section we will briefly recall the main definitions and theorems of the theory of measure and Lebesgue integration. We shall focus on the Lebesgue measure on $\mathbb{R}^n$. Appendix B gives a much fuller account of the theory, and in particular contains proofs of the results claimed below.

We start with the basic definition of a measure space.

Definition 1.2. *Given a set E, a collection of subsets $\mathcal{E}$ of E is called a σ-algebra if:*

i) $\emptyset \in \mathcal{E}$

ii) $A \in \mathcal{E}$ *implies* $A^c = \{x \in E : x \notin A\} \in \mathcal{E}$

iii) $A_n \in \mathcal{E}$ *for* $n \in \mathbb{N}$ *implies* $\cup_n A_n \in \mathcal{E}$.

The pair $(E, \mathcal{E})$ *is called a* measurable space *and elements of* $\mathcal{E}$ *are called* measurable sets. *A* measure *on* $(E, \mathcal{E})$ *is a function* $\mu : \mathcal{E} \to [0, \infty]$ *such that:*

i) $\mu(\emptyset) = 0$

ii) If $A_n \in \mathcal{E}$ *for* $n \in \mathbb{N}$ *are disjoint, then* $\mu(\cup_n A_n) = \sum_n \mu(A_n)$.

A triple $(E, \mathcal{E}, \mu)$ *is called a measure space.*

A simple example of a measure space is given by taking $\mathcal{E} = 2^E$ and $\mu(A) = \#A$. This is the counting measure. Given any collection $\mathcal{A}$ of subsets of E, we can define $\sigma(\mathcal{A})$, the σ-algebra generated by $\mathcal{A}$ to be the intersection of all σ-algebras containing $\mathcal{A}$. If E is a topological space, the *Borel* algebra is the σ-algebra generated by the open sets of E, written $\mathcal{B}(E)$.

A particular case of interest is $E = \mathbb{R}^n$, on which we can define a σ-algebra $\mathcal{M}$, and measure λ with the following properties:

i) $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{M}$

ii) If A is a rectangle, i.e. $A = (a_1, b_1] \times \cdots \times (a_n, b_n]$, then $\lambda(A) = (b_1 - a_1) \cdots (b_n - a_n)$.

iii) $A \in \mathcal{M}$ if and only for any $\epsilon > 0$ there exists an open set O and a closed set C such that $C \subset A \subset O$ and
$$\lambda(O \setminus C) < \epsilon.$$

Since any open set in $\mathbb{R}^n$ is the countable union of disjoint rectangles these conditions determine $\mathcal{M}$, λ uniquely. We note that if $\lambda(A) < \infty$, then the set C in *iii*) above may be assumed to be compact. Property *iii*) is sometimes referred to as Borel regularity. We call $\mathcal{M}$ the σ-algebra of Lebesgue measurable sets, and λ is the Lebesgue measure. For the Lebesgue measure, we often denote $\lambda(A)$ by $|A|$, and μ by dx.

Definition 1.3. *A function* $f : E \to G$ *which maps between two measurable spaces* $(E, \mathcal{E})$, $(G, \mathcal{G})$ *is* measurable *if* $f^{-1}(A) \in \mathcal{E}$ *for all* $A \in \mathcal{G}$. *Special cases include:*

a) *If* $(G, \mathcal{G}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*, we simply say* f *is a measurable function on* $(E, \mathcal{E})$*.*

b) *If* $(G, \mathcal{G}) = ([0, \infty], \mathcal{B}([0, \infty]))$ *we say* f *is a non-negative measurable function on* $(E, \mathcal{E})$*.*

c) *If* E, G *are topological spaces with their Borel algebras then we say* f *is a Borel function on* E*.*

The class of measurable functions is closed under vector space operations, products and limits.

A *simple function* is a function of the form

$$f = \sum_{k=1}^{N} a_k \mathbb{1}_{A_k}$$

for $A_k \in \mathcal{E}$ and a_k constant (typically in $[0, \infty]$, $\mathbb{R}$ or $\mathbb{C}$). All simple functions are measurable. For a non-negative simple function we define the integral

$$\mu(f) = \int_E f d\mu := \sum_{k=1}^{N} a_k \mu(A_k),$$

where $0 \cdot \infty = 0$ by convention. For a non-negative measurable function we define

$$\mu(f) = \int_E f d\mu := \sup \{\mu(g) : \ g \text{ simple and } 0 \leqslant g \leqslant f\}.$$

A measurable function is *integrable* if $\mu(|f|) < \infty$, in which case we can write $f = f^+ - f^-$ with $f^\pm$ non-negative and $\mu(f^\pm) < \infty$. Then

$$\mu(f) = \int_E f d\mu := \mu(f^+) - \mu(f^-).$$

The integral satisfies all the usual basic properties (linearity, additivity etc.), and agrees with the Riemann integral when both are defined. We can also state two important theorems for interchanging limits and integrals.

Theorem 1.5 (Monotone convergence)**.** *Let* $(f_n)_{n=1}^{\infty}$ *be an increasing sequence of non-negative measurable functions on a measure space* $(E, \mathcal{E}, \mu)$ *which converge to* f*. Then*

$$\lim_{n\to\infty} \int_E f_n d\mu = \int_E f d\mu$$

Theorem 1.6 (Dominated convergence)**.** *Let* $(f_n)_{n=1}^{\infty}$ *be a sequence of measurable functions on a measure space* $(E, \mathcal{E}, \mu)$ *such that*

i) $f_n \to f$ *pointwise ae.*

ii) $|f_n| \leqslant g$ *ae for some integrable* g*.*

Then:

$$\lim_{n\to\infty}\int_E f_n d\mu = \int_E f d\mu$$

Associated to each measure space $(E, \mathcal{E}, \mu)$ are scale of Banach spaces.

Definition 1.4. *For $1 \leqslant p < \infty$, and $f : E \to \mathbb{C}$ measurable, we define:*

$$\|f\|_{L^p} = \left(\int_E |f|^p \, d\mu\right)^{\frac{1}{p}}.$$

while for $p = \infty$ we set

$$\|f\|_{L^\infty} = \operatorname*{ess\,sup}_E |f| = \inf\{C : |f| \leqslant C \ \ ae\}.$$

The space $L^p(E, \mu)$ is then defined to be

$$L^p(E, \mu) = \{f : E \to \mathbb{C} \text{ measurable } : \|f\|_{L^p} < \infty\} / \sim$$

where we quotient by the equivalence relation $f \sim g$ if $f = g$ ae.

If E is a topological space and $\mathcal{B}(E) \subset \mathcal{E}$, we define $L^p_{loc.}(E, \mu)$ to consist of measurable functions (modulo $\sim$) such that $f\mathbb{1}_K \in L^p(E, \mu)$ for all compact K.

When our measure space is $(\mathbb{R}^n, \mathcal{M}, \lambda)$ we will often write $L^p(\mathbb{R}^n) := L^p(\mathbb{R}^n, \lambda)$.

Theorem 1.7. *The space $L^p(E, \mu)$ equipped with the norm $\|\cdot\|_{L^p}$ is a Banach space for $1 \leqslant p \leqslant \infty$.*

It is useful also to note that the set S of complex valued simple functions on E such that

$$\mu(\{x : s(x) \neq 0\}) < \infty$$

is dense in $L^p(E, \mu)$ for $1 \leqslant p < \infty$.

Exercise 1.1. Suppose $f, g : E \to \mathbb{C}$ are measurable functions on some measure space $(E, \mathcal{E}, \mu)$. Show that:

a) $\|fg\|_{L^r} \leqslant \|f\|_{L^p} \|g\|_{L^q}$ where $1 \leqslant p, q, r \leqslant \infty$ satisfy $p^{-1} + q^{-1} = r^{-1}$
[You may wish to first establish the special case $r = 1$.]

b) $\|f + g\|_{L^p} \leqslant \|f\|_{L^p} + \|g\|_{L^p}$ for $1 \leqslant p \leqslant \infty$.

Exercise 1.2. a) Suppose that $\mu(E) < \infty$. Show that if $f \in L^p(E, \mu)$, then $f \in L^q(E, \mu)$ for any $1 \leqslant q \leqslant p$, with

$$\|f\|_{L^q} \leqslant \mu(E)^{\frac{p-q}{qp}} \|f\|_{L^p}.$$

b) Suppose that $f \in L^{p_0}(E,\mu) \cap L^{p_1}(E,\mu)$ with $p_0 < p_1 \leqslant \infty$. For $0 \leqslant \theta \leqslant 1$, define p_θ by
$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}.$$
Show that $f \in L^{p_\theta}(E,\mu)$ with
$$\|f\|_{L^{p_\theta}} \leqslant \|f\|_{L^{p_0}}^{1-\theta} \|f\|_{L^{p_1}}^{\theta}.$$

c) Show that for $p_1 \neq p_2$ we have $L^{p_1}(\mathbb{R}^n) \not\subset L^{p_2}(\mathbb{R}^n)$. For which p_1, p_2 do we have $L^{p_1}_{loc.}(\mathbb{R}^n) \subset L^{p_2}_{loc.}(\mathbb{R}^n)$?

Exercise 1.3. Let $\mathcal{R}_\mathbb{Q}$ be the set of rectangles of the form $(a_1, b_1] \times \cdots \times (a_n, b_n]$ with $a_i, b_i \in \mathbb{Q}$, and let $S_\mathbb{Q}$ be the set of functions of the form
$$s(x) = \sum_{k=1}^{N} (\alpha_k + i\beta_k) \mathbb{1}_{R_k}$$
for $R_k \in \mathcal{R}_\mathbb{Q}$ and $\alpha_k, \beta_k \in \mathbb{Q}$. For $1 \leqslant p < \infty$ show that $S_\mathbb{Q}$ is dense in $L^p(\mathbb{R}^n)$ and deduce that $L^p(\mathbb{R}^n)$ is separable. Show that $L^\infty(\mathbb{R}^n)$ is not separable. *[Hint: for the last part exhibit an uncountable subset $X \subset L^\infty(\mathbb{R}^n)$ such that $\|f-g\|_{L^\infty(\mathbb{R}^n)} \geqslant 1$ for any $f, g \in X$, $f \neq g$]*.

1.4 Convolution and mollification

In this section, we are going to establish some results concerning mollification of functions in $L^p(\mathbb{R}^n)$. The final result will be to establish the density of $C_c^\infty(\mathbb{R}^n)$ in $L^p(\mathbb{R}^n)$ for $1 \leqslant p < \infty$. We first establish an important fact about the spaces $L^p(\mathbb{R}^n)$: namely that the translation operator is continuous on these spaces. More concretely, for any $z \in \mathbb{R}^n$ we set $\tau_z f(x) = f(x-z)$. We then show:

Lemma 1.8. *Suppose $p \in [1,\infty)$ and $g \in L^p(\mathbb{R}^n)$. Let $\{z_j\}_{j=1}^\infty \subset \mathbb{R}^n$ be a sequence of points such that $z_j \to 0$ as $j \to \infty$. Then:*
$$\left\|\tau_{z_j} g - g\right\|_{L^p(\mathbb{R}^n)} \to 0.$$

Proof. 1. First, suppose $g = \mathbb{1}_R$, where $R = (a_1, b_1] \times (a_2, b_2] \times \ldots \times (a_n, b_n]$ is a rectangle, with side-lengths $I_m = b_m - a_m$ for $m = 1, \ldots, n$. Now, since when a box is translated by a vector z_j each side is translated by a distance of at most $|z_j|$, and has area at most $I_{max.}^{n-1}$, where I_{max} is the longest side-length we can crudely estimate
$$\left\|\tau_{z_j} g - g\right\|_{L^p(\mathbb{R}^n)}^p \leqslant 2n\,|z_j|\,I_{\max}^{n-1}.$$
Note that this estimate requires $p < \infty$: it does not hold for $p = \infty$. We conclude that:
$$\lim_{j\to\infty} \left\|\tau_{z_j} g - g\right\|_{L^p(\mathbb{R}^n)} = 0.$$

2. Now suppose $g = \mathbb{1}_A$, where A is a measurable set of finite measure. Fix $\epsilon > 0$. By the Borel regularity of Lebesgue measure, there exists a compact $K \subset A$ and an open $U \supset A$ such that $|U \setminus K| < \epsilon^p$. Since U is open, we can write U as a union of open rectangles:
$$U = \bigcup_{\alpha \in \mathscr{A}} R_\alpha^\circ$$
Since K is compact, it is covered by a finite subset of these:
$$K \subset \bigcup_{i=1}^N R_i := B.$$
Now, note that $K \subset B \subset U$, so the symmetric difference $A \Delta B \subset U \setminus K$. Thus[3] $\|\mathbb{1}_A - \mathbb{1}_B\|_{L^p(\mathbb{R}^n)} = |A\Delta B|^{1/p} < \epsilon$. By the paragraph 1 above, we know that there exists J such that for all $j \geqslant J$ we have:
$$\left\|\tau_{z_j}\mathbb{1}_B - \mathbb{1}_B\right\|_{L^p(\mathbb{R}^n)} < \epsilon$$
Therefore:
$$\begin{aligned}\left\|\tau_{z_j}\mathbb{1}_A - \mathbb{1}_A\right\|_{L^p(\mathbb{R}^n)} &= \left\|\tau_{z_j}\mathbb{1}_A - \tau_{z_j}\mathbb{1}_B + \tau_{z_j}\mathbb{1}_B - \mathbb{1}_B + \mathbb{1}_B - \mathbb{1}_A\right\|_{L^p(\mathbb{R}^n)} \\ &\leqslant \left\|\tau_{z_j}\mathbb{1}_A - \tau_{z_j}\mathbb{1}_B\right\|_{L^p(\mathbb{R}^n)} + \left\|\tau_{z_j}\mathbb{1}_B - \mathbb{1}_B\right\|_{L^p(\mathbb{R}^n)} + \|\mathbb{1}_B - \mathbb{1}_A\|_{L^p(\mathbb{R}^n)} \\ &= 2\|\mathbb{1}_A - \mathbb{1}_B\|_{L^p(\mathbb{R}^n)} + \left\|\tau_{z_j}\mathbb{1}_B - \mathbb{1}_B\right\|_{L^p(\mathbb{R}^n)} \\ &< 3\epsilon\end{aligned}$$
for all $j \geqslant J$. Thus
$$\lim_{j\to\infty} \left\|\tau_{z_j} g - g\right\|_{L^p(\mathbb{R}^n)} = 0.$$

3. Now suppose g is a simple function, $g = \sum_{i=1}^N g_i \mathbb{1}_{A_i}$, where $g_i \in \mathbb{C}$ and A_i are measurable sets of *finite* measure. Then we have:
$$\left\|\tau_{z_j} g - g\right\|_{L^p(\mathbb{R}^n)} \leqslant \sum_{i=1}^N |g_i| \left\|\tau_{z_j}\mathbb{1}_{A_i} - \mathbb{1}_{A_i}\right\|_{L^p(\mathbb{R}^n)}$$
so as $j \to \infty$ we have:
$$\lim_{j\to\infty} \left\|\tau_{z_j} g - g\right\|_{L^p(\mathbb{R}^n)} = 0.$$

4. Now suppose that $g \in L^p(\mathbb{R}^n)$. Fix $\epsilon > 0$. Recall that there exists a simple function $\tilde{g} = \sum_{i=1}^n \tilde{g}_i \mathbb{1}_{A_i}$ with $\tilde{g}_i \in \mathbb{C}$, $|A_i| < \infty$ such that $\|g - \tilde{g}\|_{L^p(\mathbb{R}^n)} < \epsilon$. By the previous part, we can find J such that $\left\|\tau_{z_j}\tilde{g} - \tilde{g}\right\|_{L^p(\mathbb{R}^n)} < \epsilon$ for all $j \geqslant J$. Now:
$$\begin{aligned}\left\|\tau_{z_j} g - g\right\|_{L^p(\mathbb{R}^n)} &= \left\|\tau_{z_j} g - \tau_{z_j}\tilde{g} + \tau_{z_j}\tilde{g} - \tilde{g} + \tilde{g} - g\right\|_{L^p(\mathbb{R}^n)} \\ &\leqslant \left\|\tau_{z_j} g - \tau_{z_j}\tilde{g}\right\|_{L^p(\mathbb{R}^n)} + \left\|\tau_{z_j}\tilde{g} - \tilde{g}\right\|_{L^p(\mathbb{R}^n)} + \|\tilde{g} - g\|_{L^p(\mathbb{R}^n)} \\ &= 2\|g - \tilde{g}\|_{L^p(\mathbb{R}^n)} + \left\|\tau_{z_j}\tilde{g} - g\right\|_{L^p(\mathbb{R}^n)} \\ &< 3\epsilon\end{aligned}$$

[3] This is another point at which $p \neq \infty$ is crucial.

Thus, we conclude that

$$\lim_{j\to\infty} \left\|\tau_{z_j} g - g\right\|_{L^p(\mathbb{R}^n)} = 0.$$

and we're done. □

If f, g are functions mapping $\mathbb{R}^n$ to $\mathbb{C}$, then we define the convolution of f and g to be:

$$(f \star g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)dy,$$

provided the integral exists. This will happen if (for example) $f \in L^1(\mathbb{R}^n)$ and $g \in L^\infty(\mathbb{R}_n)$.

Lemma 1.9. *Suppose* $f, g, h \in C_c^\infty(\mathbb{R}^n)$. *Then:*

$$f \star g = g \star f, \qquad f \star (g \star h) = (f \star g) \star h.$$

and

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

Proof. With the change of variables $y = x - z$, we have[4]

$$(f \star g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)dy = \int_{\mathbb{R}^n} f(x-z)g(z)dz = (g \star f)(x)$$

Next, we calculate:

$$\begin{aligned}
[f \star (g \star h)](x) &= \int_{\mathbb{R}^n} f(y) \left(\int_{\mathbb{R}^n} g(z)h(x-y-z)dz \right) dy \\
&= \int_{\mathbb{R}^n} f(y) \left(\int_{\mathbb{R}^n} g(w-y)h(x-w)dw \right) dy \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(y)g(w-y)dy \right) h(x-w)dw \\
&= [(f \star g) \star h](x)
\end{aligned}$$

Above we have made the substitution $w = y + z$ to pass from the first to second line, and we have used the fact that $f, g, h \in C_c^\infty(\mathbb{R}^n)$ to invoke Fubini's theorem (Theorem B.30)

[4]If you're worried about a missing minus sign from the change of variables when n is odd, observe:

$$\int_{-\infty}^{\infty} k(x)dx = \int_{\infty}^{-\infty} k(-y)d(-y) = -\int_{\infty}^{-\infty} k(-y)dy = \int_{-\infty}^{\infty} k(-y)dy.$$

when passing from the second to third line. Finally, we calculate:

$$\begin{aligned}\int_{\mathbb{R}^n}(f\star g)(x)dx &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} f(y)g(x-y)dy\right)dx\\ &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} f(y)g(x-y)dx\right)dy\\ &= \int_{\mathbb{R}^n}\left(f(y)\int_{\mathbb{R}^n} g(z)dz\right)dy\\ &= \int_{\mathbb{R}^n} f(x)dx\int_{\mathbb{R}^n} g(z)dy.\end{aligned}$$

where again, the fact that $f, g \in C_c^\infty(\mathbb{R}^n)$ allows us to invoke Fubini. □

The assumption that the functions are smooth and compactly supported is certainly overkill in this theorem. It would be enough, for example, to consider functions in $C_c^0(\mathbb{R}^n)$, or even weaker spaces, provided we can justify the application of Fubini's theorem.

Exercise($*$). Suppose that $f, g, h \in C_c^\infty(\mathbb{R}^n)$.

a) Show that for any multi-index α, we have that $D^\alpha f \in L^p(\mathbb{R}^n)$ for $1 \leqslant p < \infty$, i.e. that

$$\|D^\alpha f\|_{L^p(\mathbb{R}^n)} = \left(\int_{\mathbb{R}^n} |D^\alpha f(x)|^p \, dx\right)^{\frac{1}{p}} < \infty.$$

b) Define

$$\begin{aligned} F \quad : \quad & \mathbb{R}^n \times \mathbb{R}^n, \\ & (x, y) \mapsto f(x)g(y-x). \end{aligned}$$

Show that $F \in L^1(\mathbb{R}^n \times \mathbb{R}^n)$.

c) For each $x \in \mathbb{R}^n$, set

$$\begin{aligned} G_x \quad : \quad & \mathbb{R}^n \times \mathbb{R}^n, \\ & (y, z) \mapsto f(y)g(z)h(x-y-z). \end{aligned}$$

Show that $G_x \in L^1(\mathbb{R}^n \times \mathbb{R}^n)$.

1.4.1 Differentiating convolutions

A remarkable property of the convolution is that the regularity of $f \star g$ is determined by the regularity of the *smoother* of f and g. This is a result of the following Lemma:

Theorem 1.10. *Suppose $f \in L^1_{loc.}(\mathbb{R}^n)$ and $g \in C_c^k(\mathbb{R}^n)$ for some $k \geqslant 0$. Then $f \star g \in C^k(\mathbb{R}^n)$ and*

$$D^\alpha(f \star g) = f \star D^\alpha g,$$

for any multiindex with $|\alpha| \leqslant k$.

Before we prove this, it's convenient to prove a technical Lemma which will streamline the proof. We introduce the difference quotient

$$\Delta_i^h f(x) = \frac{f(x+he_i)-f(x)}{h}.$$

Lemma 1.11. *a) Suppose $f \in C_c^0(\mathbb{R}^n)$ and $\{z_i\}_{i=1}^\infty \subset \mathbb{R}^n$ is a sequence with $z_i \to 0$ as $i \to \infty$. Then for any $x \in \mathbb{R}^n$:*

i) $\tau_{z_j} f(x) \to f(x)$ as $j \to \infty$.

ii) $\left|\tau_{z_j} f(x)\right| \leqslant (\sup_{\mathbb{R}^n} |f|)\, \mathbb{1}_{B_R(0)}(x)$, for some $R > 0$ and all j.

b) Suppose $f \in C_c^1(\mathbb{R}^n)$ and $\{h_j\}_{j=1}^\infty \subset \mathbb{R}$ is a sequence with $h_j \to 0$ as $j \to \infty$. Then for any $x \in \mathbb{R}^n$:

i) $\Delta_i^{h_j} f(x) \to D_i f(x)$ as $j \to \infty$.

ii) $\left|\Delta_i^{h_j} f(x)\right| \leqslant (\sup_{\mathbb{R}^n} |D_i f|)\, \mathbb{1}_{B_R(0)}(x)$, for some $R > 0$ and all j.

Proof. a) i) Recall $\tau_{z_j} f(x) = f(x - z_j)$. Clearly since $z_j \to 0$, $f(x-z_j) \to f(x)$ as $j \to \infty$ by the continuity of f.

ii) Since $z_j \to 0$, there exists some $\rho > 0$ such that $z_j \in \overline{B_\rho(0)}$ for all j. Now

$$\operatorname{supp} \tau_{z_j} f = \operatorname{supp} f + z_j \subset \operatorname{supp} f + \overline{B_\rho(0)}.$$

Since the sum of two bounded set is bounded, we conclude that there exists $R > 0$ such that $\operatorname{supp} \tau_{z_j} f \subset B_R(0)$. Thus $\tau_{z_j} f = \tau_{z_j} f \mathbb{1}_{B_R(0)}$ and we estimate:

$$\left|\tau_{z_j} f(x)\right| = \left|\tau_{z_j} f(x)\right| \mathbb{1}_{B_R(0)}(x) \leqslant \sup_{\mathbb{R}^n} |f|\, \mathbb{1}_{B_R(0)}(x).$$

b) Suppose $f \in C_c^1(\mathbb{R}^n)$ and $\{h_j\}_{j=1}^\infty \subset \mathbb{R}$ is a sequence with $h_j \to 0$ as $j \to \infty$. Then for any $x \in \mathbb{R}^n$:

i) From the definition of the difference quotient and of the partial derivative:

$$\Delta_i^{h_j} f(x) = \frac{f(x+h_j e_i)-f(x)}{h} \to D_i f(x), \qquad \text{as } j \to \infty.$$

ii) Since $h_j \to 0$, there is some $\rho > 0$ such that $|h_j| \leqslant \rho$ for all j. We have:

$$\begin{aligned} \operatorname{supp} \Delta_i^{h_j} f &\subset \operatorname{supp} \tau_{-h_j e_i} f \cup \operatorname{supp} f = (\operatorname{supp} f - h_j e_i) \cup \operatorname{supp} f \\ &\subset \left(\operatorname{supp} f + \overline{B_\rho(0)}\right) \cup \operatorname{supp} f \\ &\subset B_R(0) \end{aligned}$$

for some $R > 0$ since the union of two bounded sets is bounded. Thus $\Delta_i^{h_j} f = \Delta_i^{h_j} f \mathbb{1}_{B_R(0)}$. We also observe that by the mean value theorem, for any $h \in \mathbb{R}$, there exists $s \in \mathbb{R}$ with $|s| < |h|$ such that

$$\frac{f(x+he_i)-f(x)}{h} = D_i f(x+se_i)$$

thus

$$\left|\Delta_i^{h_j} f(x)\right| \leqslant \sup_{\mathbb{R}^n} |D_i f| \,.$$

Putting these two facts together, we readily find:

$$\left|\Delta_i^{h_j} f(x)\right| = \left|\Delta_i^{h_j} f(x)\right| \mathbb{1}_{B_R(0)}(x) \leqslant \sup_{\mathbb{R}^n} |D_i f| \, \mathbb{1}_{B_R(0)}(x).$$

□

Now, with this technical result in hand we can attack the proof our original theorem.

Proof of Theorem 1.10. 1. First we establish the result for $k = 0$. We need to show that if $f \in L^1_{loc.}(\mathbb{R}^n)$ and $g \in C^0_c(\mathbb{R}^n)$ then $f \star g$ is continuous. To show this, it suffices to show that $f \star g(x - z_j) \to f \star g(x)$ for any $x \in \mathbb{R}^n$ and any sequence $\{z_j\}_{j=1}^\infty$ with $z_j \to 0$. Now, note that

$$f \star g(x - z_j) = \int_{\mathbb{R}^n} f(y) g(x - z_j - y) dy = \int_{\mathbb{R}^n} f(y) \tau_{z_j} g(x - y) dy.$$

Now, sending $j \to \infty$, we are done, so long as we can justify interchanging the limit and the integral. Note that for any fixed x and all j:

$$\left|f(y)\tau_{z_j} g(x - y)\right| \leqslant \sup_{\mathbb{R}^n} |g| \, \mathbb{1}_{B_R(0)}(x - y) \, |f(y)|$$

for some R by the previous Lemma. Since $f \in L^1_{loc.}(\mathbb{R}^n)$ the right hand side is integrable, and so by the dominated convergence theorem:

$$\lim_{j\to\infty} f \star g(x - z_j) = \int_{\mathbb{R}^n} \lim_{j\to\infty} f(y)\tau_{z_j} g(x - y) dy = \int_{\mathbb{R}^n} f(y) g(x - y) dy = f \star g(x).$$

2. Now suppose that $f \in L^1_{loc.}(\mathbb{R}^n)$ and $g \in C^1_c(\mathbb{R}^n)$. Clearly $f \star D_i g$ is continuous by the previous argument. To show $f \star g \in C^1(\mathbb{R}^n)$, it suffices to show that for any $x \in \mathbb{R}^n$ and any sequence $\{h_j\}_{j=1}^\infty \subset \mathbb{R}$ with $h_j \to 0$ we have:

$$\lim_{j\to\infty} \Delta_i^{h_j} f \star g(x) = f \star D_i g(x).$$

Note that

$$\begin{aligned}
\Delta_i^{h_j} f \star g(x) &= \frac{f \star g(x + h_j e_i) - f \star g(x)}{h} \\
&= \int_{\mathbb{R}^n} f(y) \left(\frac{g(x + h_j e_i - y) - g(x - y)}{h} \right) dy \\
&= \int_{\mathbb{R}^n} f(y) \Delta_i^{h_j} g(x - y) dy
\end{aligned}$$

so that again we are done provided we can send $j \to \infty$ and interchange the limit and the integral. An argument precisely analogous to the previous case allows us to invoke the DCT and deduce that:

$$\lim_{j\to\infty} \Delta_i^{h_j} f \star g(x) = \int_{\mathbb{R}^n} \lim_{j\to\infty} f(y) \Delta_i^{h_j} g(x - y) dy = f \star D_i g(x).$$

3. The case where $g \in C_c^k(\mathbb{R}^n)$ with $k > 1$ now follows by a simple induction. □

Exercise(∗). Show that $f \star g \in C^k(\mathbb{R}^n)$ under the hypotheses:

a) $f \in L^1(\mathbb{R}^n)$, $g \in C^k(\mathbb{R}^n)$ with $\sup_{\mathbb{R}^n} |D^\alpha g| < \infty$ for all $|\alpha| \leqslant k$.

b) $f \in L^1(\mathbb{R}^n)$ with $\operatorname{supp} f$ compact, $g \in C^k(\mathbb{R}^n)$.

We have shown that when two functions are convolved, loosely speaking the resulting function is at least as regular as the *better* of the two original functions. It is also important to know how convolution modifies the support of a function.

Lemma 1.12. *Suppose* $f \in L^1_{loc.}(\mathbb{R}^n)$ *and* $g \in C_c^k(\mathbb{R}^n)$ *for some* $k \geqslant 0$. *Then*[5]

$$\operatorname{supp}(f \star g) \subset \operatorname{supp} f + \operatorname{supp} g.$$

Proof. Recall:

$$f \star g(x) = \int_{\mathbb{R}^n} f(y)g(x-y)dy.$$

Clearly, if $f \star g(x) \neq 0$, then there must exist $y \in \mathbb{R}^n$ such that $y \in \operatorname{supp} f$ and $x - y = z \in \operatorname{supp} g$. Thus $x = y + z$ with $y \in \operatorname{supp} f$ and $z \in \operatorname{supp} g$. This tells us that:

$$\{x \in \mathbb{R}^n : f \star g(x) \neq 0\} \subset \operatorname{supp} f + \operatorname{supp} g.$$

Since $\operatorname{supp} f$ is closed and $\operatorname{supp} g$ is compact, we know that $\operatorname{supp} f + \operatorname{supp} g$ is closed, thus

$$\operatorname{supp} f \star g = \overline{\{x \in \mathbb{R}^n : f \star g(x) \neq 0\}} \subset \operatorname{supp} f + \operatorname{supp} g,$$

which is the result we require. □

Exercise(∗). a) Prove the following identities for $r, s > 0$ and $x \in \mathbb{R}^n$:

i) $B_r(x) + B_s(0) = B_{r+s}(x)$
ii) $\overline{B_r(x)} + B_s(0) = B_{r+s}(x)$
iii) $\overline{B_r(x)} + \overline{B_s(0)} = \overline{B_{r+s}(x)}$

Suppose that $A, B \subset \mathbb{R}^n$. Show that:

b) If one of A or B is open, then so is $A + B$.

c) If A and B are both bounded, then so is $A + B$.

d) If A is closed and B is compact, then $A + B$ is closed.

e) If A and B are both compact, then so is $A + B$.

Exercise(∗). Show that if $f \in C_c^k(\mathbb{R}^n)$ and $g \in C_c^l(\mathbb{R}^n)$ then $f \star g \in C_c^{k+l}(\mathbb{R}^n)$. Conclude that $C_c^\infty(\mathbb{R}^n)$ is closed under convolution.

[5] Strictly speaking, we haven't defined the support of a measurable function. We can do this in several ways, but the simplest is to define:

$$\operatorname{supp} f = \bigcap \{E \subset \mathbb{R}^n : E \text{ is closed, and } f = 0 \text{ a.e. on } E^c\}.$$

In other words $\operatorname{supp} f$ is the smallest closed set such that f vanishes almost everywhere on its complement.

1.4.2 Approximation of the identity

An important use of the convolution is to construct smooth approximations to functions in various function spaces. The following theorem is very useful in constructing approximations:

Theorem 1.13. *Suppose $\phi \in C^\infty_c(\mathbb{R}^n)$ satisfies:*

i) $\phi \geqslant 0$

ii) $\operatorname{supp}\phi \subset B_1(0)$

iii) $\int_{\mathbb{R}^n} \phi(x)dx = 1$

Such a ϕ exists by Theorem 1.1. Define:

$$\phi_\epsilon(y) = \frac{1}{\epsilon^n}\phi\left(\frac{y}{\epsilon}\right).$$

Then:

a) If $f \in C^k_c(\mathbb{R}^n)$, then $\phi_\epsilon \star f$ is smooth, and

$$D^\alpha\left(\phi_\epsilon \star f\right) \to D^\alpha f \qquad \textit{as } \epsilon \to 0,$$

uniformly on $\mathbb{R}^n$ for any multi-index with $|\alpha| \leqslant k$.

b) If $g \in L^p(\mathbb{R}^n)$ with $1 \leqslant p < \infty$, then $\phi_\epsilon \star g$ is smooth, and

$$\phi_\epsilon \star g \to g \qquad \textit{in } L^p(\mathbb{R}^n) \qquad \textit{as } \epsilon \to 0.$$

c) Suppose $f \in C^k(\mathbb{R}^n)$ with $\sup_{\mathbb{R}^n}|D^\alpha f| < \infty$ for $|\alpha| \leqslant k$, and suppose $g \in L^1(\mathbb{R}^n)$ with $g \geqslant 0$, $\int_{\mathbb{R}^n} g(x)dx = 1$. Set $g_\epsilon(y) = \epsilon^{-n} g\left(\epsilon^{-1}y\right)$. Then $f \star g_\epsilon \in C^k(\mathbb{R}^n)$, and

$$D^\alpha\left(f \star g_\epsilon\right)(x) \to D^\alpha f(x) \qquad \textit{as } \epsilon \to 0,$$

for any $x \in \mathbb{R}^n$ and any multi-index with $|\alpha| \leqslant k$.

Proof. a) Note that the rescaling of ϕ to produce ϕ_ϵ is such that a change of variables gives:

$$\int_{\mathbb{R}^n} \phi_\epsilon(y)dy = 1.$$

By Theorem 1.10, we have that $D^\alpha(\phi_\epsilon \star f) = \phi_\epsilon \star D^\alpha f$ for any $|\alpha| \leqslant k$. Using these two facts, we calculate:

$$\begin{aligned} D^\alpha(\phi_\epsilon \star f)(x) - D^\alpha f(x) &= \int_{\mathbb{R}^n} \phi_\epsilon(y) D^\alpha f(x-y)dy - D^\alpha f(x)\int_{\mathbb{R}^n}\phi_\epsilon(y)dy \\ &= \int_{\mathbb{R}^n} \phi_\epsilon(y)\left[D^\alpha f(x-y) - D^\alpha f(x)\right]dy \\ &= \int_{B_1(0)} \phi(z)\left[D^\alpha f(x-\epsilon z) - D^\alpha f(x)\right]dz \end{aligned}$$

where in the last line we made the substitution $y = \epsilon z$, and noted that ϕ has support in $B_1(0)$, so we can restrict the range of integration. Now, since $\phi \geqslant 0$, we can estimate:

$$\begin{aligned}|D^\alpha(\phi_\epsilon \star f)(x) - D^\alpha f(x)| &\leqslant \int_{B_1(0)} \phi(z) \, |D^\alpha f(x - \epsilon z) - D^\alpha f(x)| \, dz \\ &\leqslant \sup_{z \in B_1(0)} |D^\alpha f(x - \epsilon z) - D^\alpha f(x)| \times \int_{B_1(0)} \phi(z) dz \\ &= \sup_{z \in B_1(0)} |D^\alpha f(x - \epsilon z) - D^\alpha f(x)|\end{aligned}$$

since $\int_{\mathbb{R}^n} \phi = 1$. Now, since $D^\alpha f$ is continuous and of compact support, it is uniformly continuous on $\mathbb{R}^n$. Fix $\tilde{\epsilon} > 0$. There exists δ such that for any $v, w \in \mathbb{R}^n$ with $|x - y| < \delta$, we have

$$|D^\alpha f(v) - D^\alpha f(w)| < \tilde{\epsilon}$$

For any $x \in \mathbb{R}^n$, taking $\epsilon < \delta$, and $v = x - \epsilon z$, $w = x$ with $z \in B_1(0)$ we have $|v - w| < \delta$, so:

$$|D^\alpha f(x - \epsilon z) - D^\alpha f(x)| < \tilde{\epsilon}$$

holds for any $x \in \mathbb{R}^n, z \in B_1(0)$. We have therefore shown that for any $\tilde{\epsilon} > 0$, there exists δ such that for any $\epsilon < \delta$ we have:

$$\sup_{x \in \mathbb{R}^n} |D^\alpha(\phi_\epsilon \star f)(x) - D^\alpha f(x)| < \tilde{\epsilon}.$$

This is the statement of uniform convergence on $\mathbb{R}^n$.

b) Noting that $L^p(\mathbb{R}^n) \subset L^1_{loc.}(\mathbb{R}^n)$ by an application of Hölder's inequality (see Exercise 1.2), Theorem 1.10 immediately establishes the smoothness of $\phi_\epsilon \star g$. To establish convergence as $\epsilon \to 0$, we shall require certain measure theoretic results. First we require Minkowski's Integral Identity (see Exercise 1.4). This states[6] that for $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$ a measurable function, we have the estimate:

$$\left[\int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} F(x, y) dx \right|^p dy \right]^{\frac{1}{p}} \leqslant \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} |F(x, y)|^p \, dy \right]^{\frac{1}{p}} dx$$

Now, following the calculation in the previous proof, we readily have that:

$$|(\phi_\epsilon \star g)(x) - g(x)| \leqslant \int_{\mathbb{R}^n} \phi(z) \, |g(x - \epsilon z) - g(x)| \, dz$$

[6] There is more general statement for a map $F : X \times Y \to \mathbb{C}$, which is measurable with respect to the product measure $\mu \times \nu$ where (X, μ) and (Y, ν) are measure spaces.

Integrating and applying Minkowski's integral inequality, we have:

$$\begin{aligned}
\|\phi_\epsilon \star g - g\|_{L^p(\mathbb{R}^n)} &= \left[\int_{\mathbb{R}^n} |(\phi_\epsilon \star g)(x) - g(x)|^p \, dx\right]^{\frac{1}{p}} \\
&\leqslant \left[\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} \phi(z) \, |g(x-\epsilon z) - g(x)| \, dz\right|^p dx\right]^{\frac{1}{p}} \\
&\leqslant \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} \phi(z)^p \, |g(x-\epsilon z) - g(x)|^p \, dx\right]^{\frac{1}{p}} dz \\
&= \int_{\mathbb{R}^n} \phi(z) \, \|\tau_{\epsilon z} g - g\|_{L^p(\mathbb{R}^n)} \, dz \qquad (1.3)
\end{aligned}$$

To establish our result it will suffice to set $\epsilon = \epsilon_j$, where $\{\epsilon_j\}_{j=1}^\infty \subset \mathbb{R}$ is any sequence with $\epsilon_j \to 0$, and show that $\left\|\phi_{\epsilon_j} \star g - g\right\|_{L^p(\mathbb{R}^n)} \to 0$. Note that since $\left\|\tau_{\epsilon_j z} g\right\|_{L^p(\mathbb{R}^n)} = \|g\|_{L^p(\mathbb{R}^n)}$ we have:

$$\phi(z) \left\|\tau_{\epsilon_j z} g - g\right\|_{L^p(\mathbb{R}^n)} \leqslant 2\phi(z) \, \|g\|_{L^p(\mathbb{R}^n)}$$

so the integrand is dominated uniformly in j by an integrable function. Now by Lemma 1.8, as y varies, $\tau_y : L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$ is a continuous family of bounded linear operators. This means that for each $z \in \mathbb{R}^n$ we have:

$$\lim_{j\to\infty} \left\|\tau_{\epsilon_j z} g - g\right\|_{L^p(\mathbb{R}^n)} = 0.$$

Thus we can apply the Dominated Convergence Theorem (Theorem B.26) to the integral on the right hand side of 1.3, and conclude that

$$\lim_{j\to\infty} \left\|\phi_{\epsilon_j} \star g - g\right\|_{L^p(\mathbb{R}^n)} = 0.$$

c) Again, by Theorem 1.10, we have that $D^\alpha(f \star g_\epsilon) = D^\alpha f \star g_\epsilon$ for any $|\alpha| \leqslant k$. By a change of variables, we calculate:

$$D^\alpha(f \star g_\epsilon)(x) = \int_{\mathbb{R}^n} g_\epsilon(y) D^\alpha f(x-y) dy = \int_{\mathbb{R}^n} g(z) D^\alpha f(x - \epsilon z) dz$$

Now, clearly for each fixed $x \in \mathbb{R}^n$:

$$g(z) D^\alpha f(x - \epsilon z) \to g(z) D^\alpha f(x)$$

for $z \in \mathbb{R}^n$ as $\epsilon \to 0$. Furthermore,

$$|g(z) D^\alpha f(x - \epsilon z)| \leqslant g(z) \sup_{\mathbb{R}^n} |D^\alpha f|$$

which is an integrable function of z, so by the Dominated convergence theorem, we conclude:

$$D^\alpha(f \star g_\epsilon)(x) \to D^\alpha f(x) \int_{\mathbb{R}^n} g(z) dz = D^\alpha f(x)$$

as $\epsilon \to 0$.

□

The final application of convolutions is to the construction of cut-off functions. These are often extremely useful for localising a problem to a particular region of interest for some reason or other.

Lemma 1.14. *Suppose $\Omega \subset \mathbb{R}^n$ is open, and $K \subset \Omega$ is compact. Then there exists $\chi \in C_c^\infty(\Omega)$ such that $\chi = 1$ in a neighbourhood of K.*

Proof. By Lemma 1.3, there exists $\epsilon > 0$ such that $d(K, \partial\Omega) > 4\epsilon$. We define $K_\epsilon = K + \overline{B_{2\epsilon}(0)}$. As the sum of two compact sets, K_ϵ is compact. Moreover, $K_\epsilon \subset \Omega$. Suppose ϕ_ϵ is as in Theorem 1.13. Consider:

$$\chi := \phi_\epsilon \star \mathbb{1}_{K_\epsilon}.$$

We have by Theorem 1.10 that $\chi \in C^\infty(\mathbb{R}^n)$ and from Lemma 1.12 we deduce:

$$\operatorname{supp}\chi = K_\epsilon + \operatorname{supp}\phi_\epsilon \subset K + \overline{B_{2\epsilon}(0)} + \overline{B_\epsilon(0)} = K + \overline{B_{3\epsilon}(0)} \subset \Omega.$$

Thus $\chi \in C_c^\infty(\Omega)$. Now, suppose $x \in K + B_\epsilon(0)$. Then $x + B_\epsilon(0) \subset K_\epsilon$ and so:

$$\begin{aligned}\chi(x) &= \int_{\mathbb{R}^n} \phi_\epsilon(y)\mathbb{1}_{K_\epsilon}(x-y)dy \\ &= \int_{B_\epsilon(0)} \phi_\epsilon(y)\mathbb{1}_{K_\epsilon}(x-y)dy \\ &= \int_{B_\epsilon(0)} \phi_\epsilon(y)dy = 1.\end{aligned}$$

Thus $\chi(x) = 1$ for $x \in K + B_\epsilon(0)$, which is a neighbourhood of K. □

The following exercise establishes results required for the proof of Theorem 1.13.

Exercise 1.4. a) Suppose $1 \leqslant p \leqslant \infty$ and let q satisfy $p^{-1} + q^{-1} = 1$. Show that for a measurable function $f : \mathbb{R}^n \to \mathbb{C}$:

$$\|f\|_{L^p} = \sup\left\{\int_{\mathbb{R}^n} |f(x)g(x)|\, dx : g \in L^q(\mathbb{R}^n), \|g\|_{L^q} \leqslant 1\right\}.$$

b) Now suppose $p < \infty$ and assume $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$ is integrable. Set $G(y) = \int_{\mathbb{R}^n} F(x,y)dx$. Show that if $\|g\|_{L^q} \leqslant 1$ then

$$\int_{\mathbb{R}^n} |G(y)g(y)|\, dy \leqslant \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} |F(x,y)|^p\, dy\right]^{\frac{1}{p}} dx.$$

Deduce Minkowski's integral inequality

$$\left[\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} F(x,y)dx\right|^p dy\right]^{\frac{1}{p}} \leqslant \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} |F(x,y)|^p\, dy\right]^{\frac{1}{p}} dx.$$

Exercise 1.5. Let $I = (0,1)$ and $1 \leq p < \infty$. Exhibit a sequence $(f_j)_{j=1}^\infty$ with $f_j \in L^p(I)$ such that $f_j \to 0$ in $L^p(I)$, but $f_j(x)$ does not converge for any x. Does such a sequence exist if $p = \infty$?

1.5 Lebesgue differentiation theorem

The fundamental theorem of calculus is a fundamental result in the theory of Riemann integration. It comes in two (related) flavours.

Theorem 1.15 (Fundamental Theorem of Calculus)**.** *Suppose that $f : [a,b] \to \mathbb{R}$ is continuous, and define the function $F : [a,b] \to \mathbb{R}$ by:*

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

Then F is differentiable on (a,b), and:

$$F'(x) = f(x).$$

From this one can deduce the alternative form of the Fundamental Theorem of Calculus, relating the integral of a function to its anti-derivative.

Corollary 1.16. *Suppose that $f : [a,b] \to \mathbb{R}$ is continuous and that $F : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$, differentiable on (a,b) and satisfies $F'(x) = f(x)$ for all $x \in (a,b)$. Then:*

$$\int_a^b f(t)dt = F(b) - F(a).$$

We seek to generalise Theorem 1.15 in the setting of the Lebesgue integral. First, we note that the result implies

$$\begin{aligned} f(x) &= \lim_{r\to 0} \frac{F(x+r) - F(x-r)}{2r} = \lim_{r\to 0} \frac{1}{2r} \int_{x-r}^{x+r} f(t)dt \\ &= \lim_{r\to 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} f(t)dt \end{aligned}$$

where we have used that the ball of radius r about x is simply $B_r(x) = (x-r, x+r)$ in one dimension. Rearranging, we can further conclude

$$\lim_{r\to 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} \left(f(t) - f(x)\right) dt = 0.$$

This statement is meaningful in dimensions higher that one. In fact we shall prove something slightly stronger

Theorem 1.17 (Lebesgue differentiation theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{C}$ be integrable. Then for almost every $x \in \mathbb{R}^n$ we have*

$$\lim_{r\to 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy = 0. \tag{1.4}$$

Note that it suffices for f to be defined on an open set $\Omega \subset \mathbb{R}^n$, and we obtain the same differentiability result at almost every $x \in \Omega$ by considering $f\mathbb{1}_\Omega$. We say that a point x such that (1.4) holds is a *Lebesgue point* of f. In order to establish this result, we first introduce a related quantity for which we are able to prove an estimate.

Definition 1.5. *Given an integrable function $f : \mathbb{R}^n \to \mathbb{C}$, the Hardy–Littlewood Maximal function Mf is defined to be*

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

The Hardy–Littlewood Maximal function and its generalisations are of use in many contexts in mathematics, in particular in harmonic anaysis. For our purposes, a key result is that it satisfies a *weak* L^1-bound.

Lemma 1.18 (Weak L^1-bound for Mf)**.** *Suppose $f \in L^1(\mathbb{R}^n)$ for $n \geqslant 1$. Then Mf is measurable, finite almost everywhere, and there exists a constant C_n, depending only on n such that:*

$$|\{x : Mf(x) > \lambda\}| \leqslant \frac{C_n}{\lambda} \|f\|_{L^1} \tag{1.5}$$

for all $\lambda > 0$.

Proof. Let $A_\lambda = \{x : Mf(x) > \lambda\}$. Then for each $x \in A_\lambda$ there exists a radius r_x such that

$$\frac{1}{|B_{r_x}(x)|} \int_{B_{r_x}(x)} |f(y)|\, dy > \lambda. \tag{1.6}$$

We claim A_λ is open, which implies Mf is measurable. To see this, suppose $x \in A_\lambda$ with corresponding r_x and let $(x_m)_{m=1}^\infty$ be a sequence with $x_m \to x$ and $x_m \notin A_\lambda$. Then by the dominated convergence theorem we have

$$\frac{1}{|B_{r_x}(x)|} \int_{B_{r_x}(x_m)} |f(y)|\, dy \to \frac{1}{|B_{r_x}(x)|} \int_{B_{r_x}(x)} |f(y)|\, dy,$$

however, by assumption

$$\frac{1}{|B_{r_x}(x)|} \int_{B_{r_x}(x_m)} |f(y)|\, dy \leqslant \lambda, \qquad \frac{1}{|B_{r_x}(x)|} \int_{B_{r_x}(x)} |f(y)|\, dy > \lambda,$$

a contradiction.

Fix $K \subset A_\lambda$ a compact set. Since K is covered by $\cup_{x\in A_\lambda} B_{r_x}(x)$, we can pick a finite subcover of K, say $K \subset \cup_{i=1}^N B_i$, where $B_i = B_{r_x}(x)$ for some x. By Wiener's covering Lemma (see Exercise 1.8) there is a disjoint subcollection $B_{i_1}, \dots, B_{i_k}$ such that

$$|K| \leqslant \left| \bigcup_{i=1}^N B_i \right| \leqslant 3^n \sum_{j=1}^n \left|B_{i_j}\right|.$$

Now, each B_{i_j} satisfies (1.6), so we have

$$|K| \leqslant \frac{3^n}{\lambda} \sum_{j=1}^n \int_{B_{i_j}} |f(y)|\, dy \leqslant \frac{3^n}{\lambda} \|f\|_{L^1}$$

Where in the final inequality we use that the B_{i_j} are disjoint. Since this holds for all compact $K \subset A_\lambda$, (1.5) follows. Finally, note that $\{Mf = \infty\} \subset \{Mf > \lambda\}$, which implies $|\{Mf = \infty\}| < C/\lambda$ for all λ, thus $|\{Mf = \infty\}| = 0$. $\square$

With this result in hand, we are now ready to establish the Lebesgue differentiation theorem.

Proof of Theorem 1.17. For each $\lambda > 0$ define:

$$A_\lambda = \left\{ x \in \mathbb{R}^n : \limsup_{r \to 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy > 2\lambda \right\}$$

If we can show that $|A_\lambda| = 0$ for all λ, then we will be done, as the set of $x \in \mathbb{R}^n$ which are not Lebesgue points for f is precisely $\cup_{n=1}^{\infty} A_{\frac{1}{n}}$.

Fix $\epsilon > 0$. We can find $g \in C_c^\infty(\mathbb{R}^n)$ such that $\|f - g\|_{L^1} < \epsilon$. We estimate:

$$\begin{aligned} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy &\leqslant \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - g(y)|\, dy \\ &\quad + \frac{1}{|B_r(x)|} \int_{B_r(x)} |g(y) - g(x)|\, dy + |g(x) - f(x)| \end{aligned}$$

We can bound the first term by

$$\begin{aligned} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - g(y)|\, dy &\leqslant \sup_{r>0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - g(y)|\, dy \\ &= M[f - g](x) \end{aligned}$$

Now, since g is continuous, we have

$$\limsup_{r \to 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |g(y) - g(x)|\, dy = 0,$$

hence

$$\limsup_{r \to 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)|\, dy \leqslant M[f - g](x) + |f(x) - g(x)|\,.$$

Now, if $x \in A_\lambda$, then we must either have $M[f - g](x) > \lambda$ or $|f(x) - g(x)| > \lambda$. By Lemma 1.18 we know

$$|\{x : M[f - g](x) > \lambda\}| \leqslant \frac{C_n}{\lambda} \|f - g\|_{L^1} \leqslant \frac{C_n \epsilon}{\lambda},$$

and by Tchebychev's inequality we know

$$|\{x : |f(x) - g(x)| > \lambda\}| \leqslant \frac{\|f - g\|_{L^1}}{\lambda} \leqslant \frac{\epsilon}{\lambda}$$

we conclude that

$$|A_\lambda| \leqslant \frac{1 + C_n}{\lambda} \epsilon.$$

Since ϵ was arbitrary, we conclude $|A_\lambda| = 0$, and we're done. □

Exercise 1.6. Suppose $1 \leqslant p < \infty$.

a) Suppose $f \in L^p(\mathbb{R}^n)$. Show that

$$|\{x : |f(x)| > \lambda\}| \leqslant \frac{\|f\|_{L^p}^p}{\lambda^p}.$$

This is known as Tchebychev's inequality, the $p = 1$ case is Markov's inequality.

b) We say that a measurable $f : \mathbb{R}^n \to \mathbb{C}$ is in *weak-$L^p(\mathbb{R}^n)$*, written $f \in L^{p,w}(\mathbb{R}^n)$ if there exists a constant C such that

$$|\{x : |f(x)| > \lambda\}| \leqslant \frac{C^p}{\lambda^p}.$$

Show that $L^p(\mathbb{R}^n) \subset L^{p,w}(\mathbb{R}^n)$, and that the inclusion is proper.

Exercise 1.7. Suppose that $f \in L^r(\mathbb{R}^n)$ for some $1 \leqslant r < \infty$. Show that $\|f\|_{L^\infty} = \lim_{p\to\infty} \|f\|_{L^p}$.
[Hint: you may find the estimates in Exercises 1.2 b), 1.6 a) useful.]

Exercise 1.8. a) Let $B_1, \ldots, B_N$ be a finite collection of open balls in $\mathbb{R}^n$. Show that there exists a subcollection $B_{i_1}, \ldots, B_{i_k}$ of *disjoint* balls such that

$$\bigcup_{i=1}^{N} B_i \subset \bigcup_{j=1}^{k} (3B_{i_j}),$$

where $3B$ is the ball with the same centre as B but three times the radius. Deduce

$$\left|\bigcup_{i=1}^{N} B_i\right| \leqslant 3^n \sum_{j=1}^{k} \left|B_{i_j}\right|.$$

b) (*) Suppose $\{B_j : j \in J\}$ is an arbitrary collection of balls in $\mathbb{R}^n$ such that each ball has radius at most R. Show that there exists a countable subcollection $\{B_j : j \in J'\}$, $J' \subset J$ of disjoint balls such that

$$\bigcup_{i\in J} B_i \subset \bigcup_{i\in J'} (5B_i).$$

These are Wiener and Vitali's covering Lemmas, respectively.

Exercise 1.9. Suppose $f : \mathbb{R} \to \mathbb{C}$ is integrable and let $F(x) = \int_{-\infty}^{x} f(t)dt$. Show that F is differentiable with $F'(x) = f(x)$ at each Lebesgue point $x \in \mathbb{R}$. Deduce that F is differentiable almost everywhere.

Exercise 1.10. Suppose $\phi \in L^\infty(\mathbb{R}^n)$ satisfies $\phi \geqslant 0$, supp $\phi \subset B_1(0)$, and $\int_{\mathbb{R}^n} \phi\, dx = 1$. Set $\phi_\epsilon(x) = \epsilon^{-n}\phi(\epsilon^{-1}x)$. Show that if $f \in L^1(\mathbb{R}^n)$, and x is a Lebesgue point of f,

$$\phi_\epsilon \star f(x) \to f(x), \qquad \text{as } \epsilon \to 0.$$

1.6 Littlewood's principles: Egorov's Theorem and Lusin's Theorem

In his 1944 "Lectures on the Theory of Functions", J. E. Littlewood stated three principles:

> "Every (measurable) set is nearly a finite sum of intervals; every function (of class L^p) is nearly continuous; every convergent sequence of functions is nearly uniformly convergent."

The first of these results may be stated more precisely in our language as follows:

Lemma 1.19. *Suppose $|A| < \infty$. Then for any $\epsilon > 0$ there exists a set B, which is a finite union of rectangles, such that*

$$|A \Delta B| < \epsilon.$$

This follows straightforwardly from the basic properties of Lebesgue measurable sets. The third of Littlewood's principles follows from

Theorem 1.20 (Egorov's Theorem)**.** *Suppose $(f_k)_{k=1}^{\infty}$ is a sequence of functions defined on a set $E \subset \mathbb{R}^n$ with $|E| < \infty$, and suppose that $f_k \to f$ almost everywhere on E. Then given $\epsilon > 0$ we can find a closed set $A_\epsilon \subset E$ such that $|E - A_\epsilon| \leqslant \epsilon$ and $f_k \to f$* uniformly *on A_ϵ.*

Proof. By discarding a set of measure zero if necessary, we can assume without loss of generality that $f_k(x) \to f(x)$ for all $x \in E$. For each $n, k \in \mathbb{N}$ let

$$E_k^n = \left\{ x \in E : |f_j(x) - f(x)| < \frac{1}{n}, \text{ for all } j > k \right\}.$$

Fixing n, we note that $E_k^n \subset E_{k+1}^n$ and that $\cup_{k=1}^{\infty} E_k^n = E$. By countable additivity, we have $|E_k^n|$ is an increasing sequence, with $|E_k^n| \to |E|$ as $k \to \infty$. Pick k_n such that $\left|E \setminus E_{k_n}^n\right| < 2^{-n}$. By construction we have:

$$|f_j(x) - f(x)| < \frac{1}{n}, \qquad \text{for all } j > k_n \text{ and } x \in E_{k_n}^n.$$

Now pick N such that $\sum_{n=N}^{\infty} 2^{-n} < \epsilon/2$ and let

$$A'_\epsilon = \bigcap_{n=N}^{\infty} E_{k_n}^n.$$

Now we observe

$$\left|E \setminus A'_\epsilon\right| \leqslant \sum_{n=N}^{\infty} \left|E \setminus E_{k_n}^n\right| < \epsilon/2.$$

Next. suppose that $\delta > 0$. Pick $n \geqslant N$ such that $1/n < \delta$, and note that $x \in A'_\epsilon$ implies $x \in E_{k_n}^n$. We deduce that $|f_j(x) - f(x)| < \delta$ for all $j > k_n$ and hence $f_j \to f$ uniformly on A'_ϵ. Finally, we can pick a closed set $A_\epsilon \subset A'_\epsilon$ such that $|A'_\epsilon \setminus A_\epsilon| < \epsilon/2$ and hence $|E - A_\epsilon| \leqslant \epsilon$. ☐

The final of Littlewood's principles is given flesh by

Theorem 1.21 (Lusin's Theorem)**.** *Suppose f is measurable and finite valued on E, where $E \subset \mathbb{R}^n$ with $|E| < \infty$. Then given $\epsilon > 0$ we can find a closed set $F_\epsilon \subset E$ with $|E \setminus F_\epsilon| < \epsilon$ such that $f|_{F_\epsilon}$ is continuous.*

Proof. Suppose first f is a simple function

$$f = \sum_{k=1}^{m} a_k \mathbb{1}_{A_k},$$

where $|A_k| < \infty$ and the A_k are disjoint with $E = \cup_{k=1}^m A_k$ (if necessary, we add the term $0\mathbb{1}_{f^{-1}(0)}$ to arrange this). For any $\epsilon > 0$, we can pick compact sets $K_k \subset A_k$ with

$$|A_k \setminus K_k| < \frac{\epsilon}{m}.$$

Let $B = \cup_{k=1}^m K_k$. Then $|E \setminus B| < \epsilon$. Since the sets K_k are compact and disjoint (hence $\min_{i,j} \operatorname{dist}(K_i, K_j) > 0$), and f is constant on each K_k, we have that f is continuous on B.

Now let f_n be a sequence of simple functions such that $f_n \to f$ ae. Then we can find C_n such that $|C| < 2^{-n}$ and f_n is continuous outside C_n. By Egorov's theorem, we can find a set $A_{\epsilon/3}$ such that $f_n \to f$ uniformly on $A_{\epsilon/3}$ and $\left|E \setminus A_{\epsilon/3}\right| < \epsilon/3$. Let N be sufficiently large that $\sum_{n=N}^{\infty} 2^{-n} < \epsilon/3$.

$$F'_\epsilon = A_{\epsilon/3} \setminus \bigcup_{n=N}^{\infty} C_n$$

Now, $|E \setminus F'_\epsilon| < 2\epsilon/3$ and moreover, for $n > N$ the functions f_n are continuous on F'_ϵ, so since they converge uniformly to f, we have that f is continuous on F'_ϵ. Finally, picking $F_\epsilon \subset F'_\epsilon$ closed with $|F'_\epsilon \setminus F_\epsilon| < \epsilon/3$ we're done. □

Remark. *Note that Lusin's Theorem asserts that $f|_{F_\epsilon}$ is continuous, which means that f is continuous if we think of it as defined only at points of F_ϵ. This is not the same as the statement that f (defined on E) is continuous at points of F_ϵ. For example if $f = \mathbb{1}_{\mathbb{Q}}$, then $f|_{\mathbb{R}\setminus\mathbb{Q}} = 0$ is continuous, however f is nowhere continuous.*

Chapter 2

Banach and Hilbert space analysis

2.1 Hilbert Spaces

2.1.1 Review of Hilbert space theory

We will briefly review the theory of Hilbert spaces in order to fix conventions. A (complex) inner product space is a complex vector space H, equipped with an inner product $(\cdot,\cdot) : H \times H \to \mathbb{C}$ such that

i) $(\cdot,\cdot)$ is sesquilinear[1]:

$$(x+y,z) = (x,z) + (y,z) \qquad (x,y+z) = (x,y) + (x,z)$$
$$(\alpha x, y) = \overline{\alpha}(x,y) \qquad (x,\alpha,y) = \alpha(x,y).$$

for all $x, y, z \in H, \alpha \in \mathbb{C}$

ii) $(\cdot,\cdot)$ is hermitian and positive:

$$(x,y) = \overline{(y,x)}, \qquad (x,x) \geqslant 0$$

for all $x, y \in H$, with equality in the second expression iff $x = 0$.

We define $\|\cdot\| : x \mapsto \sqrt{(x,x)}$.

Lemma 2.1. *The map $\|\cdot\| : H \to [0,\infty)$ is a norm on H, which satisfies the Cauchy-Schwarz inequality:*

$$|(x,y)| \leqslant \|x\| \, \|y\| \,, \qquad \textit{for all } x, y \in H, \ \textit{with `$=$' iff } x = \alpha y, \ \textit{for some } \alpha \in \mathbb{C},$$

and the parallelogram identity:

$$2\|x\|^2 + 2\|y\|^2 = \|x+y\|^2 + \|x-y\|^2$$

[1] There is a choice of convention over whether the first or second entry is anti-linear. Our choice here is that typically used in quantum mechanics, in pure maths the opposite convention is often used. What matters, however, is to be consistent!

Proof. Positivity and homogeneity of $\|\cdot\|$ are immediate from the definition. Fix $x, y \in H$. For any $t \in \mathbb{R}$ we have that $\|x + ty\|^2 \geqslant 0$, so expanding using we see

$$\|x\|^2 + 2t\Re(x, y) + t^2 \|y\|^2 \geqslant 0.$$

A non-negative definite quadratic must have non-positive discriminant, hence we must have $(\Re(x, y))^2 \leqslant \|x\|^2 \|y\|^2$. Replacing $x \to e^{i\theta}x$ for suitable θ, we deduce the Cauchy-Schwarz inequality. Now, we compute:

$$\|x + y\|^2 = \|x\|^2 + 2\Re(x, y) + \|y\|^2 \leqslant \|x\|^2 + 2\|x\|^2 \|y\|^2 + \|y\|^2 = (\|x\| + \|y\|)^2$$

so taking square roots the triangle inequality follows. The parallelogram law can be easily verified by expanding the right-hand side. □

Definition 2.1. *A* Hilbert space *is a complex inner product space* $(H, (\cdot, \cdot))$*, such that the associated metric* $\|\cdot\|$ *is complete.*

Thus a Hilbert space is a special case of a Banach space, however the presence of the inner product gives the space a more geometric character. In particular, we have a natural notion of 'orthogonality'. If $K \subset H$, we write

$$K^\perp = \{x \in H : (z, x) = 0, \text{ for all } z \in K\}.$$

Note that $K \cap K^\perp \subset \{0\}$ and $K \subset M \implies M^\perp \subset K^\perp$.

Lemma 2.2. *For any* $K \subset H$, $K^\perp$ *is a closed subspace.*

Proof. For any $z \in H$, the map $\Lambda_z : H \to \mathbb{C}$ given by $x \mapsto (z, x)$ is a bounded linear map. The linearity follows from properties of the inner product, and boundedness follows as:

$$|\Lambda_z x| = |(z, x)| \leqslant \|z\| \|x\|$$

by Cauchy-Schwarz. Thus Ker Λ_z is a closed subspace of H for any $z \in H$. By definition

$$K^\perp = \bigcap_{z \in K} \text{Ker } \Lambda_z,$$

so $K^\perp$ is a closed subspace. □

An important result concerns closed, convex sets in a Hilbert space (see Definition A.3)

Theorem 2.3. *Let* K *be a nonempty, closed, convex set in a Hilbert space* $(H, (\cdot, \cdot))$*. Then* K *contains a unique element of least norm.*

Proof. Let $d = \inf\{\|x\| : x \in K\}$. For any $x, y \in K$ the parallelogram identity applied to $x/2$, $y/2$ gives:

$$\frac{1}{4}\|x - y\|^2 = \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \left\|\frac{x + y}{2}\right\|^2$$

The convexity of K implies $(x+y)/2 \in K$, so

$$\|x-y\|^2 \leqslant 2\|x\|^2 + 2\|y\|^2 - 4d^2. \tag{2.1}$$

We immediately deduce that if $\|x\| = \|y\| = d$, then $x = y$, so the infimum of the distance can be attained by at most one point. Now suppose $(x_n)_{n=1}^{\infty}$ is a sequence with $x_n \in K$ and $\|x_n\| \to d$. By (2.1) we have:

$$\|x_n - x_m\|^2 \leqslant 2\|x_n\|^2 + 2\|x_m\|^2 - 4d^2$$

which implies (x_n) is Cauchy, hence $x_n \to x$ for some $x \in H$. Since K is closed, $x \in K$, and since the norm is continuous $\|x\| = d$ and we're done. □

Noting that K is closed and convex if and only if $K + y$ is closed and convex, we immediately deduce

Corollary 2.4. *Let K be a nonempty, closed, convex set and let $x \in H$. There is a unique $y \in K$ which minimises $\|x-y\|$.*

A more important corollary of this result concerns orthogonal projection onto closed linear subspaces.

Lemma 2.5. *Let L be a closed linear subspace of H. There exists a bounded linear operator $P : H \to L$ such that*

i) $Px = x$ *if* $x \in L$,

ii) $x - Px \in L^{\perp}$,

iii) $Px = 0$ *if* $x \in L^{\perp}$,

iv) $\|Px\| \leqslant \|x\|$.

Proof. For any $x \in H$, let Px be the unique point in L which minimises $\|x - Px\|$. It is immediate that $Px = x$ if $x \in L$. Now, suppose $y \in L$ and consider $\|x - Px - t\alpha y\|$ for $t \in \mathbb{R}$, where α is chosen to satisfy $\Re(\alpha y, x - Px) = |(y, x - Px)|$, $|\alpha| = 1$. We find

$$\|x - Px - t\alpha y\|^2 = \|x - Px\|^2 + 2t\,|(y, x-Px)| + t^2\|y\|^2 \geqslant \|x - Px\|^2$$

so $|(y, x-Px)| = 0$, and hence $x - Px \in L^{\perp}$. This immediately implies that if $x \in L^{\perp}$, then $Px \in L \cap L^{\perp} = \{0\}$. Next, we note that

$$\|x\|^2 = \|x - Px + Px\|^2 = \|x - Px\|^2 + \|Px\|^2,$$

from which the bound on $\|Px\|$ follows. Finally, to see that L is linear, for $\lambda \in \mathbb{C}$, $x, y \in H$ write $P(x + \lambda y) = P(x) + \lambda P(y) + z$ for some $z \in L$. We observe:

$$\begin{aligned}\|(x+\lambda y) - P(x+\lambda y)\|^2 &= \|x - P(x) + \lambda(y - P(y)) - z\|^2 \\ &= \|x - P(x) + \lambda(y - P(y))\| + \|z\|^2\end{aligned}$$

which is clearly minimised for $z = 0$, hence $P(x + \lambda y) = P(x) + \lambda P(y)$. □

An immediate corollary of the result is

Corollary 2.6. *If L is a closed linear subspace of H, then*

$$H = L \oplus L^{\perp}.$$

Proof. Given $x \in H$ we can write $x = Px + (x - Px)$ with $Px \in L$, $(x - Px) \in L^{\perp}$, and certainly $x \in L \cap L^{\perp}$ if and only if $x = 0$. □

We also have the following useful characterisation of $K^{\perp\perp} := (K^{\perp})^{\perp}$

Corollary 2.7. *If L is a closed linear subspace of H then $L^{\perp\perp} = L$. If $K \subset H$ is any set, we have $K^{\perp\perp} = \overline{\text{span } K}$. In particular, K is dense in H if and only if $K^{\perp} = \{0\}$.*

Proof. Clearly $L \subset L^{\perp\perp}$. Suppose $z \in L^{\perp\perp}$, then $z = x + y$ with $x \in L$, $y \in L^{\perp}$ by the previous Corollary. Thus $0 = (z, y) = (x + y, y) = \|y\|^2$, so $y = 0$ and $z = x \in L$, thus $L^{\perp\perp} \subset L$.

Now consider an arbitrary $K \subset H$. It is clear from the definition of the orthogonal complement that $\overline{\text{span } K} \subset K^{\perp\perp}$. Since $K \subset \overline{\text{span } K}$, we deduce $\overline{\text{span } K}^{\perp} \subset K^{\perp}$ and $K^{\perp\perp} \subset \overline{\text{span } K}^{\perp\perp} = \overline{\text{span } K}$, making use of the fact that $\overline{\text{span } K}$ is a closed subspace. □

The final result of this section shows that for a Hilbert space H, we can describe the dual space H' in a straightforward fashion. Recall that for a topological vector space X, the dual space X' is defined to be the set of continuous linear maps $\Lambda : X \to \mathbb{C}$, sometimes known as the continuous linear *functionals* on X. We saw in the proof of Lemma 2.2 that to any $z \in H$ we can associate a continuous linear map $\Lambda_z : x \mapsto (z, x)$. The following result, known as Riesz representation theorem shows that in fact *any* element of H' must be of this form.

Theorem 2.8 (Riesz representation theorem for Hilbert spaces)**.** *If $\Lambda : H \to \mathbb{C}$ is a continuous linear functional on H, then there is a unique $z \in H$ such that $\Lambda = \Lambda_z$.*

Proof. If $\Lambda x = 0$ for all x, take $z = 0$, otherwise let

$$L = \{x : \Lambda x = 0\}.$$

L is a subspace by the linearity of Λ, and it is closed by the continuity of Λ. Moreover, since $\Lambda x \neq 0$ for some x, $L^{\perp}$ cannot be trivial, as $H = L \oplus L^{\perp}$ and $L \neq H$.

Pick $y \in L^{\perp}$ such that $\|y\| = 1$ and for any $x \in H$ let

$$w = (\Lambda x)y - (\Lambda y)x.$$

Clearly $\Lambda w = 0$, so $w \in L$, and as a result $(y, w) = 0$, which implies

$$\Lambda x = (\Lambda y)(y, x) = \big((\overline{\Lambda y})y, x\big)$$

so taking $z = (\overline{\Lambda y})y$ we have established $\Lambda = \Lambda_z$. To see that z is unique, suppose $\Lambda_z = \Lambda_{z'}$, then for any $x \in H$ we have:

$$(z - z', x) = 0,$$

in particular this holds for $x = z - z'$, so that $z = z'$. □

2.1.2 The Hilbert space L^2

Given any measure space $(E, \mathcal{E}, \mu)$, the space $L^2(E, \mu)$ is naturally a Hilbert space, with inner product given by:

$$(f, g)_{L^2} := \int_E \overline{f} g d\mu.$$

We will mostly focus on the cases where E is $\mathbb{R}^n$ (or a subset) equipped with Lebesgue measure, although for some purposes it is convenient to keep the discussion more general.

Orthogonal systems of functions and their completeness

Suppose $S = \{u_j\}_{j \in J}$ is a subset of a Hilbert space H indexed by some (not necessarily countable) set J. We say S is *orthogonal* if $(u_j, u_k) = 0$ for all $j, k \in J$ with $j \neq k$. We say S is *orthonormal* if additionally $\|u_j\| = 1$ for all $j \in J$. We say that S is *complete* if $\overline{\text{Span } S} = H$, where Span S is the set of finite linear combinations of elements of S. An orthonormal set which is complete, we refer to as an *orthonormal basis*. For many purposes, we can take the set J to be $\mathbb{N}$, thanks to the following result

Theorem 2.9. *A Hilbert space H is separable if and only if it admits a countable set S which is orthonormal and complete.*

Proof. If S is countable and complete, then the set

$$\{(\alpha_1 + i\beta_1)s_1 + \cdots + (\alpha_n + i\beta_n)s_n | s_j \in S, \alpha_j, \beta_j \in \mathbb{Q}\}$$

is countable and can be seen to be dense by the completeness of S, hence H is separable. Conversely, if H is separable, then it has a countable dense subset D. By applying the Gram-Schmidt process to this set we can find a countable orthonormal set S such that Span S = Span D, and thus $\overline{\text{Span } S} = H$. □

A useful result concerning orthonormal sets is the following

Lemma 2.10. *Suppose $\{u_j\}_{j=1}^{\infty}$ is an orthonormal set. Then for any $x \in H$:*

$$\sum_{j=1}^{\infty} |(u_j, x)|^2 \leqslant \|x\|^2$$

Proof. For each j pick θ_j such that $(e^{i\theta_j} u_j, x) = |(u_j, x)|$. Consider

$$0 \leqslant \left\| x - \sum_{j=1}^{n} (e^{i\theta_j} u_j, x) e^{i\theta_j} u_j \right\|^2 = \|x\|^2 - 2\sum_{j=1}^{n} |(u_j, x)|^2 + \sum_{j=1}^{n} |(u_j, x)|^2$$

so

$$\sum_{j=1}^{n} |(x, u_j)|^2 \leqslant \|x\|^2$$

and taking the supremum over n we're done. □

Since we know that $L^2(U)$ is separable for U any subset of $\mathbb{R}^n$, we deduce that $L^2(U)$ admits a complete countable orthonormal basis. We give some examples of orthogonal sets of functions.

Example 1. *Consider $L^2([0,1])$ equipped with Lebesgue measure. The set $S = \{e^{-2\pi inx}\}_{n\in\mathbb{Z}}$ is an orthonormal set. By the Stone–Weiserstrass theorem, any function $f \in C^0([0,1])$ can be approximated uniformly by a finite linear combination of elements of S. Since we can approximate any element of $L^2([0,1])$ by a continuous function by Theorem 1.13, we deduce that S is complete. We will see another proof of the completeness of the set S later in the course.*

Example 2. *Let ψ be the function given by:*

$$\psi(x) := \begin{cases} 1 & 0 \leqslant x < \frac{1}{2} \\ -1 & \frac{1}{2} \leqslant x < 1 \\ 0 & \textit{otherwise} \end{cases}$$

and define $\psi_{n,k}$ by

$$\psi_{n,k}(x) := 2^{\frac{n}{2}}\psi(2^n x - k)$$

Then $\{\psi_{n,k}\}_{n,k\in\mathbb{Z}}$ is a complete orthonormal basis for $L^2(\mathbb{R})$ (see Exercise 1.11). This is known as the Haar system. *It is the simplest example of a* wavelet *basis, which give an (imperfect) localisation of a function in both space and frequency. Such bases are widely used in signal processing.*

Example 3. *Consider the space $L^2(\mathbb{R}, e^{-x^2}dx)$, which is a Hilbert space equipped with the Gaussian-weighted inner product*

$$(f,g) = \int_{\mathbb{R}} \overline{f(x)}g(x)e^{-x^2}dx.$$

Applying the Gram-Schmidt process to the linearly independent set $\{1, x, x^2, x^3, \ldots\}$ we can construct a sequence of polynomials $H_k(x)$ of degree k such that

$$\int_{\mathbb{R}} \overline{H_k(x)}H_l(x)e^{-x^2}dx = 0$$

for all $l < k$. For historical reasons, the normalisation is usually chosen such that the coefficient of x^k in $H_k(x)$ is 2^k, but this is purely a convention. The set $\{H_k\}_{k=0}^{\infty}$ is a complete orthogonal set for $L^2(\mathbb{R}, e^{-x^2}dx)$, known as the Hermite *polynomials. We will justify this assertion later in the course.*

Exercise 1.11. Let $S = \{\psi_{n,k}\}_{n,k\in\mathbb{Z}}$ be the Haar system, as defined in lectures.

a) Show that

$$\int_{\mathbb{R}} \psi_{n_1,k_1}(x)\psi_{n_2,k_2}(x)dx = \delta_{n_1n_2}\delta_{k_1k_2}.$$

b) Show that $\mathbb{1}_I \in \overline{\text{Span } S}$ for any finite interval I, where the closure is understood with respect to the L^2 norm.

c) Deduce that S is an orthonormal basis for $L^2(\mathbb{R})$.

2.1.3 The Radon–Nikodym Theorem

An important application of the Riesz representation theorem for Hilbert spaces in the context of measure theory is the Radon–Nikodym theorem, which is important in its own right and will moreover will be valuable when we come to study the dual spaces to the L^p spaces. We first introduce some nomenclature.

Definition 2.2. *Suppose $(E, \mathcal{E})$ is a measurable space, equipped with measures μ, ν. We say ν is* absolutely continuous *with respect to μ, written $\nu \ll \mu$, if for any measurable set A*

$$\mu(A) = 0 \qquad \Longrightarrow \nu(A) = 0.$$

We say μ, ν are mutually singular*, written $\mu \perp \nu$ if there exists a measurable set A such that*

$$\mu(A) = 0 = \nu(A^c)$$

With this definition in hand, we can state

Theorem 2.11 (Radon–Nikodym Theorem)**.** *Suppose $(E, \mathcal{E})$ is a measurable space, with* finite *measures μ, ν such that $\nu \ll \mu$. Then there exists a non-negative $w \in L^1(E, \mu)$ such that*

$$\nu(A) = \int_A w d\mu$$

for any $A \in \mathcal{E}$. In particular this implies

$$\int_E F(x) d\nu(x) = \int_E F(x) w(x) d\mu(x)$$

for any non-negative measurable F.

Proof. Let

$$\alpha = \mu + 2\nu, \qquad \beta = 2\mu + \nu,$$

then α, β are finite measures on $(E, \mathcal{E})$ in an obvious way. On the Hilbert space $H = L^2(E, \alpha) = L^2(E, \mu) \cap L^2(E, \nu)$, we consider the map $\Lambda : H \to \mathbb{C}$ given by:

$$\Lambda(f) = \int_E f d\beta.$$

Noting that $|\Lambda(f)| \leqslant \int_E |f| \, d\beta \leqslant 2 \int_E |f| \, d\alpha \leqslant 2\sqrt{\alpha(E)} \, \|f\|_{L^2(E,\alpha)}$, we see that Λ is bounded on H, and it is manifestly linear, so by Riesz representation theorem (Thm 2.8) there exists $g \in H$ such that for any $f \in H$ we have

$$\int_E f(x) d\beta(x) = \int_E f(x) g(x) d\alpha(x)$$

Rearranging, we deduce

$$\int_E f(2g - 1) d\nu = \int_E f(2 - g) d\mu. \tag{2.2}$$

for all $f \in H$. Taking $f = \mathbb{1}_{A_j}$ for $A_j = \{x \in E : g(x) < \frac{1}{2} - \frac{1}{j}\}$ we deduce $\mu(A_j) = \nu(A_j) = 0$, and so $g \geqslant \frac{1}{2}$ μ-ae and ν-ae. Similarly by considering $A_j = \{x \in E : g(x) > 2 + \frac{1}{j}\}$ we see that $g \leqslant 2$ μ-ae and ν-ae. Thus by redefining g on a set which is null with respect to all measures in the problem, we may assume $\frac{1}{2} \leqslant g(x) \leqslant 2$ for all $x \in E$. By the monotone convergence theorem, we can deduce that (2.2) holds for any non-negative measurable f.

Let $Z = \{g(x) = \frac{1}{2}\}$. Setting $f = \mathbb{1}_Z$ in (2.2) we see that $\mu(Z) = 0$. Since $\nu \ll \mu$ we deduce that $\nu(Z) = 0$, so given a non-negative measurable function F, we can define

$$f(x) = \frac{F(x)}{2g(x) - 1}, \qquad w(x) = \frac{2 - g(x)}{2g(x) - 1}$$

for all $x \in Z^c$ and set $f(x) = w(x) = 0$ otherwise. Applying (2.2) and using $\mu(Z) = \nu(Z) = 0$, we deduce

$$\begin{aligned} \int_E F(x) d\nu(x) &= \int_{E \setminus Z} F(x) d\nu(x) \\ &= \int_E f(2g - 1) d\nu = \int_E f(2 - g) d\mu \\ &= \int_{E \setminus Z} F(x) w(x) d\mu(x) = \int_E F(x) w(x) d\mu(x) \end{aligned}$$

Setting $F(x) = 1$ shows $w \in L^1(E, \mu)$. □

Exercise 1.12. (*) Suppose $(E, \mathcal{E})$ is a measurable space, with finite measures μ, ν. Show that ν may be uniquely written as $\nu = \nu_a + \nu_s$, for measures ν_a, ν_s such that $\nu_s \perp \mu$ and $\nu_a \ll \mu$.
[Hint: Return to the proof of the Radon–Nikodym theorem, but drop the assumption that $\nu \ll \mu$]

2.2 Dual spaces

Given a topological vector space[2] X, we define the dual space X' to be the set of continuous linear functions $\Lambda : X \to \mathbb{C}$. This is a vector space with the obvious operations:

$$(\Lambda_1 + \alpha \Lambda_2)(x) := \Lambda_1(x) + \alpha \Lambda_2(x), \qquad \text{for all } x \in X, \Lambda_1, \Lambda_2 \in X', \alpha \in \mathbb{C}.$$

If X is a normed space, we can equip X' with a norm by setting

$$\|\Lambda\|_{X'} = \sup_{x \in X, \|x\| = 1} |\Lambda(x)|.$$

Often (though not always), we will take X to be a Banach space.

Exercise 2.1. Let X be a normed space. Show that X' equipped with its norm forms a Banach space. If $\overline{X}$ is the completion of X with respect to the metric induced by its norm, show that $X' = \overline{X}'$.

[2] See Definition A.7

Exercise 2.2. Suppose X is a Banach space. Show that if $\Lambda \in X'$ with $\Lambda \neq 0$ then Λ is an open mapping (i.e. $\Lambda(U)$ is open whenever $U \subset X$ is open).

An important fact concerning the dual of a Banach space is that it *separates points*. That is

Lemma 2.12. *Let X be a Banach space. Suppose $x, y \in X$ with $x \neq y$. Then there exists $\Lambda \in X'$ such that $\Lambda(x) \neq \Lambda(y)$.*

We shall not prove this result at this stage, it will follow as a corollary of the Hahn–Banach theorem which we shall prove later.

Suppose X is a Banach space. Given $x \in X$, there is a natural map:

$$\begin{aligned} f_x \quad : \quad & X' \to \mathbb{C} \\ & \Lambda \mapsto \Lambda(x). \end{aligned}$$

This is a bounded linear map, thus belongs to X''. Furthermore, if $f_x(\Lambda) = f_y(\Lambda)$ for all Λ then $x = y$ by Lemma 2.12, thus we have a natural injection of X into X'' given by $x \mapsto f_x$. If this map is surjective, then we say that X is *reflexive*, and write $X = X''$ (by a slight abuse of notation).

2.2.1 The dual of $L^p(\mathbb{R}^n)$

Suppose $f \in L^p(\mathbb{R}^n)$ for some $1 \leqslant p \leqslant \infty$, and let q be such that $p^{-1} + q^{-1} = 1$. Then by Hölder's inequality, we know that if $g \in L^q(\mathbb{R}^n)$ we have:

$$\left| \int_{\mathbb{R}^n} g(x) f(x) dx \right| \leqslant \|g\|_{L^q} \|f\|_{L^p}$$

This tells us that the map $\Lambda_g : L^p(\mathbb{R}^n) \to \mathbb{C}$ given by:

$$\Lambda_g(f) := \int_{\mathbb{R}^n} g(x) f(x) dx$$

is a bounded linear map from $L^p(\mathbb{R}^n)$ to $\mathbb{C}$, thus $\Lambda_g \in L^p(\mathbb{R}^n)'$. Furthermore, it can be shown (Exercise 1.4) that $\|\Lambda_g\|_{L^{p'}} = \|g\|_{L^q}$. Thus the map

$$\begin{aligned} \kappa : \quad L^q(\mathbb{R}^n) & \quad \to \quad L^p(\mathbb{R}^n)' \\ g & \quad \mapsto \quad \Lambda_g \end{aligned}$$

is linear, isometric and injective. We see that $L^q(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)'$ in a natural way.

For the case $p = q = 2$, we know by Riesz representation theorem that in fact[3] $L^2(\mathbb{R}^n) = L^2(\mathbb{R}^n)'$. It is a very natural question to ask whether this happens for other values of p. In fact, it is true for all values of p except one.

Theorem 2.13. *[The dual of $L^p(\mathbb{R}^n)$] Let $1 \leqslant p < \infty$, and let q satisfy $p^{-1} + q^{-1} = 1$. Then $L^q(\mathbb{R}^n) = L^p(\mathbb{R}^n)'$, where we understand elements of $L^q(\mathbb{R}^n)$ as linear maps on $L^p(\mathbb{R}^n)$ according to the map κ described above.*

[3] Being pedantic, one should say that Riesz representation theorem gives an isometric bijection between $L^2(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)'$, but we will leave this as understood.

Note that this result asserts that $L^1(\mathbb{R}^n)' = L^\infty(\mathbb{R}^n)$, however it makes no statement about $L^\infty(\mathbb{R}^n)'$. In fact, $L^\infty(\mathbb{R}^n)' \neq L^1(\mathbb{R}^n)$. We also comment that the result holds more generally for the spaces $L^p(E, \mu)$, where μ is a σ-finite measure on E, but for simplicity we stick to the case of $\mathbb{R}^n$.

The main work of the proof of Theorem 2.13 has already been done in the proof of the Radon–Nikodym theorem, which is the key result we shall require. We first simplify the problem by reducing to the case of positive real linear functionals. Let $L^p(\mathbb{R}^n; \mathbb{R})$ denote the subset of $L^p(\mathbb{R}^n)$ consisting of functions taking values in $\mathbb{R}$ almost everywhere. Clearly $L^p(\mathbb{R}^n; \mathbb{R})$ is a vector space over $\mathbb{R}$, and any element $f \in L^p(\mathbb{R}^n)$ can be written uniquely as $f_r + i f_i$ with $f_r, f_i \in L^p(\mathbb{R}^n; \mathbb{R})$. Given a bounded (complex-)linear map $\Lambda : L^p(\mathbb{R}^n) \to \mathbb{C}$, we can define two bounded (real-)linear maps $\Lambda_r, \Lambda_i : L^p(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ by:

$$\Lambda_r(f) := \Re(\Lambda(f)), \quad \Lambda_i(f) := \Im(\Lambda(f))$$

and we can recover Λ from Λ_r, Λ_i by:

$$\Lambda(f_r + i f_i) = \Lambda_r(f_r) - \Lambda_i(f_i) + i \left(\Lambda_r(f_i) + \Lambda_i(f_r)\right).$$

We say that a real-linear map $u : L^p(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ is *positive* if $u(f) \geqslant 0$ for all $f \geqslant 0$. We claim that any bounded real-linear map $u : L^p(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ may be written as $u = u_+ - u_-$, where $u_\pm : L^p(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ are bounded, positive, real-linear maps (see Exercise 2.3). In view of these facts, in order to prove Theorem 2.13 it will suffice to establish:

Lemma 2.14. *Let $1 \leqslant p < \infty$, $p^{-1} + q^{-1} = 1$. Suppose $u : L^p(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ is a bounded, positive (real-)linear map. Then there exists a non-negative $g \in L^q(\mathbb{R}^n; \mathbb{R})$ with $\|g\|_{L^q} = \|u\|_{(L^p)'}$ such that:*

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

for all $f \in L^p(\mathbb{R}^n; \mathbb{R})$.

Proof. Let $\mu = e^{-|x|^2} dx$ be the Gaussian measure on the Lebesgue sets of $\mathbb{R}^n$, which has the property that $\mu(\mathbb{R}^n) < \infty$. Define, for a measurable set A:

$$\nu(A) := u(e^{-\frac{|x|^2}{p}} \mathbb{1}_A).$$

Clearly $\nu(A) \in [0, \infty]$ and $\nu(\emptyset) = 0$. Further, if $B = \cup_{n=1}^\infty A_n$ for disjoint measurable A_n, setting $B_k = \cup_{n=1}^k A_n$ we have

$$\left\| e^{-\frac{|x|^2}{p}} \mathbb{1}_B - e^{-\frac{|x|^2}{p}} \mathbb{1}_{B_k} \right\|_{L^p} = [\mu(B \setminus B_k)]^{\frac{1}{p}} \to 0$$

so by the continuity and linearity of u, we have $\nu(B_k) \to \nu(B)$. Thus ν defines a measure on the Lebesgue sets of $\mathbb{R}^n$. Further $\nu(\mathbb{R}^n) < \infty$ and moreover $\nu \ll \mu$ since if $\mu(A) = 0$ then

$$\left\| e^{-\frac{|x|^2}{p}} \mathbb{1}_A \right\|_{L^p} = [\mu(A)]^{\frac{1}{p}} = 0.$$

By the Radon–Nikodym theorem, we deduce that there exists a non-negative $G \in L^1(\mathbb{R}^n, \mu)$ such that

$$\nu(A) = \int_A G(x) d\mu = \int_A G(x) e^{-|x|^2} dx.$$

Now by linearity, we deduce that if $f = e^{-\frac{|x|^2}{p}} F$ for some simple function F then we have

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

where $g(x) = e^{-\frac{|x|^2}{q}} G(x)$. Now, functions of the form $e^{-\frac{|x|^2}{p}} F$, with F simple, are dense in $L^p(\mathbb{R}^n; \mathbb{R})$ and moreover we know from the boundedness of u that

$$\int_{\mathbb{R}^n} |f(x) g(x)| \, dx = \int_{\mathbb{R}^n} |f(x)| \, g(x) dx = u(|f|) \leqslant \|u\|_{L^{p'}} \|f\|_{L^p} .$$

This implies that

$$\sup \left\{ \int_{\mathbb{R}^n} |f(x) g(x)| \, dx : f \in L^p(\mathbb{R}^n; \mathbb{R}), \|f\|_{L^p} \leqslant 1 \right\} \leqslant \|u\|_{L^{p'}}$$

By Exercise 1.4 we deduce that $g \in L^q(\mathbb{R}^n; \mathbb{R})$ with $\|g\|_{L^q} \leqslant \|u\|_{L^{p'}}$. On the other hand $\|g\|_{L^q} \geqslant \|u\|_{L^{p'}}$ follows from Hölder and we're done. □

Exercise 2.3. Let $u : L^p(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ be a bounded, linear functional.

a) For $f \in L^p(\mathbb{R}^n; \mathbb{R}), f \geqslant 0$, define

$$\tilde{u}(f) = \sup\{u(g) : g \in L^p(\mathbb{R}^n; \mathbb{R}), 0 \leqslant g \leqslant f\}.$$

Show that $0 \leqslant \tilde{u}(f)$ and $u(f) \leqslant \tilde{u}(f) \leqslant \|u\|_{L^{p'}} \|f\|_{L^p}$, and establish

$$\tilde{u}(f + ag) = \tilde{u}(f) + a\tilde{u}(g)$$

for all $f, g \in L^p(\mathbb{R}^n; \mathbb{R})$ with $f, g \geqslant 0$ and $a \in \mathbb{R}$, $a > 0$.

b) For $f \in L^p(\mathbb{R}^n; \mathbb{R})$, define $w(f) = \tilde{u}(f^+) - \tilde{u}(f^-)$, where $f^+(x) = \max\{0, f(x)\}$, $f^-(x) = \max\{0, -f(x)\}$. Show that w is linear and bounded, and that w and $w - u$ are positive.

c) Deduce that $u = u_+ - u_-$, where $u_\pm$ are bounded, positive, linear functionals.

2.2.2 The Riesz Representation Theorem for spaces of continuous functions

Another space whose dual space can be conveniently described is $C^0_c(\mathbb{R}^n)$, the space of continuous, compactly supported, functions on $\mathbb{R}^n$ equipped with the supremum norm. By a similar reduction to §2.2.1, we can reduce to the problem of understanding positive bounded functionals on $C^0_c(\mathbb{R}^n; \mathbb{R})$.

A classical result known, somewhat confusingly, as the Riesz Representation Theorem shows that any positive functional on $C^0_c(\mathbb{R}^n; \mathbb{R})$ can be represented as integration against a suitable measure. To motivate this result, we first suppose that we are given a σ-algebra $\mathcal{M}$ on $\mathbb{R}^n$ and a measure μ such that $\mu(\mathbb{R}^n) < \infty$. We will also require that the measure space is *regular*:

Definition 2.3. *Suppose that E is a topological space, and $\mathcal{E}$ is a σ-algebra on E which contains the Borel algebra. Then a measure μ defined on $(E, \mathcal{E})$ is regular if for any $A \in \mathcal{E}$, and any $\epsilon > 0$ we can find a closed set C and an open set O such that $C \subset A \subset O$ and:*

$$\mu(O \setminus C) < \epsilon.$$

Since $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{M}$, we know that any $f \in C_c^0(\mathbb{R}^n; \mathbb{R})$ is measurable. The map $\Lambda : C_c^0(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ given by:

$$\Lambda(f) = \int_{\mathbb{R}^n} f(x) d\mu(x) \tag{2.3}$$

is then a positive, bounded linear map. Now, suppose we are given the map Λ, can we recover the measure μ? We note that, if we could set $f = \mathbb{1}_A$ for $A \in \mathcal{M}$, then

$$\text{‘}\Lambda(\mathbb{1}_A) = \int_{\mathbb{R}^n} \mathbb{1}_A(x) d\mu(x) = \mu(A)\text{’},$$

however $\mathbb{1}_A \notin C_c^0(\mathbb{R}^n; \mathbb{R})$. At least for certain sets, however, we can approximate $\mathbb{1}_A$ from below by elements of $C_c^0(\mathbb{R}^n; \mathbb{R})$. Suppose $O \subset \mathbb{R}^n$ is open, and for $k \in \mathbb{N}$ let $O_k = O \cap \{|x| < k\}$. Define:

$$\chi_k(x) := \begin{cases} 1, & x \in O_k,\ d(x, O_k^c) \geqslant k^{-1} \\ kd(x, O_k^c) & x \in O_k,\ d(x, O_k^c) < k^{-1} \\ 0 & x \in O_k^c \end{cases}$$

Then $\chi_k \in C_c^0(\mathbb{R}^n; \mathbb{R})$ and $\chi_k(x)$ increases monotonically to $\mathbb{1}_O$. Thus, by the monotone convergence theorem,

$$\mu(O) = \lim_{k \to \infty} \int_{\mathbb{R}^n} \chi_k(x) d\mu(x) = \lim_{k \to \infty} \Lambda(\chi_k).$$

This shows in particular that

$$\mu(O) = \sup\{\Lambda(g) : g \in C_c^0(\mathbb{R}^n; \mathbb{R}),\ 0 \leqslant g \leqslant \mathbb{1}_O\}. \tag{2.4}$$

Now, since μ is regular, it suffices to know how to compute $\mu(O)$ for open sets in order to find $\mu(A)$ for any $A \in \mathcal{M}$. We have shown:

Lemma 2.15. *Suppose we are given a σ-algebra $\mathcal{M}$ containing $\mathcal{B}(\mathbb{R}^n)$ on $\mathbb{R}^n$ and a regular measure μ such that $\mu(\mathbb{R}^n) < \infty$. Then $\Lambda : C_c^0(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$ given by (2.3) defines a bounded, positive, linear operator. Furthermore, μ is uniquely determined by Λ.*

Riesz Representation Theorem makes the stronger statement that *all* positive bounded linear operators take the form (2.3) for some regular measure.

Theorem 2.16. *Given a positive bounded linear operator $\Lambda : C_c^0(\mathbb{R}^n; \mathbb{R}) \to \mathbb{R}$, there exists a σ-algebra $\mathcal{M}$ on $\mathbb{R}^n$, containing $\mathcal{B}(\mathbb{R}^n)$, and a unique regular measure μ such that $\mu(\mathbb{R}^n) < \infty$ and:*

$$\Lambda(f) = \int_{\mathbb{R}^n} f(x) d\mu(x).$$

We shall not include the proof of this result here, as it is fairly long and technical, and not especially enlightening. The main idea is to use (2.4) to define μ on open sets, and then use Carathéodory's Theorem (or some equivalent approach) to complete this measure. Those interested in the proof will find it in Chapter 2 of Rudin's Real and Complex Analysis.

2.2.3 The strong, weak and weak-$*$ topologies

If X is a Banach space, then X', with the dual norm, is also a Banach space. Thus both spaces are naturally equipped with a topology which makes them topological vector spaces. For certain purposes, however, we may wish to introduce an alternative topology on X or X'. For example, a hugely useful result in the analysis of $\mathbb{R}^n$ is the Bolzano–Weierstrass theorem:

Theorem 2.17. *Let $(x_k)_{k=1}^{\infty}$ with $x_k \in \mathbb{R}^n$ be a bounded sequence. Then $(x_k)_{k=1}^{\infty}$ has a convergent subsequence.*

This result is *not* true when $\mathbb{R}^n$ is replaced by an infinite dimensional Banach space, so the closed unit ball in such a space is not compact. This is quite inconvenient for many problems: for example in the calculus of variations one often wishes to minimise some continuous function defined on a Banach or Hilbert space. Without compactness as a tool, this can be difficult to achieve.

> **Exercise 2.4.** Suppose X is a normed space, and $V \subset X$ is a closed proper subspace of X and let $0 < \alpha < 1$. Show that there exists $x \in X$ with $\|x\| = 1$ such that $\|x - y\| \geqslant \alpha$ for all $y \in V$. Deduce that the Bolzano–Weiserstrass theorem does not hold if X is an infinite dimensional Banach space.
> *[The first result above is known as Riesz' Lemma]*

One way to restore (a version of) compactness for the closed unit ball in X is to consider a different topology defined on X. In order to describe new topologies on X we will make use of seminorms. For full details of this discussion, see §A.2

Definition 2.4. *A seminorm on a vector space X over a field $\Phi = \mathbb{C}$ or $\mathbb{R}$ is a map $p : X \to \mathbb{R}$ satisfying:*

i) $p(x + y) \leqslant p(x) + p(y)$, *for all* $x, y \in X$,

ii) $p(\lambda x) = |\lambda|\, p(x)$ *for all* $x \in X$, $\lambda \in \Phi$,

iii) $p(x) \geqslant 0$ *for all* $x \in X$.

A family $\mathscr{P}$ of seminorms is said to be separating *if for every $x \in X$ with $x \neq 0$, there exists $p \in \mathscr{P}$ with $p(x) \neq 0$.*

Strictly speaking, condition *iii*) follows from *i*) and *ii*) (check this!), but we include it in the definition for convenience. Given a separating family of seminorms, we can construct a topology $\tau_{\mathscr{P}}$ as follows. First, for $p \in \mathscr{P}$, $n \in \mathbb{N}$, we define the set $V(p,n) \subset X$ by

$$V(p,n) := \left\{ x \in X : p(x) < \frac{1}{n} \right\}$$

We let $\dot{\beta}$ be the collection of finite intersections of $V(p,n)$'s, and $\beta = \{x + B : B \in \dot{\beta}\}$.

Theorem 2.18. *Let $\mathscr{P}$ be a separating family of seminorms. The collection of sets β, as described above, is a base for a Hausdorff topology $\tau_{\mathscr{P}}$ on X such that the vector space operations are continuous, and each $p \in \mathscr{P}$ is continuous.*

A topological space $(X, \tau_{\mathscr{P}})$ constructed in the manner above is known as a *locally convex topological vector space.* If $\mathscr{P} = \{p_i\}_{i=1}^{\infty}$ is countable, then the topology is a metric topology, with metric given by:

$$d(x,y) = \sum_{i=1}^{\infty} 2^{-i} \frac{p_i(x-y)}{1 + p_i(x-y)}$$

If this metric is complete, we say that $(X, \tau_{\mathscr{P}})$ is a *Fréchet* space.

> **Exercise 2.5.** Let $\mathscr{P}$ be a separating family of seminorms on a vector space X. Show that a sequence $(x_k)_{k=1}^{\infty}$ with $x_k \in X$ converges to $x \in X$ in the topology $\tau_{\mathscr{P}}$ if and only if $p(x_k - x) \to 0$ for all $p \in \mathscr{P}$.

For a Banach space X, a trivial family of separating seminorms is given by $\mathscr{P} = \{\|\cdot\|\}$. The topology $\tau_s := \tau_{\mathscr{P}}$ induced by this family is simply the usual norm topology. In this context, we sometimes refer to this as the *strong* topology on X. A sequence $(x_k)_{k=1}^{\infty}$ with $x_k \in X$ converges to x in the strong topology if

$$\|x_k - x\| \to 0.$$

An alternative topology on X is given by making use of X' to construct a family of seminorms. It is straightforward to verify that if $\Lambda \in X'$ then $p_\Lambda : x \mapsto |\Lambda(x)|$ is a seminorm. Setting

$$\mathscr{P} := \{p_\Lambda : \Lambda \in X'\}$$

we have a family of seminorms. Moreover, it is separating, since X' separates points of X. Thus $\tau_w := \tau_{\mathscr{P}}$ makes X into a locally convex topological space. This topology is known as the *weak* topology. With respect to the weak topology, the elements of X' are still continuous, however convergence of sequences in the weak topology differs from the strong topology. A sequence $(x_k)_{k=1}^{\infty}$ with $x_k \in X$ converges to x in the weak topology if

$$|\Lambda(x_k - x)| \to 0, \qquad \text{for all } \Lambda \in X'.$$

When $(x_k)_{k=1}^{\infty}$ converges to x in the weak topology, we write $x_k \rightharpoonup x$.

Now, X' is a Banach space itself in a natural fashion, and so has its own associated strong and weak topologies. It also has a further topology, known as the *weak-$*$* topology (pronounced 'weak star'). To define this topology, we note that for each $x \in X$, we can define a seminorm $p_x : X' \to \mathbb{R}$ by $p_x(\Lambda) = |\Lambda(x)|$. Setting

$$\mathscr{P} := \{p_x : x \in X\}$$

we have a family of seminorms, which is separating since if $\Lambda \in X'$, $\Lambda \neq 0$, then there exists $x \in X$ such that $\Lambda(x) \neq 0$. The associated topology we denote τ_{w*}. A sequence $(\Lambda_k)_{k=1}^{\infty}$ with $\Lambda_k \in X'$ converges to Λ in the weak-$*$ topology if:

$$|\Lambda_k(x) - \Lambda(x))| \to 0, \qquad \text{for all } x \in X.$$

When $(\Lambda_k)_{k=1}^{\infty}$ converges to Λ in the weak-$*$ topology, we write $\Lambda_k \overset{*}{\rightharpoonup} \Lambda$. Note that if X is reflexive, then $X'' = X$, and the weak and weak-$*$ topologies coincide.

> **Exercise 2.6.** Suppose that X is a Banach space, and let $(\Lambda_k)_{k=1}^{\infty}$ be a sequence with $\Lambda_k \in X'$. Show that:
>
> $$\Lambda_k \to \Lambda \quad \Longrightarrow \quad \Lambda_k \rightharpoonup \Lambda \quad \Longrightarrow \quad \Lambda_k \overset{*}{\rightharpoonup} \Lambda.$$
>
> (*) Show the stronger statement that $\tau_{w*} \subset \tau_w \subset \tau_s$, where $\tau_{w*}, \tau_w, \tau_s$ are the weak-$*$, weak and strong topologies on X' respectively.

As an example, suppose that $1 \leqslant p < \infty$. Then we know that $L^p(\mathbb{R}^n)' = L^q(\mathbb{R}^n)$ with $p^{-1} + q^{-1} = 1$. If $(f_i)_{i=1}^{\infty}$ is a sequence of functions $f_i \in L^p(\mathbb{R}^n)$, then of course $f_i \to f$ in L^p if:

$$\|f_i - f\|_{L^p} \to 0.$$

On the other hand, $f_i \rightharpoonup f$ in L^p if

$$\int_{\mathbb{R}^n} g(x) f_i(x) dx \to \int_{\mathbb{R}^n} g(x) f(x) dx, \qquad \text{for all } g \in L^q(\mathbb{R}^n).$$

If $1 < p < \infty$, $L^p(\mathbb{R}^n)$ is reflexive, so the weak-$*$ topology that arises from viewing $L^p(\mathbb{R}^n)$ as the dual of $L^q(\mathbb{R}^n)$ agrees with the weak topology. We have not identified any space X such that $X' = L^1(\mathbb{R}^n)$, so no weak-$*$ topology on L^1 is available to us. Since $L^1(\mathbb{R}^n)' = L^{\infty}(\mathbb{R}^n)$, we can consider the weak-$*$ topology on $L^{\infty}(\mathbb{R}^n)$. We have $f_i \overset{*}{\rightharpoonup} f$ in L^{∞} if:

$$\int_{\mathbb{R}^n} g(x) f_i(x) dx \to \int_{\mathbb{R}^n} g(x) f(x) dx, \qquad \text{for all } g \in L^1(\mathbb{R}^n).$$

Since we don't have a concrete realisation of $L^{\infty}(\mathbb{R}^n)'$, we don't have a simple description of weak convergence in this space (other than the abstract condition $\Lambda(f_i) \to \Lambda(f)$ for all $\Lambda \in L^{\infty}(\mathbb{R}^n)'$).

> **Exercise 2.7.** For a bounded measurable set $E \subset \mathbb{R}^n$ of positive measure, and any $f \in L^1_{loc.}(\mathbb{R}^n)$, define the mean of f on E to be:
>
> $$\fint_E f(x) dx = \frac{1}{|E|} \int_E f(x) dx.$$

Suppose $1 < p < \infty$ and let $(f_j)_{j=1}^\infty$ be a bounded sequence in $L^p(\mathbb{R}^n)$. Show that $f_j \rightharpoonup f$ for some $f \in L^p(\mathbb{R}^n)$ if and only if

$$\fint_E f_j(x)dx \to \fint_E f(x)dx$$

for all bounded measurable sets $E \subset \mathbb{R}^n$ of positive measure.

Exercise 2.8. Suppose $(H, (\cdot,\cdot))$ is an infinite dimensional Hilbert space and let $(x_i)_{i=1}^\infty$ be a sequence with $x_i \in H$.

i) Show that $x_i \rightharpoonup x$ if and only if $(y, x_i) \to (y, x)$ for all $y \in H$.

ii) Show there exists a sequence such that $x_i \rightharpoonup 0$, but $x_i \not\to 0$.

iii) Suppose $x_i \rightharpoonup x$. Show that

$$\|x\| \leqslant \liminf_{i\to\infty} \|x_i\|,$$

and $\|x_i\| \to \|x\|$ iff $x_i \to x$.

2.2.4 Compactness, Banach–Alaoglu

As we have discussed above, if X is an infinite dimensional Banach space, then the closed unit ball $\overline{B} = \{x : \|x\| \leqslant 1\}$ is not compact. This is unfortunate, and we would like to try and restore compactness in some way. There are essentially two (related) approaches: we can either restrict our attention to a subset of $\overline{B}$ for which we have compactness, or else we can weaken the topology on $\overline{B}$.

To explain this, let us recall the Arzelà–Ascoli theorem. We set $I = [0, 1]$

Theorem 2.19. *Suppose $(f_k)_{k=1}^\infty$ is a sequence of continuous functions $f_k : I \to \mathbb{C}$ which is bounded, i.e. for all k:*

$$\sup_{x\in I} |f_k(x)| \leqslant M$$

and equicontinuous: for all $\epsilon > 0$, there exists $\delta > 0$ such that for all k and all $|x - y| < \delta$ we have

$$|f_k(x) - f_k(y)| < \epsilon.$$

Then $(f_k)_{k=1}^\infty$ admits a uniformly convergent subsequence.

To put this into the language we have been discussing, recall that for $0 < \gamma \leq 1$, we say a continuous function $f : I \to \mathbb{C}$ is γ-Hölder continuous if

$$\|f\|_{C^{0,\gamma}} := \sup_{x\in I} |f(x)| + \sup_{x,y\in I, x\neq y} \frac{|f(x) - f(y)|}{|x - y|^\gamma} < \infty.$$

The set $C^{0,\gamma}(I)$, of γ-Hölder continuous functions, is a Banach space with this norm, and $C^{0,\gamma}(I) \subset C^0(I)$. A consequence of the Arzelà–Ascoli theorem is the following:

Corollary 2.20. *The closed unit ball in $C^{0,\gamma}(I)$ is compact in the C^0-topology.*

Proof. Since the C^0-topology is a metric topology, compactness is equivalent to sequential compactness. If $(f_k)_{k=1}^\infty$ is a sequence with $f_k \in C^{0,\gamma}(I)$ satisfying

$$\|f_k\|_{C^{0,\gamma}} \leqslant 1,$$

then $(f_k)_{k=1}^\infty$ is a bounded equicontinuous sequence of functions, so admits a uniformly convergent subsequence, i.e. a subsequence which converges in the C^0-topology to some f. It is a short exercise to check that $f \in C^{0,\gamma}$ with $\|f\|_{C^{0,\gamma}} \leqslant 1$. □

This gives us a paradigmatic example of a compactness result for Banach spaces: the closed unit ball is compact, but only in a weaker topology than the strong topology. The central result is the Banach–Alaoglu theorem:

Theorem 2.21. *Let X be a normed space, and let $\overline{B}' = \{\Lambda \in X' : \|\Lambda\|_{X'} \leqslant 1\}$ be the closed unit ball in X'. Then $\overline{B}'$ is compact in the weak-$*$ topology on X'.*

This result in its full generality is typically proven using Tychonoff's theorem, which relies on (a version of) the axiom of choice. We will content ourselves with the proof in the case where X is a separable Banach space for which it is possible to give a constructive proof. The majority of the applications of Banach–Alaoglu that arise in (for example) the calculus of variations or PDE are covered by this special case.

We first note that if X is a separable Banach space, then the weak-$*$ topology on $\overline{B}'$ is in fact a metric topology.

Lemma 2.22. *Let X be a separable Banach space, with a countable dense subset $D = \{x_k\}_{k=1}^\infty$. Let $\tilde{\mathscr{P}} = \{p_k\}_{k=1}^\infty$ be the family of seminorms on X' defined by:*

$$p_k : \Lambda \mapsto |\Lambda(x_k)|,$$

and let $\tau_{\tilde{\mathscr{P}}}$ be the associated topology. Then $\tau_{\tilde{\mathscr{P}}}|_{\overline{B}'} = \tau_{w*}|_{\overline{B}'}$. In particular, the weak-$*$ topology on $\overline{B}'$ is a metric topology.

Proof. From the definition of the topologies, it is immediate that every open set in $\tau_{\tilde{\mathscr{P}}}$ is open in τ_{w*}, thus $\tau_{\tilde{\mathscr{P}}}|_{\overline{B}'} \subset \tau_{w*}|_{\overline{B}'}$. For any $x \in X$, $n \in \mathbb{N}$ let

$$V(x,n) := \left\{\Lambda \in \overline{B}' : |\Lambda(x)| < \frac{1}{n}\right\}.$$

In order to show $\tau_{\tilde{\mathscr{P}}}|_{\overline{B}'} \supset \tau_{w*}|_{\overline{B}'}$ it suffices to show that for any $x \in X$, $n \in \mathbb{N}$, we can find $x_k \in D$, $m \in \mathbb{N}$ such that

$$V(x_k, m) \subset V(x,n).$$

Fix $x \in X$. For any $\epsilon > 0$, there exists $x_i \in D$ such that $\|x - x_i\| < \epsilon$. If $\Lambda \in V(x_i, m)$, then:

$$|\Lambda(x)| = |\Lambda(x - x_i) + \Lambda(x_i)| \leqslant \|\Lambda\| \, \|x - x_i\| + |\Lambda(x_i)| < \epsilon + \frac{1}{m}.$$

Taking $\epsilon < 1/(2n)$, $m > 2n$ we have $\Lambda \in V(x,n)$ and we're done. □

This result is useful in two ways. Firstly, we see that a sequence $(\Lambda_j)_{j=1}^{\infty}$ with $\Lambda_j \in \overline{B}'$ converges to Λ in the weak-$*$ topology if and only if:

$$\Lambda_j(x_k) \to \Lambda(x_k) \text{ as } j \to \infty, \text{ for all } k.$$

Secondly, since the weak-$*$ topology on $\overline{B}'$ is metric, compactness is equivalent to sequential compactness. We can establish sequential compactness with a very similar method to the proof of the Arzelà–Ascoli theorem.

Theorem 2.23. *Let X be a separable Banach space. Let $(\Lambda_j)_{j=1}^{\infty}$ be a sequence with $\Lambda_j \in \overline{B}'$. Then there exists a subsequence $(\Lambda_{j_k})_{k=1}^{\infty}$ and $\Lambda \in \overline{B}'$ such that $\Lambda_j \overset{*}{\rightharpoonup} \Lambda$.*

Proof. Let $D = \{x_k\}_{k=1}^{\infty}$ be a countable dense subset. Consider the sequence $(\Lambda_j(x_1))_{j=1}^{\infty}$. This is a uniformly bounded sequence of complex numbers, since:

$$|\Lambda_j(x_1)| \leqslant \|\Lambda_j\| \, \|x_1\| \leqslant \|x_1\| \, .$$

Thus, by Bolzano–Weierstrass, there exists a subsequence $(\Lambda_{j_k}(x_1))_{k=1}^{\infty}$ and a number $\Lambda(x_1) \in \mathbb{C}$ with $|\Lambda(x_1)| \leqslant \|x_1\|$ such that:

$$\Lambda_{j_k}(x_1) \to \Lambda(x_1).$$

We write $\Lambda_{1,k} := \Lambda_{j_k}$, then $(\Lambda_{1,j})_{j=1}^{\infty}$ is a subsequence of $(\Lambda_j)_{j=1}^{\infty}$. By a similar argument, we can find a subsequence $(\Lambda_{1,j_k})_{j=1}^{\infty}$ of $(\Lambda_{1,j})_{j=1}^{\infty}$ such that $\Lambda_{1,j_k}(x_2) \to \Lambda(x_2)$. We write $\Lambda_{2,k} := \Lambda_{1,j_k}$. Continuing in this fashion, we construct for each $l \geqslant 1$ a sequence $(\Lambda_{l,j})_{j=1}^{\infty}$, and a complex number $\Lambda(x_l)$ with $|\Lambda(x_l)| \leqslant \|x_l\|$ with the property that $(\Lambda_{l,j})_{j=1}^{\infty}$ is a subsequence of $(\Lambda_{l-1,j})_{j=1}^{\infty}$, and $\Lambda_{l,j}(x_k) \to \Lambda(x_k)$ as $j \to \infty$ for all $l \leqslant k$.

Now, consider $(\Lambda_{j,j})_{j=1}^{\infty}$. This is a subsequence of $(\Lambda_j)_{j=1}^{\infty}$ with the property that for each $x \in D$ we have:

$$\Lambda_{j,j}(x) \to \Lambda(x).$$

If we can show that there exists $\tilde{\Lambda} \in \overline{B}'$ with $\tilde{\Lambda}(x) = \Lambda(x)$ for all $x \in D$, then we are done. We first claim that $\Lambda : D \to \mathbb{C}$ is uniformly continuous. Fix $\epsilon > 0$, and suppose $x, y \in D$ with $\|x - y\| < \frac{\epsilon}{3}$. Since $\Lambda_{j,j}(x) \to \Lambda(x)$ and $\Lambda_{j,j}(y) \to y$, there exists k such that for all $j \geqslant k$ we have $|\Lambda_{j,j}(x) - \Lambda(x)| < \frac{\epsilon}{3}$ and $|\Lambda_{j,j}(y) - \Lambda(y)| < \frac{\epsilon}{3}$. For such a j we estimate:

$$|\Lambda(x) - \Lambda(y)| \leqslant |\Lambda(x) - \Lambda_{j,j}(x)| + |\Lambda_{j,j}(y) - \Lambda(y)| + |\Lambda_{j,j}(x - y)| < \epsilon.$$

we conclude that $\Lambda : D \to \mathbb{C}$ is continuous. Thus Λ extends to a continuous function $\tilde{\Lambda} : X \to \mathbb{C}$. We abuse notation and drop the tilde at this point. Next, we claim Λ is linear. Suppose $x, y \in X$ and $a \in \mathbb{C}$ and for $z = x + ay$ estimate:

$$\begin{aligned} |\Lambda(z) - \Lambda(x) - a\Lambda(y)| \leqslant & \left|\Lambda(z) - \Lambda(z')\right| + \left|\Lambda(x) - \Lambda(x')\right| + |a| \left|\Lambda(y) - \Lambda(y')\right| \\ & + \left|\Lambda(z') - \Lambda_{j,j}(z')\right| + \left|\Lambda(x') - \Lambda_{j,j}(x')\right| + |a| \left|\Lambda(y') - \Lambda_{j,j}(y')\right| \\ & + \left|\Lambda_{j,j}(z' - x' - ay')\right| . \end{aligned}$$

By choosing $x', y', z' \in D$ sufficiently close to x, y, z respectively we may arrange that the first and final line are arbitrarily small. Taking j sufficiently large we see that the middle line can also be made arbitrarily small. We conclude that $\Lambda(x + ay) = \Lambda(x) + a\Lambda(y)$. Thus $\Lambda : X \to \mathbb{C}$ is a continuous linear map. Finally, since D is dense in X we have:

$$\|\Lambda\| = \sup_{x\in X, \|x\|\leqslant 1} |\Lambda(x)| = \sup_{x\in D, \|x\|\leqslant 1} |\Lambda(x)| \leqslant 1.$$

Hence $\Lambda \in \overline{B}'$. □

As a corollary, we find the following compactness result for the Lebesgue spaces:

Corollary 2.24. *Suppose $1 < p \leqslant \infty$, and let $(f_j)_{j=1}^\infty$ be a sequence of functions $f_j \in L^p(\mathbb{R}^n)$ satisfying*

$$\|f_j\|_{L^p} \leqslant K.$$

Then there exists $f \in L^p(\mathbb{R}^n)$ and a subsequence $(f_{j_k})_{k=1}^\infty$ such that $\|f\|_{L^p} \leqslant K$ and

$$\int_{\mathbb{R}^n} g(x) f_{j_k}(x) dx \to \int_{\mathbb{R}^n} g(x) f(x) dx$$

for all $g \in L^q(\mathbb{R}^n)$, where $p^{-1} + q^{-1} = 1$.

Proof. Apply the previous result to f_j/K. □

Note that we do not have a corresponding compactness result for $L^1(\mathbb{R}^n)$. It is possible to gain some compactness by considering $L^1(\mathbb{R}^n)$ as a subspace of the dual space of $C_c^0(\mathbb{R}^n)$, however the limiting objects constructed in this way are typically measures, not elements of $L^1(\mathbb{R}^n)$.

Exercise 2.9. Construct a bounded sequence $(f_i)_{i=1}^\infty$ of functions $f_i \in L^1(\mathbb{R})$ such that no subsequence is weakly convergent.

2.3 Hahn–Banach

The next result we shall cover is the Hahn–Banach theorem. This modest-seeming result permits us to extend a bounded linear functional defined on a subspace, M, of a vector space X into a bounded linear functional defined on the whole space. While it seems like this should be straightforward, in full generality it requires the axiom of choice (or at least some method of transfinite induction). We will proceed modestly by first showing that we can extend a linear functional in one direction.

We first show that we can reduce to the case of a real vector space. Suppose X is a complex vector space. We note that X is also a real vector space in a natural fashion. A real-linear map $\ell : X \to \mathbb{R}$ is a map satisfying:

$$\ell(x + ay) = \ell(x) + a\ell(y), \qquad \text{for all } x, y \in X, a \in \mathbb{R}.$$

Suppose $\Lambda : X \to \mathbb{C}$ is a complex-linear map, then $\ell(x) = \Re(\Lambda(x))$ defines a real-linear map on X. Conversely, given a real-linear map $\ell : X \to \mathbb{R}$, we have that:

$$\Lambda(x) = \ell(x) - i\ell(ix)$$

defines a complex-linear map $\Lambda : X \to \mathbb{C}$, with $\Re(\Lambda(x)) = \ell(x)$. Thus provided we can establish a result which allows us to extend linear functionals in a bounded fashion on a real vector space, we immediately have a corresponding result for the complex setting.

At this point we should address our use of the word 'bounded' above. When X is a Banach space, then we have already discussed what it means for a functional to be bounded. It turns out to be useful to consider a slightly more general notion of boundedness at this stage, however. For this we introduce

Definition 2.5. *A sublinear functional on a real vector space X is a map $p : X \to \mathbb{R}$ satisfying*

$$p(x+y) \leqslant p(x) + p(y), \qquad p(tx) = tp(x),$$

for any $x, y \in X$ and $t \geq 0$.

For example if $\Lambda : X \to \mathbb{R}$ linear, then $p(x) = |\Lambda(x)|$ is sublinear. Any semi-norm (hence any norm) is sublinear, but the converse doesn't hold, so be careful!

We first show that we can extend a linear functional ℓ defined on a subspace in one direction, maintaining a bound by a sublinear functional p. We will work with one-sided bounds of the form $\ell(x) \leqslant p(x)$, but we note that this implies the two-sided bound $-p(-x) \leqslant \ell(x) \leqslant p(x)$.

Lemma 2.25. *Let X be a real vector space, $p : X \to \mathbb{R}$ sublinear and $M \subset X$ a subspace. Suppose $\ell : M \to \mathbb{R}$ is linear and satisfies $\ell(y) \leqslant p(y)$ for all $y \in M$. Fix $x \in X \setminus M$, then setting $\tilde{M} = span\ \{M, x\}$, there exists a linear operator $\tilde{\ell} : \tilde{M} \to \mathbb{R}$ such that*

$$\tilde{\ell}(z) \leqslant p(z) \quad \text{for all } z \in \tilde{M}.$$

and

$$\ell(y) = \tilde{\ell}(y), \quad \text{for all } y \in M.$$

Proof. Any $z \in \tilde{M}$ can be uniquely written as $z = \lambda x + y$ for $y \in M$, so to define the extension $\tilde{\ell}$, by linearity it suffices to specify $\tilde{\ell}(x) = a$ as then $\tilde{\ell}(\lambda x + y) = \lambda a + \ell(y)$. Suppose $y, z \in M$, then

$$\ell(y) + \ell(z) = \ell(y+z) \leqslant p(y+z) \leqslant p(y-x) + p(z+x)$$

and hence

$$\ell(y) - p(y-x) \leqslant p(z+x) - \ell(z). \tag{2.5}$$

Let

$$a = \sup_{y \in M} \left(\ell(y) - p(y-x)\right).$$

This is well defined by (2.5) and further we deduce

$$\ell(y) - a \leqslant p(y-x), \qquad \ell(z) + a \leqslant p(z+x)$$

for all $y, z \in M$. If $\lambda > 0$, replace z with $\lambda^{-1}y$ and multiply by λ. If $\lambda < 0$ replace y with $-\lambda^{-1}y$ and multiply by $-\lambda$ to deduce

$$\ell(y) + a\lambda \leqslant p(y + \lambda x),$$

holds for all $y \in M, \lambda \in \mathbb{R}$. □

Now, if dim X/M is finite, this result allows us to iteratively extend $\ell : M \to \mathbb{R}$ to the whole space in a finite number of steps. If X is infinite, but *separable*, then it's possible to construct an extension inductively (try it!). If, however, X is not infinite, then (speaking loosely) it's not possible to exhaust X with a countable number of finite extensions. We require some way to make an inductive type argument in a non-countable setting. There are several approaches to this, all of which require the axiom of choice. We shall use Zorn's Lemma[4]. For this we require some background.

> **Exercise($*$).** Suppose X is a *separable* real Banach space. Prove the Hahn–Banach theorem on X *without* invoking the axiom of choice through Zorn's Lemma (or equivalent).

2.3.1 Zorn's Lemma

Zorn's Lemma is a statement concerning *partial orderings* of a set S.

Definition 2.6. *Let S be a set. Then a* partial order *on S is a binary relation $\leq$ satisfying, for any $a, b, c \in S$:*

i) $a \leq a$ for all $a \in S$. (Reflexivity)

ii) If $a \leq b$ and $b \leq a$, then $a = b$. (Antisymmetry)

iii) If $a \leq b$ and $b \leq c$, then $a \leq c$. (Transitivity)

A set with a partial order is called a partially ordered set, or *poset*. Note that we do not assert that for any a, b either $a \leq b$ or $b \leq a$. If this does hold, we say $\leq$ is a *total order*.

A subset T of a partially ordered set which is totally ordered is called a *chain*. An element $u \in S$ is an *upper bound* for $T \subset S$ if $a \leq u$ for all $a \in T$. A *maximal element* of S is an element $m \in S$ such that $m \leq x$ implies $x = m$.

Example 4. *a) If S is any set, then the power set 2^S is a poset, with $\leq$ given by inclusion, i.e. $A \leq B$ iff $A \subset B$. S is a maximal element.*

b) The real numbers with their usual order is a totally ordered set, with no maximal element.

c) The collection S of open balls in $\mathbb{R}^n$ is a poset with order given by inclusion. The subset

$$T = \{B_r(0) \subset \mathbb{R}^n : 0 < r \leqslant 1\}$$

is a chain. $B_1(0)$ is a maximal element, and $B_2(0)$ is an upper bound.

Zorn's Lemma can now be stated as:

[4] For a good discussion of how and why Zorn's Lemma is useful, see Prof. Gower's blog: https://gowers.wordpress.com/2008/08/12/how-to-use-zorns-lemma/

Proposition 1. *Let $(S, \leq)$ be a partially ordered set in which every chain has an upper bound. Then $(S, \leq)$ contains at least one maximal element.*

We shall not prove this claim. In fact, it is equivalent (within the Zermelo-Fraenkel framework) to the axiom of choice, so we can reasonably treat Zorn's Lemma as an axiom itself. The proof of the Hahn–Banach theorem we give below is a typical application of Zorn's Lemma.

Theorem 2.26 (Hahn–Banach)**.** *Let X be a real vector space, $p : X \to \mathbb{R}$ sublinear, and $M \subset X$ a subspace. Suppose $\ell : M \to \mathbb{R}$ is linear and satisfies $\ell(y) \leqslant p(y)$ for all $y \in M$. Then there exists a linear operator $\tilde{\ell} : X \to \mathbb{R}$ such that*

$$\tilde{\ell}(z) \leqslant p(z) \quad \text{for all } z \in X.$$

and

$$\ell(y) = \tilde{\ell}(y), \quad \text{for all } y \in M.$$

Proof. We consider the set S of extensions of ℓ to a linear subspace of X. That is a pair $(N, \ell^*) \in S$ if:

i) N is a linear subspace of X containing M.

ii) $\ell^* : N \to \mathbb{R}$ is a linear map.

iii) $\ell^*(x) \leqslant p(x)$ for all $x \in N$

iv) $\ell^*(y) = \ell(y)$ for all $y \in M$

S is a poset, with the partial ordering given by $(N_1, \ell_1) \leq (N_2, \ell_2)$ if N_1 is a subspace of N_2 and $\ell_1(x) = \ell_2(x)$ for all $x \in N_1$. Suppose that T is a totally ordered subset of S. We define $(\mathcal{N}, L) \in S$ by:

$$\mathcal{N} = \bigcup_{(N,\ell^*) \in T} N,$$

and for any $x \in \mathcal{N}$, we define $L(x) = \ell^*(x)$, where $(N, \ell^*) \in T$ with $x \in N$. This is well defined since T is totally ordered, and moreover we have $(N, \ell^*) \leq (\mathcal{N}, L)$ for all $(N, \ell^*) \in T$, thus T has an upper bound. By Zorn's Lemma, S has a maximal element, $(\mathscr{N}, \tilde{\ell})$. We claim that $\mathscr{N} = X$. Suppose not, then there exists $x \in X \setminus \mathscr{N}$ and we can extend $\tilde{\ell}$ to a functional $\tilde{\ell}^*$ on $\mathscr{N}^* = \text{span}\,\{\mathscr{N}, x\}$ by Lemma 2.25. Then $(\mathscr{N}, \tilde{\ell}) \leq (\mathscr{N}^*, \tilde{\ell}^*)$, but $(\mathscr{N}, \tilde{\ell}) \neq (\mathscr{N}^*, \tilde{\ell}^*)$, contradicting the maximality of $(\mathscr{N}, \tilde{\ell})$. Thus $\tilde{\ell} : X \to \mathbb{R}$ is the extension we seek. □

Notice that this proof of the Hahn–Banach theorem is *non-constructive*: while we assert the existence of at least one extension, the proof provides no mechanism to construct a particular example. This is typical of proofs which invoke the axiom of choice through Zorn's Lemma (or otherwise).

Corollary 2.27. *Let X be a Banach space over Φ, where $\Phi = \mathbb{R}$, or $\mathbb{C}$, and $M \subset X$ be a subspace. Let $\Lambda : M \to \Phi$ be a bounded linear operator. Then there exists a bounded linear operator $\tilde{\Lambda} : X \to \Phi$ with $\|\Lambda\|_{M'} = \|\tilde{\Lambda}\|_{X'}$ such that $\Lambda(y) = \tilde{\Lambda}(y)$ for all $y \in M$.*

Proof. If $\Phi = \mathbb{R}$, then we may apply the previous result with $p(x) = \|\Lambda\|\|x\|$.

If $\Phi = \mathbb{C}$, then recall that we may write $\Lambda(x) = \ell(x) - i\ell(ix)$ for a real-linear map $\ell(x) = \Re(\Lambda(x))$. Further, by noting that $|\Lambda(x)| = \ell(e^{i\theta}x)$ for suitable θ, we can see that for any subspace $N \subset X$

$$\sup_{x\in N, \|x\|\leqslant 1} |\Lambda(x)| = \sup_{x\in N, \|x\|\leqslant 1} |\ell(x)|,$$

and we may apply the $\Phi = \mathbb{R}$ result to ℓ. □

We will now establish some more geometric consequences of the Hahn–Banach theorem that go by the name of separation theorems. The first of these is related to the hyperplane separation theorem, which states that given two disjoint convex sets in $\mathbb{R}^n$ we may find a co-dimension one plane such that the sets are on opposite sides of the plane. The theorem (as with many of our results on Banach spaces) can be generalised to other topological vector spaces. Those interested in more general statements may wish to consult Rudin's "Functional Analysis".

Theorem 2.28. *Suppose A and B are disjoint, nonempty, convex sets in a real or complex Banach space X.*

a) If A is open, there exist $\Lambda \in X'$ and $\gamma \in \mathbb{R}$ such that

$$\Re(\Lambda x) < \gamma \leqslant \Re(\Lambda y) \tag{2.6}$$

for all $x \in A$, $y \in B$. If B is further assumed to be open the second inequality may be taken to be strict.

b) If A is compact, B is closed then there exist $\Lambda \in X'$ and $\gamma_1, \gamma_2 \in \mathbb{R}$ such that

$$\Re(\Lambda x) < \gamma_1 < \gamma_2 < \Re(\Lambda y).$$

Proof. We first observe that it suffices to establish the result for real scalars. If we have done so then for X a complex Banach space we may find a real-linear $\ell : X \to \mathbb{R}$ which separates A and B as required, and we may then set $\Lambda(x) = \ell(x) - i\ell(ix)$.

a) Pick $a_0 \in A, b_0 \in B$ and let $x_0 = b_0 - a_0$. Let $C = A - B + x_0$. This is a convex neighbourhood of 0 in X, and since A, B are disjoint $x_0 \notin C$. Let $p(x) = \inf\{t > 0 : t^{-1}x \in C\}$. By Exercise 2.10 this is a sublinear function satisfying $p(x) \leqslant k\|x\|$ for some $k > 0$ and $p(y) < 1$ for $y \in C$. Since $x_0 \notin C$ we have $p(x_0) \geqslant 1$.

   Let M be the subspace generated by x_0 and on this space define the linear functional $f(tx_0) = t$. If $t \geqslant 0$ then $f(tx_0) = t \leqslant tp(x_0) = p(tx_0)$, while if $t < 0$ then $f(tx_0) = t < 0 \leqslant p(tx_0)$, so $f \leqslant p$ on M. By the Hahn–Banach theorem we can extend f to a linear functional Λ satisfying $-k\|x\| \leqslant -p(-x) \leqslant \Lambda(x) \leqslant p(x) \leqslant k\|x\|$, so $\Lambda \in X'$.

   Now suppose $a \in A, b \in B$. Then

$$\Lambda a - \Lambda b + 1 = \Lambda(a - b + x_0) \leqslant p(a - b + x_0) < 1$$

since $a - b + x_0 \in C$. Thus $\Lambda a < \Lambda b$. This implies that $\Lambda(A)$ and $\Lambda(B)$ are disjoint convex subsets of $\mathbb{R}$. $\Lambda(A)$ is open since every non-constant linear functional is an open mapping by Exercise 2.2. We may take γ to be the right end-point of $\Lambda(A)$ and (2.6) follows. If B is open, then so is $\Lambda(B)$ and we can replace the $\leqslant$ in (2.6) with $<$.

b) Since now A is compact and B is closed, we have $\inf\{\|a - b\| : a \in A, b \in B\} = d > 0$. Let $V = B_{\frac{d}{2}}(0)$ and consider $A + V$. This is open, convex and disjoint from B. Applying part *a*) with $A + V$ in place of A shows there exists $\Lambda \in X'$ such that $\Lambda(A + V)$ and $\Lambda(B)$ are disjoint convex subsets of $\mathbb{R}$ with $\Lambda(A + V)$ to the left of $\Lambda(B)$. Since $\Lambda(A + V)$ is open and $\Lambda(A)$ is a compact subset of $\Lambda(A + V)$, the result follows. □

This result immediately gives the proof of Lemma 2.12, which states that X' separates points in X: we set $A = \{x\}$, $B = \{y\}$ and apply part *b*). Another consequence is

Corollary 2.29. *Suppose M is a subspace of a Banach space X and $x_0 \in X$. If x_0 is not in the closure of M then there exists $\Lambda \in X'$ such that $\Lambda x_0 = 1$ and $\Lambda x = 0$ for every $x \in M$.*

Proof. Applying part *b*) of the previous Theorem with $A = \{x_0\}$ and $B = \overline{M}$, there exists $\Lambda \in X'$ such that Λx_0 and $\Lambda(M)$ are disjoint. But $\Lambda(M)$ must be a proper subspace of the scalar field, so must be $\{0\}$. The desired functional can be obtained by dividing Λ by Λx_0. □

Exercise 2.10. Let X be a Banach space and suppose $A \subset X$ is a convex neighbourhood of 0. For $x \in X$ define $\mu_A(x) = \inf\{t > 0 : t^{-1}x \in A\}$. Show that μ_A is sublinear and satisfies $\mu_A(x) \leqslant k\|x\|$ for some $k > 0$. Show further that $\mu_A(y) < 1$ for $y \in A$.
μ_A is called the Minkowski functional of A

Exercise 2.11. Let $\{x_1, \dots x_n\}$ be a set of linearly independent elements of a Banach space X. Let $a_1, \dots, a_n \in \mathbb{C}$. Show that there exists $\Lambda \in X'$ such that $\Lambda(x_i) = a_i$, for $i = 1, \dots, n$.

Exercise 2.12. Let M be a vector subspace of the Banach space X, and suppose that $K \subset X$ is open, convex and disjoint from M. Show that there exists a co-dimension one subspace $N \subset X$ which contains M and is disjoint from K.
This is Mazur's theorem.

Exercise 2.13. let X be a reflexive Banach space, and suppose $Y \subset X$ is a closed subspace. Show that Y is reflexive.

Chapter 3

Test functions and distributions

3.1 The space $\mathscr{D}(\Omega)$

Given an open set Ω, we are familiar with $C^\infty(\Omega)$ and $C_c^\infty(\Omega)$ as sets and we can equip them with the algebraic structure of a vector space. We want to discuss notions of convergence and continuity in these spaces, and for this we shall require a topology. The topologies we require are locally convex, so can in principle be described by a family of semi-norms, however in the case of $C_c^\infty(\Omega)$, it turns out to be quite subtle to do this. Appendix A develops the topology in detail, for those who are interested. We shall simply quote the following result:

Theorem 3.1. *The set $C_c^\infty(\Omega)$ can be endowed with a topology τ, such that:*

i) The vector space operations of addition and scalar multiplication are continuous with respect to τ.

ii) A sequence $\{\phi_j\}_{j=1}^\infty \subset C_c^\infty(\Omega)$ tends to zero with respect to the topology τ if there exists a compact $K \subset \Omega$ such that $\operatorname{supp} \phi_j \subset K$ for all $j \in \mathbb{N}$ and for each multi-index α we have:

$$\sup_{x \in K} |D^\alpha \phi_j| \to 0,$$

as $j \to \infty$. Similarly, $\phi_j \to \phi$ with respect to τ if $\phi_j - \phi \to 0$.

We denote the set $C_c^\infty(\Omega)$ equipped with the topology τ by $\mathscr{D}(\Omega)$.

This topology is not a metric topology, so the description of the convergent sequences is not the whole story, but for the purposes that we require, it will suffice. In particular, for linear maps from $\mathscr{D}(\Omega)$ into a locally convex vector space sequential continuity is equivalent to continuity.

Example 5. *Suppose $\phi \in \mathscr{D}(\Omega)$. Let δ be such that $\tau_x \phi \in \mathscr{D}(\Omega)$ for $|x| < \delta$. If $\{x_l\}_{l=1}^\infty \subset \mathbb{R}^n$ is a sequence with $|x| < \delta$, and $x_l \to 0$, then*

$$\tau_{x_l} \phi \to \phi, \qquad \text{as } l \to \infty.$$

To see why this is so, recall that there exists $\epsilon > 0$ *such that* $\operatorname{supp}\phi + B_{2\epsilon}(0) \subset \Omega$. *Suppose that* $|x| < \epsilon$. *Then*

$$\operatorname{supp}\tau_x\phi = \operatorname{supp}\phi + x \subset \overline{\operatorname{supp}\phi + B_\epsilon(0)} \subset \operatorname{supp}\phi + B_{2\epsilon}(0) \subset \Omega.$$

Thus for i *large enough,* $\operatorname{supp}\tau_{x_l}\phi \subset K := \overline{\operatorname{supp}\phi + B_\epsilon(0)}$, *where* K *is a compact subset of* Ω. *Now for any multi-index* α, $D^\alpha\phi$ *is a continuous function defined on a compact set, hence is uniformly continuous. In particular this implies that*

$$\sup_K |D^\alpha\phi(y + x_l) - D^\alpha\phi(y)| \to 0, \qquad \text{as } x_l \to 0,$$

which immediately gives us that $\tau_{x_l}\phi \to \phi$ *in* $\mathscr{D}(\Omega)$.

Example 6. *Suppose* $\phi \in \mathscr{D}(\Omega)$. *For* $h > 0$ *sufficiently small, we define the* forward difference quotient*:*

$$\Delta_i^h\phi = \frac{1}{h}\left(\tau_{-he_i}\phi - \phi\right)$$

with $\{e_i\}_{i=1}^n$ *the standard basis on* $\mathbb{R}^n$. *Then*

$$\Delta_i^h\phi \to D_i\phi, \qquad \text{as } h \to 0.$$

By the same argument as for the previous example, there exists a compact $K \subset \Omega$ *such that* $\operatorname{supp}\Delta_i^h\phi \subset K$ *for* h *sufficiently small. By the mean value theorem, for each* $x \in K$, *there exists* $t_x \in (0, h)$ *such that*

$$D^\alpha\Delta_i^h\phi(x) = \frac{D^\alpha\phi(x + he_i) - D^\alpha\phi(x)}{h} = D_iD^\alpha\phi(x + t_xe_i)$$

Fix $\epsilon > 0$. *Since* $D_iD^\alpha\phi$ *is continuous on* K *(hence uniformly continuous), there exists* $\delta > 0$, *independent of* x *such that If* $t_x < \delta$ *we have*

$$|D_iD^\alpha\phi(x + t_xe_i) - D_iD^\alpha\phi(x)| < \epsilon.$$

If we take $h < \delta$, *then* $t_x < \delta$ *for all* x *and we conclude:*

$$\sup_{x\in K} |D_iD^\alpha\phi(x + t_xe_i) - D_iD^\alpha\phi(x)| < \epsilon,$$

which implies

$$\sup_{x\in K} \left|D^\alpha\Delta_i^h\phi(x) - D^\alpha D_i\phi(x)\right| \to 0 \qquad \text{as } h \to 0.$$

Example 7. *Fix* $\phi \in \mathscr{D}(\mathbb{R})$ *with* $\phi(x) \not\equiv 0$. *The sequence:*

$$\phi_j(x) = \frac{1}{j}\phi(x - j), \qquad j = 1, 2, \ldots$$

does NOT *converge in* $\mathscr{D}(\mathbb{R})$. *We have that*

$$\sup_{\mathbb{R}} |D^\alpha\phi| \to 0, \qquad \text{as } j \to \infty,$$

but there is no compact set which contains the support of ϕ_j *for all* j.

3.2 The space $\mathscr{E}(\Omega)$

The set $C^\infty(\Omega)$ can be given a topology in a fairly natural way. We recall that given $\Omega \subset \mathbb{R}^n$ open, we can find a sequence of compact sets $K_i \subset \Omega$ such that $K_i \subset K_{i+1}^\circ$ and $\Omega = \cup K_i$. We define a family of semi-norms on $C^\infty(\Omega)$ by $\mathscr{P} = \{p_N\}_{N=1}^\infty$, where for $N = 0, 1, 2, \dots$

$$p_N(\phi) = \sup_{x \in K_N, |\alpha| \leq N} |D^\alpha \phi(x)| \,.$$

This is a separating family of seminorms, giving rise to a topology $\tau_{\mathscr{P}}$ such that vector space operations are continuous. We denote the set $C_c^\infty(\Omega)$ equipped with the topology $\tau_{\mathscr{P}}$ by $\mathscr{E}(\Omega)$. We can characterise the convergent sequences:

Theorem 3.2. *A sequence $\{\phi_j\}_{j=1}^\infty \subset \mathscr{E}(\Omega)$ converges to zero if for every compact $K \subset \Omega$ and for each multi-index α we have:*

$$\sup_{x \in K} |D^\alpha \phi_j| \to 0,$$

as $j \to \infty$. Similarly, $\phi_j \to \phi$ if $\phi_j - \phi \to 0$.

Since $\mathscr{P}$ is countable, the topology of $\mathscr{E}(\Omega)$ comes from a translation invariant metric, and moreover one can verify that it is complete, hence $\mathscr{E}(\Omega)$ is a *Fréchet* space.

Example 8. *Recall that $C_c^\infty(\Omega) \subset C^\infty(\Omega)$. If $\{\phi_i\}_{i=1}^\infty \subset C_c^\infty(\Omega)$ tends to 0 in $\mathscr{D}(\Omega)$, then $\phi_i \to 0$ in $\mathscr{E}(\Omega)$. In fact, we can say more: the inclusion map $\iota : \mathscr{D}(\Omega) \hookrightarrow \mathscr{E}(\Omega)$ is continuous.*

Example 9. *Fix $\phi \in \mathscr{D}(\mathbb{R})$ with $\phi(x) \not\equiv 0$, and consider the sequence:*

$$\phi_j(x) = j\phi(x - j), \qquad j = 1, 2, \dots$$

This converges to 0 in $\mathscr{E}(\mathbb{R})$. For any compact K, $\operatorname{supp} \phi_j \cap K = \emptyset$ for j sufficiently large, i.e., the support of ϕ_j eventually leaves any compact set. This shows that the topology of $\mathscr{D}(\Omega)$ is not *simply the induced topology of $C_c^\infty(\Omega)$ thought of as a subspace of $\mathscr{E}(\Omega)$.*

Exercise(∗). a) Suppose $\phi \in \mathscr{E}(\mathbb{R}^n)$. Let $\{x_l\}_{l=1}^\infty \subset \mathbb{R}^n$ be a sequence with $x_l \to 0$. Show that

$$\tau_{x_l}\phi \to \phi, \qquad \text{as } l \to \infty.$$

in $\mathscr{E}(\mathbb{R}^n)$, where τ_x is the translation operator defined in equation (1.2).

b) Suppose $\phi \in \mathscr{E}(\mathbb{R}^n)$, show that

$$\Delta_i^h \phi \to D_i \phi, \qquad \text{as } h \to 0,$$

in $\mathscr{E}(\mathbb{R}^n)$, where Δ_i^h is the difference quotient defined in Example 6.

3.3 The space $\mathscr{S}(\mathbb{R}^n)$

The spaces $\mathscr{D}(\Omega)$ and $\mathscr{E}(\Omega)$ are both defined on arbitrary open sets in $\mathbb{R}^n$. The final space of functions that we wish to consider is a subspace of $\mathscr{E}(\mathbb{R}^n)$ consisting of functions which are *rapidly decreasing* near infinity.

Definition 3.1. *A function $\phi \in C^\infty(\mathbb{R}^n)$ is said to be* rapidly decreasing *if:*

$$\sup_{x\in\mathbb{R}^n} \left|(1+|x|)^N D^\alpha \phi(x)\right| < \infty$$

for all multi-indices α and all $N \in \mathbb{N}$.

Notice that rapidly decreasing functions and their derivatives decay faster than any inverse power of $|x|$ as $|x| \to \infty$.

Example 10. *i) Suppose $\phi \in C_c^\infty(\mathbb{R}^n)$, then ϕ is rapidly decreasing.*

ii) The function $x \mapsto e^{-|x|^2}$ is rapidly decreasing.

The set of rapidly decreasing functions can be endowed with a topology as follows. We define a family of semi-norms by $\mathscr{P} = \{p_N\}_{N=1}^\infty$, where

$$p_N(\phi) = \sup_{x\in\mathbb{R}^n, |\alpha|\le N} \left|(1+|x|)^N D^\alpha \phi(x)\right|.$$

This is a separating family of seminorms, giving rise to a topology $\tau_{\mathscr{P}}$ such that vector space operations are continuous. We denote the set of rapidly decreasing functions equipped with the topology $\tau_{\mathscr{P}}$ by $\mathscr{S}(\mathbb{R}^n)$ or $\mathscr{S}$. This is known as the *Schwartz class* of functions. We can characterise the convergent sequences:

Theorem 3.3. *A sequence $\{\phi_j\}_{j=1}^\infty$ of rapidly decreasing functions tends to zero in $\mathscr{S}$ if for every multi-index α and $N \in \mathbb{N}$ we have:*

$$\sup_{x\in\mathbb{R}^n} \left|(1+|x|)^N D^\alpha \phi_j(x)\right| \to 0,$$

as $j \to \infty$. Similarly, $\phi_j \to \phi$ with respect to τ if $\phi_j - \phi \to 0$.

As for $\mathscr{E}(\mathbb{R}^n)$, the topology on $\mathscr{S}$ is induced by a complete translation invariant metric, so that $\mathscr{S}$ is a *Fréchet space*. The topology is not induced by a norm, so it cannot be given a Banach space structure.

Lemma 3.4. *The spaces $\mathscr{D}(\mathbb{R}^n)$, $\mathscr{S}$ and $\mathscr{E}(\mathbb{R}^n)$ satisfy:*

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

Moreover, the inclusion map is continuous in each case.

Exercise 3.1. a) Show that $\mathscr{S}$ is a vector subspace of $\mathscr{E}(\mathbb{R}^n)$. Show that if $\{\phi_j\}_{j=1}^\infty$ is a sequence of rapidly decreasing functions which tends to zero in $\mathscr{S}$, then $\phi_j \to 0$ in $\mathscr{E}(\mathbb{R}^n)$.

b) Show that $\mathscr{D}(\mathbb{R}^n)$ is a vector subspace of $\mathscr{S}$. Show that if $\{\phi_j\}_{j=1}^{\infty}$ is a sequence of compactly supported functions which tends to zero in $\mathscr{D}(\mathbb{R}^n)$ then $\phi_j \to 0$ in $\mathscr{S}$.

c) Give an example of a sequence $\{\phi_j\}_{j=1}^{\infty} \subset C_c^{\infty}(\mathbb{R}^n)$ such that

i) $\phi_j \to 0$ in $\mathscr{S}$, but ϕ_j has no limit in $\mathscr{D}(\mathbb{R}^n)$.

ii) $\phi_j \to 0$ in $\mathscr{E}(\mathbb{R}^n)$, but ϕ_j has no limit in $\mathscr{S}$.

Exercise 3.2. For each $X \in \{\mathscr{D}(\mathbb{R}^n), \mathscr{S}, \mathscr{E}(\mathbb{R}^n)\}$, suppose $\phi \in X$ and establish:

a) If $x_l \in \mathbb{R}^n$, $x_l \to 0$, then

$$\tau_{x_l}\phi \to \phi, \qquad \text{in } X \text{ as } l \to \infty,$$

where τ_x is the translation operator defined by $\tau_x \phi(y) := \phi(y - x)$.

b) If $h_l \in \mathbb{R}$, $h_l \to 0$, then

$$\Delta_i^{h_l}\phi \to D_i\phi, \qquad \text{in } X \text{ as } l \to \infty,$$

in X, where $\Delta_i^h \phi := h^{-1}\left[\tau_{-he_i}\phi - \phi\right]$ is the difference quotient.

3.4 Distributions

The theory of distributions (sometimes called generalised functions) allows us to consider familiar functions as sitting within a larger class of objects, which are in some sense easier to manipulate, and in which certain equations are easier to solve. This is a familiar idea in the context of complex numbers, which are introduced to extend the real numbers such that every polynomial has a root.

To motivate the idea, recall the general linear PDE of order k,:

$$Lu := \sum_{|\alpha| \leqslant k} a_\alpha D^\alpha u = f, \tag{3.1}$$

where we'll assume that $a_\alpha \in C^\infty(\Omega)$ for an open $\Omega \subset \mathbb{R}^n$. We want to extend the notion of a solution for this PDE to include the situation where u and f need not be of class $C^k(\Omega)$. Let's denote by X the space to which our generalised solution u and right hand side f should belong. What properties do we require for X so that we can at least make sense of the PDE (3.1)? We can start to make a list of desirable properties:

i) The smooth functions $C^\infty(\Omega)$ should be included in X in such a way that we can recover them (i.e. the inclusion map $\iota : C^\infty(\Omega) \hookrightarrow X$ should be injective).

ii) X should be a vector space over $\mathbb{C}$, and the vector space operations should be compatible with the inclusion ι.

iii) We need to be able to multiply elements of X by elements of $C^\infty(\Omega)$.

iv) We need to be able to differentiate elements of X. Moreover whatever definition for 'differentiation' we come up with, it should be compatible with the inclusion ι.

v) We need some topology on X such that the above operations are continuous.

The idea that we shall pursue is to consider the space of distributions to be the dual space to that space of test functions, i.e. we take as our space $\mathscr{D}'(\Omega)$ the continuous dual space of X:

Definition 3.2. *A distribution $u \in \mathscr{D}'(\Omega)$ is a linear functional on the space of test functions*

$$u : \mathscr{D}(\Omega) \to \mathbb{C}$$

which is continuous with respect to the topology of $\mathscr{D}(\Omega)$.

We will state (but not prove) a criterion for continuity. Those interested in the proof of this result will find it in the Appendix.

Theorem 3.5. *Let $u : \mathscr{D}(\Omega) \to \mathbb{C}$ be a linear map. The following are equivalent:*

i) u is continuous with respect to the topology of $\mathscr{D}(\Omega)$.

ii) For each sequence $\{\phi_j\}_{j=1}^{\infty} \subset \mathscr{D}(\Omega)$ with $\phi_j \to 0$ in $\mathscr{D}(\Omega)$, we have:

$$\lim_{j\to\infty} u\left[\phi_j\right] = 0.$$

iii) For each compact $K \subset \Omega$, there exists $N \in \mathbb{N}$ and a constant C such that:

$$|u[\phi]| \leqslant C \sup_{x\in K} \sum_{|\alpha|\leqslant N} \left|D^\alpha \phi(x)\right|, \qquad \text{for all } \phi \in C_c^\infty(K). \tag{3.2}$$

Remark. *1. By the linearity of u, condition ii) is equivalent to:*

$$\lim_{j\to\infty} u\left[\phi_j\right] = u\left[\phi\right], \qquad \text{for all } \phi_j \to \phi \text{ in } \mathscr{D}(\Omega).$$

2. If there exists a single $N \in \mathbb{N}$ such that (3.2) holds for all compact $K \subset \Omega$ (possibly with C depending on K), then we say that u has finite order*. The least such N is called the* order *of u.*

For a general topological space (as opposed to a metric space) in order to establish that a function is continuous, it is necessary to consider open sets and their pull-backs etc. It is not usually enough to simply check continuity for sequences. The reason that we can get away with it in this case is somewhat complicated, but boils down to the fact that although the topology of $\mathscr{D}(\Omega)$ does not arise as a metric topology, it is in some sense the limit of a sequence of spaces which are metric.

We can think of a distribution as an operation which swallows a smooth function of compact support and produces a real number. Let's look at two important examples:

Example 11. *a) (The Dirac delta) For $x \in \Omega$ we define the distribution:*

$$\delta_x \phi := \phi(x) \qquad \forall\ \phi \in \mathscr{D}(\Omega).$$

b) If $f \in L^1_{loc.}$, we can define the distribution T_f by:

$$T_f \phi := \int_\Omega f(x)\phi(x)dx, \qquad \forall\ \phi \in \mathscr{D}(\Omega).$$

c) For $\phi \in \mathscr{D}(\mathbb{R})$, we define:

$$P.V. \left(\frac{1}{x}\right)[\phi] = \lim_{\epsilon \to 0} \left[\int_{-\infty}^{-\epsilon} \frac{\phi(x)}{x} dx + \int_{\epsilon}^{\infty} \frac{\phi(x)}{x} dx\right].$$

This is clearly linear (assuming the limit exists). By a change of variables, we can re-write:

$$P.V. \left(\frac{1}{x}\right)[\phi] = \lim_{\epsilon \to 0} \int_{\epsilon}^{\infty} \frac{\phi(x) - \phi(-x)}{x} dx$$

Note that

$$\frac{\phi(x) - \phi(-x)}{x} = \int_{-1}^{1} \phi'(xt)dt$$

so that:

$$\left|\frac{\phi(x) - \phi(-x)}{x}\right| = \left|\int_{-1}^{1} \phi'(xt)dt\right| \leqslant 2 \sup_{\mathbb{R}^n} |\phi'|.$$

From this, we conclude that the limit $\epsilon \to 0$ above is well defined, and moreover, if $\phi \in C^\infty_c(B_R(0))$ then:

$$\left|P.V. \left(\frac{1}{x}\right)[\phi]\right| \leqslant 4R \sup_{\mathbb{R}^n} |\phi'|.$$

We conclude that $P.V. \left(\frac{1}{x}\right)$ defines a distribution of order at most one.

Exercise($*$). a) Show that δ_x, as defined in Example 11 is continuous and linear, hence a distribution. Find the order.

b) Show that T_f, as defined in Example 11 is continuous and linear, hence a distribution. Find the order.

c) By constructing a suitable sequence of smooth functions show that the order of $P.V. \left(\frac{1}{x}\right)$ is one.

3.5 Functions as distributions

Let's have a look at how we're doing with our 'wish list' of properties. For Property *i)* we will take inspiration from part *a)* of the example above and define

$$\begin{aligned} \iota\ :\ C^\infty(\Omega) &\to \mathscr{D}'(\Omega), \\ f &\mapsto T_f. \end{aligned}$$

In order to check that this is injective, we need to show that if $T_f = T_g$ then $f = g$. We shall prove this by showing that for any distribution of the form T_f, it is possible to recover f by applying T_f to appropriately shifted and scaled bump functions. Recall that in the proof of Theorem 1.10, we introduced the bump functions ϕ_ϵ. The idea will be to make use of our previous results about convolutions. We require a bit of notation first. Recall that if $\phi : \mathbb{R}^n \to \mathbb{C}$, and $x \in \mathbb{R}^n$ we set $\tau_x \phi(y) = \phi(y - x)$. We introduce the spatial inversion operator $\check{\cdot}$ defined by $\check{\phi}(y) = \phi(-y)$. By convention, the translation operator acts first, so that $\tau_x \check{\phi}(y) = \phi(x - y)$.

Theorem 3.6. *Suppose $f \in C^k(\Omega)$, and let ϕ_ϵ be as in Theorem 1.13. Define for any x with $d(x, \partial\Omega) \geqslant \epsilon$:*

$$f_\epsilon(x) := T_f[\tau_x \check{\phi}_\epsilon].$$

Then for any compact subset $K \subset \Omega$ and any $|\alpha| \leqslant k$ we have:

$$\sup_{x \in K} |D^\alpha f_\epsilon(x) - D^\alpha f(x)| \to 0$$

as $\epsilon \to 0$.

Proof. Fix $K \subset \Omega$ compact. Recall that by Lemma 1.14, there exists $\chi \in C_c^\infty(\Omega)$ such that $\chi = 1$ on $K + B_\delta(0)$ for some $\delta > 0$. Take $\epsilon < \delta$. Then, since $\operatorname{supp} \phi_\epsilon \subset B_\epsilon(0)$ we have for $x \in K$:

$$\begin{aligned} f_\epsilon(x) &= \int_\Omega f(y) \tau_x \check{\phi}_\epsilon(y) dy \\ &= \int_\Omega f(y) \phi_\epsilon(x - y) dy \\ &= \int_{\mathbb{R}^n} \chi(y) f(y) \phi_\epsilon(x - y) dy \\ &= \phi_\epsilon \star (\chi f). \end{aligned}$$

Here we have used the fact that when $|x - y| < \epsilon$ we have $\chi = 1$ to insert the cut-off function without altering the integral. Now, since χ is smooth, $\chi f \in C_c^k(\mathbb{R}^n)$, so as $\epsilon \to 0$, we have by Theorem 1.13 that for any $|\alpha| \leqslant k$:

$$D^\alpha(\phi_\epsilon \star (\chi f)) \to D^\alpha(\chi f)$$

uniformly on $\mathbb{R}^n$ as $\epsilon \to 0$. In particular we have uniform convergence on K, so that:

$$\sup_{x \in K} |D^\alpha(\phi_\epsilon \star (\chi f))(x) - D^\alpha(\chi f)(x)| \to 0.$$

Since for $x \in K$ we have $\phi_\epsilon \star (\chi f)(x) = f_\epsilon(x)$ and $\chi(x) = 1$, this is the result we require. ☐

This result immediately tells us that our map ι is injective.

Corollary 3.7. *Suppose $f, g \in C^0(\Omega)$. If $T_f = T_g$ then $f \equiv g$. In particular this implies $\iota : C^\infty(\Omega) \hookrightarrow \mathscr{D}'(\Omega)$ is injective.*

Exercise(∗). Suppose $f \in L^1_{loc.}(\Omega)$. Take ϕ_ϵ as in Theorem 1.13, and define for x with $d(x, \partial\Omega) \geqslant \epsilon$:

$$f_\epsilon(x) := T_f \left[\tau_x \check{\phi_\epsilon} \right].$$

Show that for any compact $K \subset \Omega$:

$$\|f_\epsilon - f\|_{L^1(K)} \to 0$$

as $\epsilon \to 0$.
[Hint: follow the proof of Theorem 3.6, but use part b) of Theorem 1.13]

3.6 Derivatives of distributions

Things are looking good for Property *ii*) because the dual space to a vector space is naturally a vector space. If $u_1, u_2 \in \mathscr{D}'(\Omega)$ we define the sum $u_1 + u_2 \in \mathscr{D}'(\Omega)$ by:

$$(u_1 + u_2)\phi = u_1\phi + u_2\phi \qquad \forall\ \phi \in \mathscr{D}(\Omega).$$

It's easy to check that this is linear and continuous from the properties of u_i.

How about Property *iii*)? First, let's notice that if $a \in C^\infty(\Omega)$ and $\phi \in C^\infty_c(\Omega)$ then $a\phi \in C^\infty_c(\Omega)$. We can therefore define for $u \in \mathscr{D}'(\Omega)$ the product $au \in \mathscr{D}'(\Omega)$ by

$$(au)\phi = u[a\phi] \qquad \forall\ \phi \in \mathscr{D}(\Omega).$$

Now let's consider Property *iv*). We want to find a definition for the derivative of a distribution which gives the right answer when the distribution in question arises from a smooth function f by the map $\iota : f \mapsto T_f$. If $f \in C^\infty(\Omega)$, then certainly $D_i f \in C^\infty(\Omega)$. Let's consider $T_{D_i f}$. We have:

$$T_{D_i f}\phi = \int_\Omega [D_i f](x)\phi(x)dx$$

Since $\phi \in C^\infty_c(\Omega)$, we can integrate by parts in this integral without picking up any boundary terms to find

$$\begin{aligned} T_{D_i f}\phi &= -\int_\Omega f(x)[D_i\phi](x)dx \\ &= -T_f\left[D_i\phi\right]. \end{aligned}$$

Motivated by this, we *define* the D^α derivative of a distribution u to be the distribution $D^\alpha u$ which acts on a test function $\phi \in C^\infty_c(\Omega)$ as:

$$(D^\alpha u)[\phi] := (-1)^{|\alpha|} u\left[D^\alpha \phi\right].$$

The derivative we have defined on distributions *extends* the usual derivative for functions defined as the linear approximation to the function at a point. The advantage of the distributional derivative is that it is defined for any distribution. We'll work out a couple of examples:

Example 12. *a) The derivative of the Dirac delta acts on a test function $\phi \in C_c^\infty(\Omega)$ by:*

$$(D_i\delta_x)\,\phi = -\delta_x\,[D_i\phi] = -\frac{\partial \phi}{\partial x_i}(x), \qquad \forall\, \phi \in C_c^\infty(\mathbb{R}^n).$$

b) Consider the Heaviside function $H : \mathbb{R} \to \mathbb{R}$ defined by

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

This function is certainly not *differentiable at $x = 0$. We can define the distribution $T_H \in \mathscr{D}'(\mathbb{R})$ to be*

$$T_H\phi = \int_{\mathbb{R}} H(x)\phi(x)dx$$

for $\phi \in C_c^\infty(\mathbb{R})$. We then compute:

$$\begin{aligned} (D_xT_H)\,\phi &= -T_H\,[D_x\phi] \\ &= -\int_{\mathbb{R}} H(x)\phi'(x)dx \\ &= -\int_0^\infty \phi'(x)dx \\ &= \phi(0) \end{aligned}$$

Here, we've used the fact that ϕ has compact support as well as the fact that H vanishes for $x < 0$. Thus, we can say that

$$D_xT_H = \delta_0.$$

Thus we see that the theory of distributions allows us to give some sort of meaning to the derivative of a functions whose classical derivative does not exist.

Exercise(∗). a) Show that if $f_1, f_2 \in C^0(\Omega)$ and $a \in C^\infty(\Omega)$, then

$$aT_{f_1} + T_{f_2} = T_{af_1+f_2}$$

b) Show that if $f \in C^k(\Omega)$ then

$$D^\alpha T_f = T_{D^\alpha f}$$

for $|\alpha| \leqslant k$. Deduce that $\iota \circ D^\alpha = D^\alpha \circ \iota$.

c) Deduce that if $f \in C^k(\Omega)$ then

$$\sum_{|\alpha|\leqslant k} a_\alpha D^\alpha T_f = T_{Lf}.$$

where

$$Lf = \sum_{|\alpha|\leqslant k} a_\alpha D^\alpha f$$

Exercise 3.3. Suppose $u \in \mathscr{D}'(\mathbb{R})$ satisfies $Du = 0$. Show that u is a constant distribution, i.e. there exists $\lambda \in \mathbb{C}$ such that:

$$u[\phi] = \lambda \int_{\mathbb{R}} \phi(x)dx, \qquad \text{for all } \phi \in \mathscr{D}(\mathbb{R}).$$

(*) Extend the result to $\mathbb{R}^n$ for $n > 1$.
[Hint: Fix $\phi_0 \in \mathscr{D}(\mathbb{R})$ and show that any $\phi \in \mathscr{D}(\mathbb{R})$ may be written as $\phi(x) = \psi'(x) + c_\phi \phi_0(x)$ for some $\psi \in \mathscr{D}(\mathbb{R})$, $c_\phi \in \mathbb{C}$.]

A useful result allows us to infer regularity of a function from the regularity of its distributional derivatives:

Theorem 3.8. *Suppose that $f \in C^0(\Omega)$ defines the distribution T_f in the usual way and suppose moreover that for any multi-index α with $|\alpha| \leqslant k$ there exists $g^\alpha \in C^0(\Omega)$ such that*

$$D^\alpha T_f = T_{g^\alpha}$$

where D^α is the distributional derivative. Then in fact $f \in C^k(\Omega)$ and $D^\alpha f = g^\alpha$ in the sense of classical derivatives.

Proof. First we show the result for $k = 1$. Let us fix a compact subset K of Ω and let ϕ_ϵ be as in Theorem 1.13. Let us define for $x \in K$ and ϵ sufficiently small:

$$f_\epsilon(x) = T_f[\tau_x \check{\phi}_\epsilon] = \int_\Omega f(y)\phi_\epsilon(x-y)dy.$$

We know from Theorem 3.6 that $f_\epsilon \to f$ uniformly on K as $\epsilon \to 0$. Let us calculate

$$\begin{aligned} D_i f_\epsilon(x) &= \int_\Omega f(y) \frac{\partial}{\partial x_i} \phi_\epsilon(x-y)dy \\ &= -\int_\Omega f(y) \frac{\partial}{\partial y_i} \phi_\epsilon(x-y)dy \\ &= (D_i T_f)\left[\tau_x \check{\phi}_\epsilon\right] \\ &= (T_{g_i})\left[\tau_x \check{\phi}_\epsilon\right] \end{aligned}$$

Now, as $\epsilon \to 0$, we know that $(T_{g_i})\left[\tau_x \check{\phi}_\epsilon\right] \to g_i(x)$ uniformly on K. Thus we have

$$f_\epsilon \to f, \quad D_i f_\epsilon \to g_i$$

uniformly on K as $\epsilon \to 0$. This implies that $f \in C^1(K)$, $D_i f = g_i$. Since this holds for any compact set we have that $f \in C^1(\Omega)$ with $D_i f = g_i$. By repeated application of the $k = 1$ result, we can establish that the result holds for all k. □

This tells us that the distributional derivative is essentially equivalent to the classical derivative wherever both are defined and continuous. One should be careful, however. There are examples of continuous functions whose derivative vanishes Lebesgue-almost everywhere, but whose distributional derivative is not the zero distribution. You may wish to look up the Cantor function.

3.7 Convergence of distributions

Property v) of our 'wish list' is the only one that we've not yet addressed directly. We wish to give $\mathscr{D}'(\Omega)$ a topology. In fact, as the continuous dual of a topological vector space, it naturally carries the weak-$\star$ topology. For convenience, we recall the main features of this topology:

Theorem 3.9. *The vector space $\mathscr{D}'(\Omega)$ inherits a topology from $\mathscr{D}(\Omega)$, the weak-$\star$ topology, such that*

i) The vector space operations on $\mathscr{D}'(\Omega)$ are continuous.

ii) A sequence $\{u_j\}_{j=1}^\infty \subset \mathscr{D}'(\Omega)$ converges to zero in $\mathscr{D}'(\Omega)$ if

$$u_j[\phi] \to 0, \qquad \text{for all } \phi \in \mathscr{D}(\Omega).$$

as $j \to \infty$. Similarly $u_j \to u$ for $u \in \mathscr{D}'(\Omega)$ if $u_j - u \to 0$.

Example 13. *a) Suppose $\{f_j\}_{j=1}^\infty \subset C^0(\Omega)$ is a sequence of functions such that $f_j \to 0$ uniformly on any compact $K \subset \Omega$. Then $T_{f_j} \to 0$ in $\mathscr{D}'(\Omega)$. To see this, note that for any $\phi \in \mathscr{D}(\Omega)$, there exists a K such that $\operatorname{supp}\phi \subset K$. Then*

$$\left|T_{f_j}[\phi]\right| = \left|\int_K f_j(y)\phi(y)dy\right| \leqslant |K| \sup_K |f_j\phi|$$

but the right hand side is tending to zero, so $T_{f_j}[\phi] \to 0$ for any $\phi \in \mathscr{D}(\Omega)$.

b) A similar argument shows that if $\{f_j\}_{j=1}^\infty \subset L^1_{loc.}(\Omega)$ is a sequence such that $\|f_j\|_{L^1(K)} \to 0$ for any compact $K \subset \Omega$, then $T_{f_j} \to 0$ in $\mathscr{D}'(\Omega)$.

c) Let $\Omega = \mathbb{R}$. Define a distribution as follows. For $\phi \in \mathscr{D}(\Omega)$, set:

$$u[\phi] = \sum_{m=-\infty}^{\infty} \phi^{(|m|)}(m)$$

For any given ϕ this sum will only have finitely many non-zero terms. It is straightforward to verify that u is itself a distribution, in fact it is an example of a distribution of infinite order. Consider the sequence of distributions

$$u_M = \sum_{m=-M}^{M} D^{|m|}\delta_m$$

Let $\phi \in \mathscr{D}(\Omega)$ be any test function. Then $\operatorname{supp}\phi \subset B_R(0)$, so for $M \geqslant R$, we have:

$$u_M[\phi] = \sum_{m=-M}^{M} \phi^{(|m|))}(m) = u[\phi]$$

Thus we can write:

$$u = \sum_{m=-\infty}^{\infty} D^{|m|}\delta_m,$$

where the sum converges in $\mathscr{D}(\Omega)$.

d) Consider the bump functions $\tau_x\check{\phi}_\epsilon$ for $x \in \Omega$, where ϕ_ϵ is as constructed in Theorem 1.13. Then:

$$T_{\tau_x\check{\phi}_\epsilon} \to \delta_x$$

in $\mathscr{D}'(\Omega)$ as $\epsilon \to 0$. To see this, recall that for $\psi \in \mathscr{D}(\Omega)$, and ϵ sufficiently small:

$$T_{\tau_x\check{\phi}_\epsilon}[\psi] = \int_\Omega \phi_\epsilon(y)\psi(x-y)dy = \int_{\mathbb{R}^n} \phi_\epsilon(y)\psi(x-y)dy = \phi_\epsilon \star \psi(x).$$

By Theorem 1.13 we have $T_{\tau_x\check{\phi}_\epsilon}[\psi] \to \psi(x) = \delta_x[\psi]$. Since ψ was arbitrary the result follows.

3.8 Convolutions and the fundamental solution

What is the advantage of introducing distributions? Taking together various of the properties we've considered above, we can formulate the following proposition.

Proposition 2. *Suppose that $f \in C^0(\Omega)$ and that $T \in \mathscr{D}'(\Omega)$ satisfies the distributional equation:*

$$\sum_{|\alpha|\leqslant k} a_\alpha D^\alpha T = T_f,$$

for $a_\alpha \in C^\infty(\Omega)$ and that moreover there exist functions $u_\alpha \in C^0(\Omega)$ such that $D^\alpha T = T_{u_\alpha}$. Then $u := u_0 \in C^k(\Omega)$ is a classical solution of the equation

$$\sum_{|\alpha|\leqslant k} a_\alpha D^\alpha u = f.$$

This gives us a new approach to finding a classical solution for a linear PDE. We first show that there exists a distribution which solves the PDE and then worry about whether it is in fact a classical solution.

Now we are going to specialise to the case of linear operators of constant coefficients defined on $\mathbb{R}^n$. We will take L to be the partial differential operator

$$L := \sum_{|\alpha|\leqslant k} a_\alpha D^\alpha,$$

where a_α are now assumed to be constant. We wish to find solutions to the distributional equation

$$Lu = T_f$$

since by Proposition 2 we hope that this will lead to classical solutions $w \in C^k(\Omega)$ of the equation

$$\sum_{|\alpha|\leqslant k} a_\alpha D^\alpha w = f,$$

A powerful approach to finding a distributional solution to a PDE is to first construct a fundamental solution. For this, we will require the notion of a convolution. We will specialise to the case where $\Omega = \mathbb{R}^n$, i.e. to distributions defined on all of space.

It's useful to introduce the notion of the support of a distribution.

Definition 3.3. *A distribution $u \in \mathscr{D}'(\Omega)$ is supported in the closed set $K \subset \Omega$ if*

$$u[\phi] = 0 \qquad \forall\ \phi \in C_c^\infty(\Omega \setminus K)$$

in other words if u gives zero when applied to any test function that vanishes on K. The support of u, supp u is the set:

$$\text{supp } u = \bigcap \{K : u \text{ supported in } K\}.$$

As an intersection of closed sets, this is closed. If there exists a compact K such that u is supported in K then we say that u has compact support.

With this definition it is easy to see that for $f \in C^0(\Omega)$, supp $T_f =$ supp f where the support of the function f is defined in the usual way as the closure of the set on which f is non-zero. Also, one can easily check that if u is supported in K, then so is $D^\alpha u$ for any multi-index α.

> **Exercise($*$).** Show that
>
> $$\text{supp } \delta_x = \{x\}.$$
>
> Deduce that there is no function $f \in C^0(\Omega)$ such that $\delta_x = T_f$.

Suppose that $f \in C^0(\mathbb{R}^n)$ and $\phi \in C_c^\infty(\mathbb{R}^n)$. Recall that the convolution of the two functions is defined to be

$$(f \star \phi)(x) = \int_{\mathbb{R}^n} f(y)\phi(x-y)dy.$$

So that

$$(f \star \phi)(x) = T_f\left[\tau_x \check{\phi}\right].$$

Now, we can see that it is straightforward to define the convolution of any distribution $u \in \mathscr{D}'(\mathbb{R}^n)$ with a test function $\phi \in \mathscr{D}(\mathbb{R}^n)$ by the formula:

$$(u \star \phi)(x) := u\left[\tau_x \check{\phi}\right].$$

Lemma 3.10 (Properties of convolutions)**.** *Suppose $u, u_i \in \mathscr{D}'(\mathbb{R}^n)$ and $\phi \in \mathscr{D}(\mathbb{R}^n)$. Then*

i) If

$$u_1 \star \phi = u_2 \star \phi \qquad \text{for all } \phi \in \mathscr{D}(\mathbb{R}^n),$$

then $u_1 = u_2$.

ii) $u \star \phi \in C^\infty(\mathbb{R}^n)$ and

$$D^\alpha(u \star \phi) = u \star D^\alpha \phi = D^\alpha u \star \phi. \tag{3.3}$$

iii) We have:

$$\text{supp}\, u \star \phi \subset \text{supp}\, u + \text{supp}\, \phi.$$

In particular, if u has compact support, then $u \star \phi$ has compact support, and $u \star \phi \in \mathscr{D}(\mathbb{R}^n)$.

Proof. i) Notice that $u[\phi] = (u \star \check{\phi})(0)$, so we deduce that

$$u_1\phi = (u_1 \star \tilde{\phi})(0) = (u_2 \star \tilde{\phi})(0) = u_2\phi$$

for any test function ϕ, thus $u_1 = u_2$.

ii) We calculate

$$\begin{aligned}\Delta_i^h [u \star \phi](x) = \frac{(u \star \phi)(x + he_i) - (u \star \phi)(x)}{h} &= \frac{1}{h}\left(u\left[\tau_{(x+he_i)}\check{\phi}\right] - u\left[\tau_x\check{\phi}\right]\right) \\ &= u\left[\frac{1}{h}\left(\tau_{(x+he_i)}\check{\phi} - \tau_x\check{\phi}\right)\right] \\ &= u[\Delta_i^h \tau_x \check{\phi}]\end{aligned}$$

here we use the linearity of u. Now, if $\phi \in C_c^\infty(\mathbb{R}^n)$, then

$$\begin{aligned}\lim_{h\to 0} \Delta_i^h \tau_x \check{\phi} &= \lim_{h \to 0} \frac{\phi(x + he_i - y) - \phi(x - y)}{h} \\ &= D_i\phi(x - y) \\ &= \left(\tau_x(\check{D_i\phi})\right)(y)\end{aligned}$$

with convergence in the topology of $\mathscr{D}(\mathbb{R}^n)$. As a result, using the continuity of the distribution, we have that

$$\lim_{h\to 0} \Delta_i^h [u \star \phi](x) = u\left[\tau_x(\check{D_i\phi})\right] = (u \star D_i\phi)(x).$$

Repeating the argument for higher derivatives, we conclude that the first equality of (3.3) holds. To get the second equality, we calculate:

$$\begin{aligned}D^\alpha \left[\tau_x\check{\phi}\right](y) &= \frac{\partial^{|\alpha|}}{\partial y^\alpha}[\phi(x - y)] \\ &= (-1)^{|\alpha|}(D^\alpha\phi)(x - y) \\ &= (-1)^{|\alpha|}\left[\tau_x(\check{D^\alpha\phi})\right](y)\end{aligned}$$

Now, using the definition of the derivative of a distribution:

$$\begin{aligned}(D^\alpha u \star \phi)(x) &= D^\alpha u\left[\tau_x\check{\phi}\right] = (-1)^{|\alpha|}u\left[D^\alpha\left(\tau_x\check{\phi}\right)\right] \\ &= (-1)^{|\alpha|}u\left[(-1)^{|\alpha|}\left[\tau_x(\check{D^\alpha\phi})\right]\right] \\ &= u\left[\tau_x(\check{D^\alpha\phi})\right] \\ &= (u \star D^\alpha\phi)(x).\end{aligned}$$

iii) Suppose for $x \in \mathbb{R}^n$ that $u \star \phi(x) \neq 0$. Then we must have

$$\operatorname{supp} u \cap \operatorname{supp} \tau_x\check{\phi} \neq \emptyset.$$

In particular, there is $z \in \operatorname{supp} u$ such that $\tau_x \check{\phi}(z) = \phi(x - z) \neq 0$. Thus $x - z \in \operatorname{supp} \phi$ for $z \in \operatorname{supp} u$, and we conclude:

$$\{x : u \star \phi(x) \neq 0\} \subset \operatorname{supp} u + \operatorname{supp} \phi.$$

Since $\operatorname{supp} \phi$ is compact and $\operatorname{supp} u$ is closed, $\operatorname{supp} u + \operatorname{supp} \phi$ is closed, and so:

$$\operatorname{supp} u \star \phi = \overline{\{x : u \star \phi(x) \neq 0\}} \subset \operatorname{supp} u + \operatorname{supp} \phi. \qquad \square$$

We would like to define the convolution of two distributions. To do this, we recall that if $f, g, h \in C^0_c(\mathbb{R}^n)$ then

$$(f \star g) \star h = f \star (g \star h), \tag{3.4}$$

so that the convolution is associative. Motivated by this, we define:

Definition 3.4. *Suppose $u_1, u_2 \in \mathscr{D}'(\mathbb{R}^n)$ are distributions and that u_2 has compact support. The convolution $u_1 \star u_2$ is the unique distribution which satisfies*

$$(u_1 \star u_2) \star \phi = u_1 \star (u_2 \star \phi)$$

for all test functions $\phi \in \mathscr{D}(\mathbb{R}^n)$.

Notice that the fact that u_2 has compact support is required for this definition to make sense, as it ensures $(u_2 \star \phi)$ is a test function. Since we can recover a distribution from its convolution with an arbitrary test function, this defines $u_1 \star u_2$.

Exercise($*$). a) Show that for $f, g \in C^0_c(\mathbb{R}^n)$:

$$T_{f \star g} = T_f \star T_g.$$

b) Show that convolution is linear in both of its arguments, i.e. if $u_i \in \mathscr{D}'(\mathbb{R}^n)$ and u_3, u_4 have compact support then

$$(u_1 + a u_2) \star u_3 = u_1 \star u_3 + a u_2 \star u_3$$

and

$$u_1 \star (u_3 + a u_4) = u_1 \star u_3 + a u_1 \star u_4$$

where $a \in \mathbb{C}$ is a constant.

Exercise 3.4. Let $X \in \{\mathscr{D}(\mathbb{R}^n), \mathscr{S}, \mathscr{E}(\mathbb{R}^n)\}$. For $u \in X'$, $x \in \mathbb{R}^n$, define $\tau_x u$ by $\tau_x u[\phi] = u[\tau_{-x}\phi]$ for all $\phi \in X$, and let $\Delta^h_i u = h^{-1}[\tau_{-he_i} u - u]$. Show that $\Delta^h_i u \to D_i u$ as $h \to 0$ in the weak-$*$ topology of X'.

Exercise 3.5. Suppose $u \in \mathscr{D}'(\mathbb{R})$ satisfies $xu = 0$. Show that $u = c\delta_0$ for some $c \in \mathbb{C}$. Find the most general $u \in \mathscr{D}'(\mathbb{R})$ which satisfies $x^k u = 0$ for some $k \in \mathbb{N}$.

Theorem 3.11. *Suppose that $u_1, u_2 \in \mathscr{D}'(\mathbb{R}^n)$ are distributions and that u_2 has compact support. Then*

$$D^\alpha(u_1 \star u_2) = u_1 \star D^\alpha u_2 = D^\alpha u_1 \star u_2 \tag{3.5}$$

Proof. We will consider the convolution of the distribution $D^\alpha(u_1 \star u_2)$ with an arbitrary test function ϕ and use the definition of the convolution, together with the previous Lemma to shuffle the derivatives around:

$$\begin{aligned} D^\alpha(u_1 \star u_2) \star \phi &= (u_1 \star u_2) \star D^\alpha \phi \\ &= u_1 \star (u_2 \star D^\alpha \phi) \\ &= u_1 \star (D^\alpha u_2 \star \phi) \\ &= (u_1 \star D^\alpha u_2) \star \phi, \end{aligned}$$

which establishes the first equality in (3.5) since ϕ was arbitrary. For the second equality, we note

$$\begin{aligned} (u_1 \star D^\alpha u_2) \star \phi &= u_1 \star (D^\alpha u_2 \star \phi) \\ &= u_1 \star D^\alpha (u_2 \star \phi) \\ &= D^\alpha u_1 \star (u_2 \star \phi) \\ &= (D^\alpha u_1 \star u_2) \star \phi, \end{aligned}$$

again since ϕ was arbitrary we're done. □

Exercise(∗). a) Show that if $\phi \in \mathscr{D}(\mathbb{R}^n)$ then

$$\delta_0 \star \phi = \phi$$

b) Show that if $u \in \mathscr{D}'(\mathbb{R}^n)$ has compact support, then

$$\delta_0 \star u = u$$

Now let us introduce the notion of a fundamental solution to a linear PDE. We say that a distribution G is a *fundamental solution* of the partial differential operator with constant coefficients L

$$L := \sum_{|\alpha| \leqslant k} a_\alpha D^\alpha$$

where a_α are constants if it satisfies the distributional equation:

$$LG = \delta_0$$

The reason that this is useful is the following:

Lemma 3.12. *Suppose that $G \in \mathscr{D}'(\mathbb{R}^n)$ is a fundamental solution of L and let $u_0 \in \mathscr{D}'(\mathbb{R}^n)$ be a distribution of compact support. Then the distribution $u := G \star u_0$ solves the distributional equation*

$$\sum_{|\alpha| \leqslant k} a_\alpha D^\alpha u = u_0.$$

Proof. First note that the linearity and differentiability properties of the convolution imply that

$$\begin{aligned} L(G\star u_0) &= \sum_{|\alpha|\leqslant k} a_\alpha D^\alpha (G\star u_0) \\ &= \sum_{|\alpha|\leqslant k} a_\alpha (D^\alpha G\star u_0) && \text{[Lemma 3.10]} \\ &= \left(\sum_{|\alpha|\leqslant k} a_\alpha D^\alpha G\right)\star u_0 && \text{[Linearity]} \\ &= (LG)\star u_0 \end{aligned}$$

Now, use the definition of the fundamental solution to obtain

$$L(G\star u_0) = \delta_0 \star u_0 = u_0$$

since the convolution of a distribution with δ_0 gives back the distribution (Exercise 3.8). □

Thus, once we can find a fundamental solution, we can essentially solve the equation $Lu = u_0$ for an arbitrary right hand side. Rather crudely, we can think of the Dirac delta distribution as an identity. Then the fundamental solution provides an inverse to the operator L.

3.8.1 An example: Poisson's equation

Consider the following classical PDE problem. Given $f \in C_c^2(\mathbb{R}^3)$, find a $w \in C^2(\mathbb{R}^3)$ such that:

$$\Delta w = f.$$

Here the Laplace operator is given by:

$$\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{3.6}$$

Following the procedure outlined above, we will first turn the equation into the distributional PDE, and seek $u \in \mathscr{D}'(\mathbb{R}^3)$ such that:

$$\Delta u = T_f. \tag{3.7}$$

If we can find a G such that

$$\Delta G = \delta_0,$$

then we can write down a solution of (3.7) by convolution. There are several ways to find such a G. We will note two facts (which we will not attempt to prove at this stage)

- The Laplace operator and δ_0 are both invariant under rotations about the origin.

- The Laplace operator is elliptic. In particular, it has the property of elliptic regularity. Roughly speaking this means that if a distribution u satisfies $\Delta u = 0$ on some open set, then u is a smooth function (thought of as a distribution).

Based on these observations, it is reasonable to suspect that we can write $G = T_g$ for $g \in C^\infty(\mathbb{R}^3 \setminus \{0\})$ a radial function $g = g(r)$. Since g must satisfy the Laplace equation away from the origin, and is spherically symmetric, we have (on changing to polar coordinates):

$$\frac{d^2}{dr^2}(rg) = 0, \qquad r > 0.$$

Thus:

$$g = \frac{A}{r} + B.$$

Clearly, since we can add a constant to any solution of (3.6), B is arbitrary and we choose it to be 0. Let us proceed leaving A arbitrary. We shall eventually see that $A = -(4\pi)^{-1}$. Note that $g \in L^1_{loc.}(\mathbb{R}^3)$, so defines a distribution in the natural way.

We wish to show that for a suitable choice of A, the distribution $G = T_g$ satisfies (3.7). To show this, we take $\phi \in \mathscr{D}(\mathbb{R}^3)$ to be arbitrary, and choose $R > 0$ such that $\operatorname{supp}\phi \subset B_R(0)$. We calculate:

$$\begin{aligned}\Delta T_g[\phi] &= T_g \Delta[\phi] \\ &= \int_{\mathbb{R}^3} g(x)\Delta\phi(x)dx \\ &= \lim_{\epsilon\to 0}\int_{B_R(0)\setminus B_\epsilon(0)} g(x)\Delta\phi(x)dx.\end{aligned}$$

In the last line, we use the dominated convergence theorem to justify the limit. The reason that we have inserted this limit is that on $B_R(0) \setminus B_\epsilon(0)$ the integrand is smooth, so we are entitled to apply the divergence theorem. Note that for ψ_1, ψ_2 smooth functions, we have the identity:

$$\nabla \cdot (\psi_1 \nabla \psi_2 - \psi_2 \nabla \psi_1) = \psi_1 \Delta \psi_2 - \psi_2 \Delta \psi_1.$$

Integrating this identity over $B_R(0) \setminus B_\epsilon(0)$ with $\psi_1 = g$, $\psi_2 = \phi$ and applying the divergence theorem, we have:

$$\begin{aligned}\int_{B_R(0)\setminus B_\epsilon(0)} g(x)\Delta\phi(x)dx &= \int_{\partial(B_R(0)\setminus B_\epsilon(0))}\left[g\frac{\partial\phi}{\partial n} - \phi\frac{\partial g}{\partial n}\right]d\sigma + \int_{B_R(0)\setminus B_\epsilon(0)}\Delta g(x)\phi(x)dx \\ &= \int_{\partial B_\epsilon(0)}\left[\phi\frac{\partial g}{\partial n} - g\frac{\partial\phi}{\partial n}\right]d\sigma\end{aligned}$$

In passing from the first to second line we have noted that $\Delta g = 0$ away from the origin, and also that $\phi = 0$ in a neighbourhood of $\partial B_R(0)$. The change of sign comes from the fact that $\partial B_\epsilon(0)$ is an inner boundary. Now, we estimate:

$$\sup_{y\in\partial B_\epsilon(0)}\left|g(y)\frac{\partial\phi}{\partial n}(y)\right| = \sup_{y\in\partial B_\epsilon(0)}\frac{|A|}{\epsilon}\left|\frac{\partial\phi}{\partial n}(y)\right| \leqslant \frac{|A|}{\epsilon}\sup_{\mathbb{R}^3}|D\phi|.$$

Thus:

$$\left| \int_{\partial B_\epsilon(0)} g \frac{\partial \phi}{\partial n} d\sigma \right| \leqslant 4\pi\epsilon^2 \times \frac{|A|}{\epsilon} \sup_{\mathbb{R}^3} |D\phi| \to 0,$$

as $\epsilon \to 0$. For the other term, note that on $\partial B_\epsilon(0)$, $\frac{\partial}{\partial n} = \frac{\partial}{\partial r}$, so that for $y \in B_\epsilon(0)$ we have:

$$\phi(y)\frac{\partial g}{\partial n}(y) = \frac{-A}{\epsilon^2}\phi(y).$$

We therefore have:

$$\int_{\partial B_\epsilon(0)} \phi \frac{\partial g}{\partial n} d\sigma = (-4\pi A) \times \frac{1}{|\partial B_\epsilon(0)|} \int_{\partial B_\epsilon(0)} \phi(y) d\sigma_y.$$

Now, for any $\delta > 0$, since ϕ is continuous, there exists $\epsilon > 0$ such that $|\phi(y) - \phi(0)| < \delta$ for all $y \in B_\epsilon(0)$. We estimate:

$$\begin{aligned} \left| \frac{1}{|\partial B_\epsilon(0)|} \int_{\partial B_\epsilon(0)} \phi(y) d\sigma_y - \phi(0) \right| &= \left| \frac{1}{|\partial B_\epsilon(0)|} \int_{\partial B_\epsilon(0)} \left(\phi(y) - \phi(0)\right) d\sigma_y \right| \\ &\leqslant \frac{1}{|\partial B_\epsilon(0)|} \int_{\partial B_\epsilon(0)} |\phi(y) - \phi(0)|\, d\sigma_y \\ &< \delta \times \frac{1}{|\partial B_\epsilon(0)|} \int_{\partial B_\epsilon(0)} d\sigma_y = \delta. \end{aligned}$$

Thus, we conclude:

$$\lim_{\epsilon \to 0} \frac{1}{|\partial B_\epsilon(0)|} \int_{\partial B_\epsilon(0)} \phi(y) d\sigma_y = \phi(0).$$

Putting this all together, we have:

$$\Delta T_g [\phi] = (-4\pi A) \times \phi(0) = (-4\pi A)\, \delta_0 [\phi].$$

Thus if $A = -(4\pi)^{-1}$, we deduce that $\Delta T_g = \delta_0$, and so $G = T_g$ is a fundamental solution.

We conclude that if $f \in C_c^0(\mathbb{R}^3)$, then a solution of the distributional equation (3.7) is given by:

$$u = T_g \star T_f = T_{g \star f}.$$

If $f \in C_c^2(\mathbb{R}^3)$, the we know that $g \star f \in C^2(\mathbb{R}^3)$, and moreover that $w = g \star f$ is a solution of the classical equation (3.6). Thus, the solution we seek is:

$$w(x) = \frac{-1}{4\pi} \int_{\mathbb{R}^3} \frac{f(y)}{|x-y|} dy.$$

By examining this integral more carefully, it is possible to show that assuming $f \in C_c^2(\mathbb{R}^3)$ in order to get $w \in C^2(\mathbb{R}^3)$ is overkill. In fact, it suffices to have $f \in C_0^{0,\alpha}(\mathbb{R}^3)$ for some $0 < \alpha \leqslant 1$, where a function belongs to the Hölder space $C_0^{0,\alpha}(\mathbb{R}^3)$ if it has compact support, and there exists a constant C such that:

$$|f(x) - f(y)| \leqslant C\, |x - y|^\alpha,$$

holds for any $x, y \in \mathbb{R}^3$. This is the subject of the Schauder estimates for elliptic PDE (see "Elliptic Partial Differential Equations of Second Order", Gilbarg and Trudinger).

3.9 Distributions of compact support

Recall that as well as the space $\mathscr{D}(\Omega)$ of test functions, we also defined the space $\mathscr{E}(\Omega)$ consisting of smooth functions on Ω where the topology is such that a sequence $\{\phi_j\}_{j=1}^{\infty} \subset \mathscr{E}(\Omega)$ converges to zero in $\mathscr{E}(\Omega)$ if for any compact $K \subset \Omega$ and any multiindex α we have:

$$\sup_{x\in K} |D^\alpha \phi_j(x)| \to 0.$$

It is natural to define $\mathscr{E}'(\Omega)$ to be the set of continuous linear maps $\mathscr{E}(\Omega) \to \mathbb{C}$. Since the topology of $\mathscr{E}(\Omega)$ is induced by a metric, continuity is equivalent to sequential continuity. For a linear map $u : \mathscr{E}(\Omega) \to \mathbb{C}$ to belong to $\mathscr{E}'(\Omega)$, it is enough that:

$$\lim_{j\to\infty} u\,[\phi_j] = 0$$

for any sequence $\{\phi_j\}_{j=1}^{\infty} \subset \mathscr{E}(\Omega)$ which converges to zero in $\mathscr{E}(\Omega)$. Notice that if $u \in \mathscr{E}'(\Omega)$, then since $\mathscr{D}(\Omega)$ is a subspace of $\mathscr{E}(\Omega)$, $u : \mathscr{D}(\Omega) \to \mathbb{C}$ is a linear map. Moreover, we know that if $\{\phi_j\}_{j=1}^{\infty}$ is a sequence tending to zero in $\mathscr{D}(\Omega)$, then it also tends to zero in $\mathscr{E}(\Omega)$. Thus $u \in \mathscr{E}'(\Omega)$ is naturally an element of $\mathscr{D}'(\Omega)$, we have:

$$\mathscr{E}'(\Omega) \subset \mathscr{D}'(\Omega).$$

We are justified then in referring to elements of $\mathscr{E}'(\Omega)$ as distributions.

We can give a useful characterisation of continuity for linear maps from $\mathscr{E}(\Omega)$ to $\mathbb{C}$ as follows:

Lemma 3.13. *Suppose $u : \mathscr{E}(\Omega) \to \mathbb{C}$ is a linear map. Then u is continuous if and only if there is some compact $K \subset \Omega$, $N \in \mathbb{N}$ and $C > 0$ such that:*

$$|u[\phi]| \leqslant C \sup_{x\in K;|\alpha|\leqslant N} |D^\alpha \phi(x)|\,, \qquad \text{for all } \phi \in \mathscr{E}(\Omega). \tag{3.8}$$

Proof. First we show that (3.8) implies that u is continuous. Pick a sequence $\{\phi_j\}_{j=1}^{\infty} \subset \mathscr{E}(\Omega)$ which converges to zero in $\mathscr{E}(\Omega)$. This means that for all α and any compact $K' \subset \Omega$ we have

$$\sup_{x\in K'} |D^\alpha \phi(x)| \to 0.$$

In particular, this holds with $K' = K$ and for all α with $|\alpha| \leqslant N$, so as $j \to \infty$ we have:

$$|u[\phi_j]| \leqslant C \sup_{x\in K;|\alpha|\leqslant N} |D^\alpha \phi_j(x)| \to 0,$$

and so u is continuous.

To show the opposite implication, we assume that (3.8) does not hold for any K, N, C. We take an exhaustion of Ω by compact sets K_i, $i = 1, 2, \ldots$ such that $K_i \subset K_{i+1}^\circ$ and $\Omega = \cup_i K_i$ (see Lemma A.6). Then since (3.8) does not hold for any K, N, C, in particular it does not hold for $K = K_j$, $N = j$ and $C = j$. Thus there must exist $\phi_j \in \mathscr{E}(\Omega)$ such that:

$$|u[\phi_j]| > j \sup_{x\in K_j;|\alpha|\leqslant j} |D^\alpha \phi_j(x)|\,.$$

We define:

$$\psi_j(x) = \frac{\phi_j(x)}{|u[\phi_j]|}$$

Clearly $\psi_j \in \mathscr{E}(\Omega)$. We claim that $\psi_j \to 0$ in $\mathscr{E}(\Omega)$. To see this, fix a compact $K \subset \Omega$ and a multiindex α. For sufficiently large j, $K \subset K_j$ and $j \geqslant |\alpha|$. Thus:

$$\sup_{x\in K} |D^\alpha \phi_j(x)| \leqslant \sup_{x\in K_j; |\beta|\leqslant j} \left| D^\beta \phi_j(x) \right|$$

as a result, we can estimate:

$$\begin{aligned}\sup_{x\in K} |D^\alpha \psi_j(x)| &= \frac{1}{|u[\phi_j]|} \sup_{x\in K} |D^\alpha \phi_j(x)| \\ &< \frac{\sup_{x\in K} |D^\alpha \phi_j(x)|}{j \sup_{x\in K_j; |\beta|\leqslant j} |D^\beta \phi_j(x)|} < \frac{1}{j}\end{aligned}$$

We conclude that $D^\alpha \psi_j$ tends to zero on K, but since α and K were arbitrary, this implies $\psi_j \to 0$ in $\mathscr{E}(\Omega)$. However, $u[\psi_j] \not\to 0$ since $|u[\psi_j]| = 1$ by construction. Thus u is not continuous. This establishes that if u is continuous, then (3.8) must hold for some K, N, C. □

With this result in hand, we can give some examples of distributions $u \in \mathscr{E}'(\Omega)$.

Example 14. *i) If $f \in C^0_c(\Omega)$, then defining as usual:*

$$T_f[\phi] = \int_\Omega \phi(x) f(x) dx, \qquad \textit{for all } \phi \in \mathscr{E}(\Omega)$$

we have $T_f \in \mathscr{E}'(\Omega)$, since:

$$|T_f[\phi]| \leqslant \int_\Omega |f(x)|\, dx \sup_{y\in \operatorname{supp} f} |\phi(y)| \,.$$

If $f \in C^0(\Omega)$ but $\operatorname{supp} f$ is not compact, then $T_f \notin \mathscr{E}(\Omega)$.

ii) If $x \in \Omega$, then then setting:

$$\delta_x[\phi] = \phi(x) \qquad \textit{for all } \phi \in \mathscr{E}(\Omega),$$

we have $\delta_x \in \mathscr{E}'(\Omega)$. If K is any compact set containing x, then:

$$|\delta_x[\phi]| \leqslant \sup_{y\in K} |\phi(y)| \,.$$

iii) The map $u : \mathscr{D}(\mathbb{R}) \to \mathbb{C}$

$$u[\phi] = \sum_{m=-\infty}^{\infty} \phi(m) \tag{3.9}$$

does not define a distribution in $\mathscr{E}'(\mathbb{R})$. Indeed, the sum need not converge for any given element of $\mathscr{E}(\mathbb{R})$. For example, the constant function $\phi(x) = 1$ belongs to $\mathscr{E}(\mathbb{R})$, but the sum in (3.9) does not converge for this test function.

In these examples we see that elements of $\mathscr{E}'(\Omega)$ have compact support, while distributions with non-compact support do not appear to make sense when applied to elements of $\mathscr{E}(\Omega)$. In fact this is a more general result:

Theorem 3.14. *Suppose $u \in \mathscr{E}'(\Omega)$. Then $u \in \mathscr{D}'(\Omega)$ and u has compact support. Conversely, suppose that $u \in \mathscr{D}(\Omega)$ has compact support. Then there exists a unique $\tilde{u} \in \mathscr{E}'(\Omega)$ such that*

$$\tilde{u}[\phi] = u[\phi] \qquad \text{for all } \phi \in \mathscr{D}(\Omega).$$

We say that $\tilde{u}$ is the extension of u as a linear map on $\mathscr{E}(\Omega)$.

Proof. Suppose $u \in \mathscr{E}'(\Omega)$. We have already argued that $u \in \mathscr{D}'(\Omega)$ in a natural fashion, so it remains to show that $\operatorname{supp} u$ is compact. By Lemma 3.13 there exists some compact $K \subset \Omega$ and $N \in \mathbb{N}$, $C > 0$ such that:

$$|u[\phi]| \leqslant C \sup_{x \in K; |\alpha| \leqslant N} |D^\alpha \phi(x)|, \qquad \text{for all } \phi \in \mathscr{E}(\Omega).$$

Now suppose that $\operatorname{supp} \phi \subset \Omega \setminus K$. From the estimate above, we have that $u[\phi] = 0$. Thus $\operatorname{supp} u \subset K$ and we must have that $\operatorname{supp} u$ is compact.

Now suppose that $u \in \mathscr{D}'(\Omega)$ has compact support. By Lemma 1.14 we know that there exists $\chi \in C^\infty_c(\Omega)$ such that $\chi = 1$ on $\operatorname{supp} u$. For $\phi \in \mathscr{E}(\Omega)$, we define:

$$\tilde{u}[\phi] = u[\chi\phi].$$

This makes sense because $\chi\phi$ is compactly supported in Ω, so $u[\chi\phi]$ is defined. If $\phi_j \to 0$ in $\mathscr{E}(\Omega)$, then $\chi\phi_j \to 0$ in $\mathscr{D}(\Omega)$ (see Exercise 3.9). Thus $\tilde{u} \in \mathscr{E}'(\Omega)$. We also note that if $\phi \in \mathscr{D}(\Omega)$, then $\chi\phi - \phi$ has support in $\Omega \setminus \operatorname{supp} u$. Thus:

$$0 = u\left[\chi\phi - \phi\right] = \tilde{u}[\phi] - u[\phi],$$

so that $\tilde{u}$ and u agree on $\mathscr{D}(\Omega)$. It remains to show that $\tilde{u}$ is unique. Suppose $\tilde{v} \in \mathscr{E}'(\Omega)$ satisfies

$$\tilde{u}[\phi] = \tilde{v}[\phi] \qquad \text{for all } \phi \in \mathscr{D}(\Omega).$$

Let $\psi \in \mathscr{E}(\Omega)$ be arbitrary. We can find $\phi_j \in \mathscr{D}(\Omega)$ such that $\phi_j \to \psi$ in $\mathscr{E}(\Omega)$ (see Exercise 3.9). We have, using the continuity of $\tilde{u}$, $\tilde{\psi}$:

$$\tilde{u}[\psi] = \lim_{j\to\infty} \tilde{u}[\phi_j] = \lim_{j\to\infty} \tilde{v}[\phi_j] = \tilde{v}[\psi].$$

Thus $\tilde{u} = \tilde{v}$, since ψ was arbitrary. □

Exercise(∗). a) Suppose that $\{\phi_j\}_{j=1}^\infty \subset \mathscr{E}(\Omega)$ is a sequence such that $\phi_j \to \phi$ in $\mathscr{E}(\Omega)$, and $\chi \in \mathscr{D}(\Omega)$. Show that

$$\chi\phi_j \to \chi\phi \quad \text{in } \mathscr{D}(\Omega).$$

b) Show that if $\psi \in \mathscr{E}(\Omega)$, then there exists a sequence $\{\phi_j\}_{j=1}^\infty \subset \mathscr{D}(\Omega)$ such that $\phi_j \to \psi$ in $\mathscr{E}(\Omega)$.
[Hint: Take an exhaustion of Ω by compact sets and apply Lemma 1.14]

3.10 Tempered distributions

The final class of distributions that we shall consider are the *tempered* distributions. The space of tempered distributions arises as the continuous dual of $\mathscr{S}$, the Schwartz space of rapidly decreasing functions. Recall that $\phi \in \mathscr{S}$ if $\phi \in C^\infty(\mathbb{R}^n)$ and for any multiindex and any $N \in \mathbb{N}$ we have:

$$\sup_{x\in\mathbb{R}^n} \left|(1+|x|)^N D^\alpha \phi(x)\right| < \infty.$$

We say that a sequence $\{\phi_j\}_{j=1}^\infty \subset \mathscr{S}$ tends to zero if:

$$\sup_{x\in\mathbb{R}^n} \left|(1+|x|)^N D^\alpha \phi_j(x)\right| \to 0$$

for all $N \in \mathbb{N}$ and all multiindices α.

We define $\mathscr{S}'$ to be the continuous dual space of $\mathscr{S}$. That is to say, $u \in \mathscr{S}'$ if $u : \mathscr{S} \to \mathbb{C}$ is a continuous linear map. Since the topology of $\mathscr{S}$ can be induced by a metric, again sequential continuity is equivalent to continuity. For a linear map $u : \mathscr{S} \to \mathbb{C}$ to belong to $\mathscr{S}'$, it is enough that:

$$\lim_{j\to\infty} u[\phi_j] \to 0,$$

for any sequence $\{\phi_j\}_{j=1}^\infty \subset \mathscr{S}$ which converges to zero in $\mathscr{S}$.

Lemma 3.15. *Suppose $u : \mathscr{S} \to \mathbb{C}$ is a linear map. Then u is continuous if and only if there exist $N, k \in \mathbb{N}$ and $C > 0$ such that:*

$$|u[\phi]| \leqslant C \sup_{x\in\mathbb{R}^n;|\alpha|\leqslant k} \left|(1+|x|)^N D^\alpha \phi(x)\right|, \quad \textit{for all } \phi \in \mathscr{S}. \tag{3.10}$$

Proof. See Exercise 3.6 □

Example 15. *i) Suppose $f \in L^1_{loc.}(\mathbb{R}^n)$ and there exist $C > 0, N \in \mathbb{N}$ such that*

$$|f(x)| \leqslant C(1+|x|)^N.$$

Then $T_f \in \mathscr{S}'$

ii) The map:

$$\phi \mapsto \int_{\mathbb{R}^n} e^{|x|^2}\phi(x)dx$$

does not define a tempered distribution.

iii) For $\phi \in C^\infty(\mathbb{R})$, and $N \in \mathbb{N}$ we set:

$$u[\phi] = \sum_{m=-\infty}^{\infty} m^N \phi(m).$$

The sum converges for $\phi \in \mathscr{S}$, and defines a tempered distribution.

From these examples, and Lemma 3.15, we see that the tempered distributions are those that don't grow too much near infinity.

Exercise 3.6. Suppose $u : \mathscr{S} \to \mathbb{C}$ is a linear map. Show that u is continuous if and only if there exist $N, k \in \mathbb{N}$ and $C > 0$ such that:

$$|u[\phi]| \leqslant C \sup_{x \in \mathbb{R}^n; |\alpha| \leqslant k} \left|(1 + |x|)^N D^\alpha \phi(x)\right|, \qquad \text{for all } \phi \in \mathscr{S}.$$

Exercise 3.7. Suppose $u \in \mathscr{D}'(\mathbb{R}^n)$ is *positive*, i.e. $u[\phi] \geq 0$ for all $\phi \in \mathscr{D}(\mathbb{R}^n)$ with $\phi \geq 0$. Show that u has order 0. (*) Deduce that $u[\phi] = \int_{\mathbb{R}^n} \phi d\mu$ for some regular measure μ.

Exercise($*$). Let $(a_j)_{j=-\infty}^{\infty}$ be a sequence of complex numbers. Define for $\phi \in C^\infty(\mathbb{R})$:

$$u[\phi] = \sum_{j=-\infty}^{\infty} a_j \phi(j)$$

provided that the sum converges. Give necessary and sufficient conditions on a_j such that: a) $u \in \mathscr{E}'(\mathbb{R})$, b) $u \in \mathscr{S}'$, c) $u \in \mathscr{D}'(\mathbb{R})$.

Chapter 4

The Fourier Transform and Sobolev Spaces

4.1 The Fourier transform on $L^1(\mathbb{R}^n)$

The Fourier transform is an extremely powerful tool across the full range of mathematics. Loosely speaking, the idea is to consider a function on $\mathbb{R}^n$ as a superposition of plane waves with different frequencies. For $f \in L^1(\mathbb{R}^n)$, we define the Fourier transform $\hat{f} : \mathbb{R}^n \to \mathbb{C}$ by:

$$\mathcal{F}[f](\xi) = \hat{f}(\xi) := \int_{\mathbb{R}^n} f(x) e^{-ix\cdot\xi} dx.$$

Sine $\left|f(x)e^{-ix\cdot\xi}\right| \leqslant |f(x)|$, the integral is absolutely convergent, and $\hat{f}(\xi)$ makes sense for each $\xi \in \mathbb{R}^n$.

Example 16. *i) Suppose $f \in L^1(\mathbb{R})$ is the "top hat" function, defined by:*

$$f(x) = \begin{cases} 1 & -1 < x < 1, \\ 0 & |x| \geqslant 1. \end{cases}$$

We calculate:

$$\hat{f}(\xi) = \int_{-1}^{1} e^{-ix\xi} dx = \left[\frac{e^{-ix\xi}}{-i\xi}\right]_{-1}^{1} = 2\frac{\sin\xi}{\xi}$$

Notice that $\hat{f}(\xi)$ is continuous (in fact smooth) on $\mathbb{R}$. We also have $\hat{f}(\xi) \to 0$ as $\xi \to \infty$.

ii) Suppose $f \in L^1(\mathbb{R})$ is defined by:

$$f(x) = \begin{cases} e^{x} & x < 0, \\ e^{-x} & x \geqslant 0. \end{cases}$$

Then:

$$\begin{aligned}\hat{f}(\xi) &= \int_{-\infty}^{0} e^{x(1-i\xi)}dx + \int_{0}^{\infty} e^{x(-1-i\xi)}dx \\ &= \left[\frac{e^{x(1-i\xi)}}{1-i\xi}\right]_{-\infty}^{0} + \left[\frac{e^{x(-1-i\xi)}}{-1-i\xi}\right]_{0}^{\infty} \\ &= \frac{1}{1-i\xi} + \frac{1}{1+i\xi} = \frac{2}{1+\xi^2}\end{aligned}$$

Again, notice that $\hat{f}$ is smooth and decays for large ξ.

iii) Consider $g \in L^1(\mathbb{R})$ given by

$$g(x) = \frac{1}{1+x^2}.$$

We have:

$$\hat{g}(\xi) = \int_{-\infty}^{\infty} \frac{e^{-ix\xi}}{1+x^2}dx$$

We can consider this as a limit of contour integrals:

$$\hat{g}(\xi) = \lim_{R\to\infty} \int_{\gamma_R} \frac{e^{-iz\xi}}{1+z^2}dx.$$

Where $\gamma_R = \{\Im(z) = 0,\ |\Re(z)| < R\}$. For $\xi \geqslant 0$, we can close the contour with a semi-circle in the lower half-plane, and we pick up a contribution from the pole at $z = -i$. The contribution from the curved part of the contour tends to zero as $R \to \infty$ by Jordan's lemma, and we find:

$$\hat{g}(\xi) = \pi e^{-\xi}, \quad \xi \geqslant 0.$$

For $\xi < 0$, we close the contour in the upper half-plane, picking up a contribution from the pole at $z = i$ and again discard the contribution from the curved part of the contour in the limit. We find:

$$\hat{g}(\xi) = \pi e^{\xi}, \quad \xi < 0.$$

In conclusion, we have:

$$\hat{g}(\xi) = \begin{cases} \pi e^{\xi} & \xi < 0, \\ \pi e^{-\xi} & \xi \geqslant 0. \end{cases}$$

iv) Consider now for $x \in \mathbb{R}^n$ the Gaussian $f(x) = e^{-\frac{1}{2}|x|^2}$. We calculate:

$$\begin{aligned}\hat{f}(\xi) &= \int_{\mathbb{R}^n} e^{-\frac{1}{2}|x|^2 - i\xi\cdot x}dx \\ &= \int_{\mathbb{R}^n} e^{-\frac{1}{2}(x-i\xi)\cdot(x-i\xi)-\frac{1}{2}|\xi|^2}dx \\ &= e^{-\frac{1}{2}|\xi|^2}\left(\int_{\mathbb{R}} e^{-\frac{1}{2}(x_1-i\xi_1)^2}dx_1\right)\cdots\left(\int_{\mathbb{R}} e^{-\frac{1}{2}(x_n-i\xi_n)^2}dx_n\right)\end{aligned}$$

By shifting a contour in the complex plane, which is justified since e^{-z^2} is entire and rapidly decaying as z approaches infinity along any line parallel to the real axis, we can show that:

$$\int_{\mathbb{R}} e^{-\frac{1}{2}(x_1 - i\xi_1)^2} dx_1 = \int_{\mathbb{R}} e^{-\frac{1}{2}x_1^2} dx_1 = \sqrt{2\pi}.$$

We deduce that:

$$\hat{f}(\xi) = (2\pi)^{\frac{n}{2}} e^{-\frac{1}{2}|\xi|^2}$$

Notice that this is (as a function) equal to f up to a factor.

We will make some casual observations at this stage. In all of our examples, we saw that the Fourier transformed function decays towards infinity. For examples iii), iv) we see very rapid (exponential) decay of the Fourier transform, while in examples i), ii) the decay is only polynomial. In all examples the transformed function is continuous. In examples i), ii), iv) it is in fact smooth, while for iii) the Fourier transform has a discontinuous first derivative. Reflecting on this, one sees that these two features appear to be dual to one another: if f is smooth, then $\hat{f}$ has rapid decay towards infinity. If f decays rapidly near infinity, then $\hat{f}$ is smooth. This is in fact a general feature of the Fourier transform smoothness and decay are dual to one another under the transform.

We shall now make some of these observations more precise.

Lemma 4.1 (Riemann-Lebesgue Lemma)**.** *Suppose $f \in L^1(\mathbb{R}^n)$. Then $\hat{f} \in C^0(\mathbb{R}^n)$ with the estimate:*

$$\sup_{\xi \in \mathbb{R}^n} \left|\hat{f}(\xi)\right| \leqslant \|f\|_{L^1} \tag{4.1}$$

and moreover $\hat{f}(\xi) \to 0$ as $|\xi| \to \infty$.

Proof. To establish the continuity of $\hat{f}$, we use the continuity of the exponential, together with the dominated convergence theorem. Let $\{\xi_j\}_{j=1}^{\infty}$ be any sequence with $\xi_j \to \xi$ as $j \to \infty$. Recalling the definition of the integral, we have:

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

Now, clearly for $x \in \mathbb{R}^n$ we have:

$$f(x)e^{-ix\cdot\xi_j} \to f(x)e^{-ix\cdot\xi}, \qquad \text{as } j \to \infty$$

so we have pointwise convergence of the integrand. We can also estimate:

$$\left|f(x)e^{-ix\cdot\xi_j}\right| \leqslant |f(x)|$$

so the integrand is dominated by an integrable function, since $f \in L^1(\mathbb{R}^n)$. Applying the Dominated Convergence Theorem, we conclude:

$$\hat{f}(\xi_j) \to \hat{f}(\xi), \qquad \text{as } j \to \infty.$$

This implies that $\hat{f}(\xi)$ is continuous. We can readily estimate:

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

This establishes the first part of the Lemma. To establish the second part, we make use of an approximation argument. Chapter 1 that if $f \in L^1(\mathbb{R}^n)$, we can approximate f by an element of $C_0^\infty(\mathbb{R}^n)$. Given $\epsilon > 0$, there exists $f_\epsilon \in C_0^\infty(\mathbb{R}^n)$ with

$$\|f - f_\epsilon\|_{L^1} < \frac{\epsilon}{2}.$$

Now, in the integral for $\hat{f}_\epsilon$ we can integrate by parts::

$$\begin{aligned}
\hat{f}_\epsilon(\xi) &= \int_{\mathbb{R}^n} f_\epsilon(x)e^{-ix\cdot\xi}dx \\
&= \int_{\mathbb{R}^n} f_\epsilon(x)\,\mathrm{div}\left(\frac{\xi}{-i\,|\xi|^2}e^{-ix\cdot\xi}\right)dx \\
&= -\int_{\mathbb{R}^n}\frac{\xi}{-i\,|\xi|^2}\cdot Df_\epsilon(x)e^{-ix\cdot\xi}dx
\end{aligned}$$

so that for each $i = 1, \ldots, n$ we have, by the Cauchy-Schwarz inequality:

$$\begin{aligned}
\left|\hat{f}_\epsilon(\xi)\right| &= \left|\int_{\mathbb{R}^n}\frac{\xi}{i\,|\xi|^2}\cdot Df_\epsilon(x)e^{-ix\cdot\xi}dx\right| \\
&\leqslant \int_{\mathbb{R}^n}\left|\frac{\xi}{i\,|\xi|^2}\cdot Df_\epsilon(x)e^{-ix\cdot\xi}\right|dx \\
&\leqslant \int_{\mathbb{R}^n}\frac{1}{|\xi|}\,|Df_\epsilon(x)|\,dx \\
&= \frac{1}{|\xi|}\Big\|\,|Df_\epsilon(x)|\,\Big\|_{L^1}
\end{aligned} \tag{4.2}$$

From this, we conclude that there exists $R > 0$ such that if $|\xi| > R$, we have $\left|\hat{f}_\epsilon(\xi)\right| < \frac{\epsilon}{2}$. For $|\xi| > R$ we calculate:

$$\begin{aligned}
\left|\hat{f}(\xi)\right| &= \left|\hat{f}(\xi) - \hat{f}_\epsilon(\xi) + \hat{f}_\epsilon(\xi)\right| \\
&\leqslant \left|\hat{f}_\epsilon(\xi)\right| + \left|\hat{f}(\xi) - \hat{f}_\epsilon(\xi)\right| \\
&\leqslant \left|\hat{f}_\epsilon(\xi)\right| + \|f - f_\epsilon\|_{L^1} < \epsilon.
\end{aligned}$$

In the last line, we have used (4.1), together with the linearity of the Fourier transform. Since $\epsilon > 0$ was arbitrary, we have shown that $\left|\hat{f}(\xi)\right| \to 0$. □

Remark. *The argument above is another example of an approximation argument where one first proves the result on a suitably nice dense subset then extends to the full space by* *continuity. In this case, we are using the fact that the Fourier transform is a bounded (hence continuous) linear operator from $L^1(\mathbb{R}^n)$ to the Banach space of continuous functions decaying at infinity equipped with the uniform norm. The dense set is $C_0^\infty(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$.*

Another tactic which we used here was to use integration by parts to exploit the rapid oscillations in the $e^{-ix\cdot\xi}$ factor when $|\xi|$ is large.

One might be tempted to infer from (4.2) that $\left|\hat{f}(\xi)\right| \leqslant C(1+|\xi|)^{-1}$. While this is true for each f_ϵ approximating f, in general the constant C will grow larger and larger as $\epsilon \to 0$, so we cannot quite come to this conclusion.

Exercise(∗). For $\xi \in \mathbb{R}^n$, define $e_\xi(x) = e^{i\xi\cdot x}$. Show that $T_{e_\xi} \in \mathscr{S}'$, and that:

$$T_{e_\xi} \to 0, \qquad \text{as } |\xi| \to \infty$$

in the weak-∗ topology of $\mathscr{S}'$.

We shall prove some important properties of the Fourier transform. Recall that $\tau_y f(x) = f(x-y)$, and introduce the character $e_y(x) = e^{iy\cdot x}$.

Lemma 4.2 (Properties of the Fourier transform). *i) Suppose $f \in L^1(\mathbb{R}^n)$, $x \in \mathbb{R}^n$, $\lambda > 0$ and $f_\lambda(y) = \lambda^{-n} f(\lambda^{-1}y)$. Then*

$$\hat{f}_\lambda(\xi) = \hat{f}(\lambda\xi) \qquad \widehat{(e_x f)}(\xi) = \tau_x \hat{f}(\xi) \qquad \widehat{\tau_x f}(\xi) = e_{-x}(\xi)\hat{f}(\xi)$$

ii) Suppose $f, g \in L^1(\mathbb{R}^n)$. Then $f \star g \in L^1(\mathbb{R}^n)$ and:

$$\widehat{f \star g}(\xi) = \hat{f}(\xi)\hat{g}(\xi).$$

Proof. i) Writing out the expression for $\hat{f}_\lambda(\xi)$, and changing the integration variable to $z = \lambda^{-1}x$, we see

$$\hat{f}_\lambda(\xi) = \int_{\mathbb{R}^n} f_\lambda(x) e^{-i\xi\cdot x} dx = \int_{\mathbb{R}^n} f(\lambda^{-1}x) e^{-i\xi\cdot x}\lambda^{-n} dx = \int_{\mathbb{R}^n} f(y) e^{-i\lambda\xi\cdot z} dz = \hat{f}(\lambda\xi).$$

Next, we calculate:

$$\widehat{(e_x f)}(\xi) = \int_{\mathbb{R}^n} e^{ix\cdot y} f(y) e^{-i\xi\cdot y} dy = \int_{\mathbb{R}^n} f(y) e^{-i(\xi - x)\cdot y} dy = \tau_x \hat{f}(\xi).$$

Finally, we have:

$$\widehat{\tau_x f}(\xi) = \int_{\mathbb{R}^n} f(y-x) e^{-i\xi\cdot y} dy = \int_{\mathbb{R}^n} f(z) e^{-i\xi\cdot(z+x)} dz = e^{-i\xi\cdot x}\int_{\mathbb{R}^n} f(z) e^{-i\xi\cdot z} dz = e_{-x}(\xi)\hat{f}(\xi),$$

where we have used the substitution $z = y - x$.

ii) First we show that $f \star g \in L^1(\mathbb{R}^n)$. To see this, we first estimate:

$$|f \star g(x)| = \left|\int_{\mathbb{R}^n} f(y) g(x-y) dy\right| \leqslant \int_{\mathbb{R}^n} |f(y) g(x-y)|\, dy$$

Integrating and applying Tonelli's theorem, we have:

$$\begin{aligned}
\|f\star g\|_{L^1} &\leqslant \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}|f(y)g(x-y)|\,dy\right)dx\\
&= \int_{\mathbb{R}^n}|f(y)|\left(\int_{\mathbb{R}^n}|g(x-y)|\,dx\right)dy\\
&= \int_{\mathbb{R}^n}|f(y)|\,\|g\|_{L^1}\,dy = \|f\|_{L^1}\,\|g\|_{L^1}
\end{aligned}$$

Now, we can calculate the Fourier transform:

$$\begin{aligned}
\widehat{f\star g}(\xi) &= \int_{\mathbb{R}^n} f\star g(x)e^{-i\xi\cdot x}dx\\
&= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} f(y)g(x-y)dy\right)e^{-i\xi\cdot x}dx\\
&= \int_{\mathbb{R}^n} f(y)\left(\int_{\mathbb{R}^n} g(x-y)e^{-i\xi\cdot x}dx\right)dy\\
&= \int_{\mathbb{R}^n} f(y)\widehat{\tau_y g}(\xi)dy\\
&= \int_{\mathbb{R}^n} f(y)\widehat{g}(\xi)e^{-i\xi\cdot y}dy = \hat{f}(\xi)\hat{g}(\xi)
\end{aligned}$$

□

Exercise($*$). Calculate the Fourier transform of the following functions $f \in L^1(\mathbb{R})$:

a) $f(x) = \dfrac{\sin x}{1+x^2}$.

b) $f(x) = \dfrac{1}{\epsilon^2+x^2}$, for $\epsilon > 0$ a constant.

c) $f(x) = \sqrt{\dfrac{\sigma}{t}}e^{-\sigma\frac{(x-y)^2}{t}}$, where $\sigma > 0$, $t > 0$ and y are constants.

*d) $f(x) = \dfrac{1}{\cosh x}$.

We saw with the examples that there is a duality between the decay of a function and the regularity of its Fourier transform and vice versa. To make this more precise we prove the following result, which tells us, roughly speaking, that the Fourier transform swaps coordinate functions x_j multiplying f for derivatives iD_j acting on $\hat{f}$.

Theorem 4.3. *i) Suppose $f \in C^1(\mathbb{R}^n)$ and that $f, D_j f \in L^1(\mathbb{R}^n)$ for all $j = 1, \dots n$. Then*

$$\widehat{D_j f}(\xi) = i\xi_j\hat{f}(\xi)$$

ii) Suppose $(1+|x|)f \in L^1(\mathbb{R}^n)$. *Then* $\hat{f} \in C^1(\mathbb{R}^n)$, *and:*

$$D_j \hat{f}(\xi) = -i\,\widehat{x_j f}(\xi)$$

Proof. i) We again appeal to an approximation result. For $f \in C^1(\mathbb{R}^n)$ with $f, D_i f \in L^1(\mathbb{R}^n)$, then for any $\epsilon > 0$ there exists $f_\epsilon \in C^1_0(\mathbb{R}^n)$ such that $\|f - f_\epsilon\|_{L^1} < \epsilon$ and $\|D_j f - D_j f_\epsilon\|_{L^1} < \epsilon$. Integrating by parts, we readily calculate:

$$\begin{aligned}\widehat{D_j f_\epsilon}(\xi) &= \int_{\mathbb{R}^n} D_j f_\epsilon(x) e^{-i\xi\cdot x} dx \\ &= -\int_{\mathbb{R}^n} f_\epsilon(x) D_j(e^{-i\xi\cdot x}) dx \\ &= i\xi_j \int_{\mathbb{R}^n} f_\epsilon(x) e^{-i\xi\cdot x} dx\end{aligned}$$

so that $\widehat{D_j f_\epsilon}(\xi) = i\xi_j \hat{f}_\epsilon(\xi)$. Now, we calculate:

$$\begin{aligned}\left|\widehat{D_j f}(\xi) - i\xi_j \hat{f}(\xi)\right| &= \left|\widehat{D_j f}(\xi) - \widehat{D_j f_\epsilon}(\xi) + i\xi_j \hat{f}_\epsilon(\xi) - i\xi_j \hat{f}(\xi)\right| \\ &\leqslant \|D_j f - D_j f_\epsilon\|_{L^1} + |\xi|\,\|f - f_\epsilon\|_{L^1} \\ &\leqslant \epsilon(1+|\xi|)\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we must have that $\left|\widehat{D_j f}(\xi) - i\xi_j \hat{f}(\xi)\right| = 0$, and the result follows.

ii) From the condition on f it is clear that $x_j f \in L^1(\mathbb{R}^n)$, so $-i\,\widehat{x_j f}$ is continuous. It suffices to prove then that:

$$\Delta_j^{h_k} \hat{f}(\xi) \to -i\,\widehat{x_j f}(\xi), \qquad \text{as } k \to \infty$$

for any sequence $\{h_k\}_{k=1}^\infty \subset \mathbb{R}$ with $h_k \to 0$. We calculate:

$$\Delta_j^{h_k} \hat{f}(\xi) = \frac{1}{h_k}\left(\hat{f}(\xi + h_k e_j) - \hat{f}(\xi)\right) = \int_{\mathbb{R}^n} f(x) e^{-ix\cdot\xi}\left(\frac{e^{-ix_j h_k} - 1}{h_k}\right) dx.$$

Now for $x \in \mathbb{R}^n$ we have:

$$f(x)e^{-ix\cdot\xi}\left(\frac{e^{-ix_j h_k} - 1}{h_k}\right) \to -ix_j f(x) e^{-ix\cdot\xi}$$

as $k \to \infty$. Noting that $\left|e^{i\theta} - 1\right| = 2\left|\sin\frac{\theta}{2}\right| \leqslant \theta$ for any $\theta \in \mathbb{R}$, we have that:

$$\left|f(x)e^{-ix\cdot\xi}\left(\frac{e^{-ix_j h_k} - 1}{h_k}\right)\right| \leqslant |x_j f(x)|$$

where the right hand side is integrable. By the Dominated Convergence Theorem, we have:

$$\lim_{k\to\infty} \Delta_j^{h_k} \hat{f}(\xi) = \int_{\mathbb{R}^n} -ix_j f(x) e^{-ix\cdot\xi} dx = -i\,\widehat{x_j f}(\xi).$$

We deduce that $\hat{f} \in C^1(\mathbb{R}^n)$.

□

Exercise($*$). Suppose $f \in C^1(\mathbb{R}^n)$ and that $f, D_j f \in L^1(\mathbb{R}^n)$. Fix $\epsilon > 0$. Show that there exists $f_\epsilon \in C_0^1(\mathbb{R}^n)$ such that

$$\|f - f_\epsilon\|_{L^1} + \|D_j f - D_j f_\epsilon\|_{L^1} < \frac{\epsilon}{2}.$$

Corollary 4.4. *i) Suppose $f \in C^k(\mathbb{R}^n)$ and $D^\alpha f \in L^1(\mathbb{R}^n)$ for $|\alpha| \leqslant k$. Then there is some constant $C_k > 0$ depending only on k such that:*

$$\sup_{\xi \in \mathbb{R}^n} \left|(1+|\xi|)^k \hat{f}(\xi)\right| \leqslant C_k \sum_{|\alpha| \leqslant k} \|D^\alpha f\|_{L^1}$$

ii) Suppose $(1+|x|)^k f \in L^1(\mathbb{R}^n)$. Then $\hat{f} \in C^k(\mathbb{R}^n)$ and for any $|\alpha| \leqslant k$ we have:

$$\sup_{\xi \in \mathbb{R}^n} \left|D^\alpha \hat{f}(\xi)\right| \leqslant \left\|(1+|x|)^k f\right\|_{L^1}$$

iii) The Fourier transform is a continuous linear map from $\mathscr{S}$ into $\mathscr{S}$:

$$\mathcal{F} : \mathscr{S} \to \mathscr{S}.$$

Proof. i) First we note the algebraic fact that for any k there is some constant C_k such that[1]:

$$(1+|\xi|)^k \leqslant C_k \sum_{|\alpha| \leqslant k} |\xi^\alpha|$$

holds for any $\xi \in \mathbb{R}^n$. Repeatedly applying the part $i)$ of Theorem 4.3 we know that:

$$i^{|\alpha|} \xi^\alpha \hat{f}(\xi) = \widehat{D^\alpha f}(\xi).$$

We therefore have:

$$(1+|\xi|)^k \left|\hat{f}(\xi)\right| \leqslant C_k \sum_{|\alpha| \leqslant k} \left|i^{|\alpha|} \xi^\alpha \hat{f}(\xi)\right| = C_k \sum_{|\alpha| \leqslant k} \left|\widehat{D^\alpha f}(\xi)\right|$$

taking the supremum over $\xi \in \mathbb{R}^n$ and applying the estimate (4.1) we conclude

$$\sup_{\xi \in \mathbb{R}^n} \left|(1+|\xi|)^k \hat{f}(\xi)\right| \leqslant C_k \sum_{|\alpha| \leqslant k} \|D^\alpha f\|_{L^1}.$$

ii) By iterating part $ii)$ of Theorem 4.3 we have that for $|\alpha| \leqslant k$:

$$D^\alpha \hat{f}(\xi) = (-i)^{|\alpha|} \widehat{x^\alpha f}(\xi).$$

Taking the supremum of the absolute value over $\xi \in \mathbb{R}^n$ and applying the estimate (4.1) we have:

$$\sup_{\xi \in \mathbb{R}^n} \left|D^\alpha \hat{f}(\xi)\right| \leqslant \|x^\alpha f\|_{L^1} \leqslant \left\|(1+|x|)^k f\right\|_{L^1}$$

[1] recall that $\xi^\alpha := \xi_1^{\alpha_1} \xi_2^{\alpha_2} \cdots \xi_n^{\alpha_n}$

iii) Note that if:

$$\sup_{x\in\mathbb{R}^n} (1+|x|)^N |f(x)| < K,$$

we have:

$$\|f\|_{L^1} = \int_{\mathbb{R}^n} |f(x)|\, dx \leqslant K \int_{\mathbb{R}^n} \frac{1}{(1+|x|)^N} dx < \infty$$

provided $N > n$. Thus in particular if $f \in \mathscr{S}$ then there exists some constant C_n such that:

$$\left\|(1+|x|)^M D^\alpha f\right\|_{L^1} \leqslant C_n \sup_{x\in\mathbb{R}^n} (1+|x|)^{M+n+1} |D^\alpha f(x)|$$

for all $M \in \mathbb{N}$ and all multi-indices α. Applying the previous two parts we conclude that $\hat{f} \in C^\infty(\mathbb{R}^n)$ and:

$$\sup_{\xi\in\mathbb{R}^n, |\beta|\leqslant M} (1+|\xi|)^N \left|D^\beta \hat{f}(\xi)\right| \leqslant C_{N,M,n} \sup_{x\in\mathbb{R}^n, |\alpha|\leqslant N} (1+|x|)^{M+n+1} |D^\alpha f(x)|$$

For some constant $C_{N,M,n}$ depending only on N, M, n. Thus $\hat{f} \in \mathscr{S}$. Moreover, if $\{f_j\}_{j=1}^\infty \subset \mathscr{S}$ is a sequence with $f_j \to 0$ in $\mathscr{S}$, then $\hat{f}_j \to 0$ in $\mathscr{S}$, so that $\mathcal{F}$ is continuous.

□

Notice that while the Fourier transform maps $\mathscr{S}$ to itself, the same is not true of $\mathscr{D}(\mathbb{R}^n)$. Suppose $f \in C_0^\infty(\mathbb{R}^n)$, then provided $\operatorname{supp} f \subset K$ for K a compact set we have:

$$\hat{f}(\xi) = \int_K f(x) e^{-ix\cdot\xi} dx$$

By repeatedly differentiating, it is possible to show that $\hat{f}$ is in fact *real analytic*, and hence $\hat{f}$ cannot vanish on any open set without vanishing everywhere. In particular, $\hat{f}$ cannot vanish outside a compact set.

Exercise 3.8. Suppose $f \in L^1(\mathbb{R}^n)$, with $\operatorname{supp} f \subset B_R(0)$ for some $R > 0$.

a) Show that $\hat{f} \in C^\infty(\mathbb{R}^n)$ and for any multi-index:

$$\sup_{\xi\in\mathbb{R}^n} \left|D^\alpha \hat{f}(\xi)\right| \leqslant R^{|\alpha|} \|f\|_{L^1}$$

b) (*)Show that $\hat{f}$ is real analytic, with an infinite radius of convergence, i.e.:

$$\hat{f}(\xi) = \sum_\alpha D^\alpha \hat{f}(0) \frac{\xi^\alpha}{\alpha!}$$

holds for all $\xi \in \mathbb{R}^n$. Deduce that if $\hat{f}(\xi)$ vanishes on an open set, it must vanish everywhere.

You may assume the following form of Taylor's theorem. Suppose $g \in C^{k+1}(\overline{B_r(0)})$. *Then for* $x \in B_r(0)$:

$$g(x) = \sum_{|\alpha| \leqslant k} D^\alpha g(0) \frac{x^\alpha}{\alpha!} + \sum_{|\beta|=k+1} R_\beta(x) x^\beta$$

where the remainder $R_\beta(x)$ *satisfies the following estimate in* $B_r(0)$:

$$|R_\beta(x)| \leqslant \frac{1}{\beta!} \max_{|\alpha|=|\beta|} \max_{y \in \overline{B_r(0)}} |D^\alpha g(y)| \,.$$

Exercise 3.9. Recall that $L^\infty(\mathbb{R}) = L^1(\mathbb{R})'$. Consider the sequence $(f_n)_{n=1}^\infty$, where $f_n \in L^\infty(\mathbb{R})$ is given by $f_n(x) = \sin(nx)$. Show that $f_n \overset{*}{\rightharpoonup} 0$. Show that $f_n^2 \overset{*}{\rightharpoonup} g$ for some $g \in L^\infty(\mathbb{R})$ which you should find.

To complete this section, we are going to establish the invertibility of the Fourier transform, under some reasonable assumptions on f and $\hat{f}$. In particular, this will permit us to show that $\mathcal{F} : \mathscr{S} \to \mathscr{S}$ is in fact a bijection.

Theorem 4.5 (Fourier inversion theorem)**.** *Suppose* $f \in L^1(\mathbb{R}^n)$, *and assume* $\hat{f} \in L^1(\mathbb{R}^n)$, *then for almost every* x:

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{ix\cdot\xi} d\xi. \tag{4.3}$$

Proof. We shall establish the result by looking at the limit $\epsilon \to 0$ of

$$I_\epsilon(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{-\frac{1}{2}\epsilon^2|\xi|^2} e^{ix\cdot\xi} d\xi.$$

in two different ways. Firstly note that for $\xi \in \mathbb{R}^n$ we have:

$$\hat{f}(\xi) e^{-\frac{1}{2}\epsilon^2|\xi|^2} e^{ix\cdot\xi} \to \hat{f}(\xi) e^{ix\cdot\xi}.$$

Moreover, we can estimate

$$\left| \hat{f}(\xi) e^{-\frac{1}{2}\epsilon^2|\xi|^2} e^{ix\cdot\xi} \right| \leqslant \left| \hat{f}(\xi) \right|$$

so that the integrand is dominated by an integrable function. Thus by the Dominated Convergence Theorem we have:

$$I_\epsilon(x) \to \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{ix\cdot\xi} d\xi, \qquad \text{as } \epsilon \to 0.$$

On the other hand, we have, using Fubini's thorem:

$$\begin{aligned} I_\epsilon(x) &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(y) e^{-i\xi\cdot y} dy \right) e^{-\frac{1}{2}\epsilon^2|\xi|^2} e^{ix\cdot\xi} d\xi \\ &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} f(y) \left(\int_{\mathbb{R}^n} e^{-\frac{1}{2}\epsilon^2|\xi|^2} e^{-i\xi\cdot(y-x)} d\xi \right) dy \\ &= \int_{\mathbb{R}^n} f(y) \frac{1}{\epsilon^n (2\pi)^{\frac{n}{2}}} e^{-\frac{|y-x|^2}{2\epsilon^2}} dy \\ &= f \star \psi_\epsilon(x) \end{aligned}$$

where $\psi_\epsilon(x) = \epsilon^{-n}\psi(\epsilon^{-1}x)$ for

$$\psi(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}|x|^2}.$$

Note that $\psi \in C^\infty(\mathbb{R}^n)$, $\psi(x) \geqslant 0$ and

$$\int_{\mathbb{R}^n} \psi(x)dx = 1$$

so by Theorem 1.13, *b*) we have that:

$$f \star \psi_\epsilon \to f,$$

in $L^1(\mathbb{R}^n)$, thus we must have that

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{ix\cdot\xi} d\xi$$

for almost every x. □

Note that by the Riemann Lebesgue Lemma the map

$$x \mapsto \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{ix\cdot\xi} d\xi$$

is continuous. Thus under the conditions of the theorem, if f is additionally assumed to be continuous, then we can upgrade the almost everywhere convergence to convergence everywhere. Alternatively, our result shows that if both $f \in L^1(\mathbb{R}^n)$ and $\hat{f} \in L^1(\mathbb{R}^n)$, then f must be almost everywhere equal to a continuous function.

We can summarise the inversion formula quite neatly by noting that (on a suitable f):

$$\mathcal{F}^2 f = (2\pi)^n \check{f}.$$

An immediate corollary of the above result is that $\mathcal{F} : \mathscr{S} \to \mathscr{S}$ is a bijection, and that $\mathcal{F}^{-1} : \mathscr{S} \to \mathscr{S}$ is continuous.

Exercise($*$). Consider the following ODE problem. Given $f : \mathbb{R} \to \mathbb{C}$, find ϕ such that:

$$-\phi'' + \phi = f. \tag{4.4}$$

a) Show that if $f \in \mathscr{S}$, there is a unique $\phi \in \mathscr{S}$ solving (4.4), and give an expression for $\hat{\phi}$.

b) Show that

$$\phi(x) = \int_{\mathbb{R}} f(y) G(x-y) dy$$

where

$$G(x) = \begin{cases} \frac{1}{2}e^x & x < 0, \\ \frac{1}{2}e^{-x} & x \geqslant 0. \end{cases}$$

Exercise(∗). Suppose $f \in L^1(\mathbb{R}^3)$ is a radial function, i.e. $f(Rx) = f(x)$, whenever $R \in SO(3)$ is a rotation.

a) Show that $\hat{f}$ is radial.

b) Suppose that $\xi = (0, 0, \zeta)$. By writing the Fourier integral in polar coordinates, show that

$$\hat{f}(\xi) = \int_{r=0}^{\infty} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} f(r) e^{-i\zeta r \cos\theta} r^2 \sin\theta d\theta dr d\phi.$$

c) Making the substitution $s = \cos\theta$, and using the fact that $\hat{f}$ is radial, deduce:

$$\hat{f}(\xi) = 4\pi \int_0^{\infty} f(r) \frac{\sin r\,|\xi|}{r\,|\xi|} r^2 dr$$

for any $\xi \in \mathbb{R}^n$.

4.2 The Fourier transform on $L^2(\mathbb{R}^n)$

Having defined the Fourier transform acting on functions in $L^1(\mathbb{R}^n)$, we are going to extend it to act on more general functions (and eventually distributions). Firstly, we shall see how the Fourier transform extends very nicely to act on functions in $L^2(\mathbb{R}^n)$. As we have already seen, this is a particularly nice function space because it is a *Hilbert space*. We recall the inner product:

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

which induces the norm via:

$$\|f\|_{L^2} = (f, f)^{\frac{1}{2}}$$

and moreover it is complete, which means that all Cauchy sequences converge in $L^2(\mathbb{R}^n)$.

We shall first establish that the Fourier transform maps $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ into $L^2(\mathbb{R}^n)$, and moreover show that the L^2 inner product is preserved by the Fourier transform (up to multiplication by a constant).

Theorem 4.6 (Parseval's Formula)**.** *Suppose $f, g \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$. Then $\hat{f}, \hat{g} \in L^2(\mathbb{R}^n)$ and moreover:*

$$(f, g) = \frac{1}{(2\pi)^n} (\hat{f}, \hat{g}).$$

Proof. We will again use a density argument to prove this result. First suppose that $f, g \in \mathscr{S}$. Then using the Fourier Inversion Theorem (Theorem 4.5) and Fubini's theorem we can calculate:

$$\begin{aligned}
(f,g) &= \int_{\mathbb{R}^n} \overline{f}(x) g(x) dx \\
&= \int_{\mathbb{R}^n} \overline{f}(x) \left(\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{g}(\xi) e^{ix\cdot\xi} d\xi \right) dx \\
&= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \overline{f}(x) e^{ix\cdot\xi} dx \right) \hat{g}(\xi) d\xi \\
&= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \overline{\left(\int_{\mathbb{R}^n} f(x) e^{-ix\cdot\xi} dx \right)} \hat{g}(\xi) d\xi \\
&= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \overline{\hat{f}(\xi)} \hat{g}(\xi) d\xi = \frac{1}{(2\pi)^n} (\hat{f}, \hat{g})
\end{aligned}$$

Now suppose that $f, g \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$. By Theorem 1.13 part *b*), there exists a sequence $\{f_j\}_{j=1}^{\infty} \subset C_0^{\infty}(\mathbb{R}^n) \subset \mathscr{S}$ such that:

$$\|f_j - f\|_{L^1} + \|f_j - f\|_{L^2} < \frac{1}{j}$$

and similarly for g. We know that:

$$\sup_{\xi \in \mathbb{R}^n} \left| \hat{f}_j(\xi) - \hat{f}(\xi) \right| \leqslant \|f_j - f\|_{L^1} < \frac{1}{j}$$

so that $\hat{f}_j \to f$ uniformly on $\mathbb{R}^n$. We also have by the calculation above:

$$\left\| \hat{f}_j - \hat{f}_k \right\|_{L^2} = (2\pi)^{\frac{n}{2}} \|f_j - f_k\|_{L^2} .$$

Now since $f_j \to f$ in $L^2(\mathbb{R}^n)$, we have that $\{f_j\}$ is a Cauchy sequence in $L^2(\mathbb{R}^n)$. Thus $\hat{f}_j$ is a Cauchy sequence in $L^2(\mathbb{R}^n)$. By the completeness of $L^2(\mathbb{R}^n)$, we have that $\hat{f}_j$ converges in $L^2(\mathbb{R}^n)$ and hence $\hat{f} \in L^2(\mathbb{R}^n)$. Furthermore, we know that

$$(f_j, g_j) = \frac{1}{(2\pi)^n} (\hat{f}_j, \hat{g}_j)$$

since each of the sequences $\{f_j\}, \{g_j\}, \{\hat{f}_j\}, \{\hat{g}_j\}$ converge in $L^2(\mathbb{R}^n)$, we can take the limit[2] $j \to \infty$ to conclude:

$$(f,g) = \frac{1}{(2\pi)^n} (\hat{f}, \hat{g}) \qquad \square$$

Thus we have shown that the Fourier transform $\mathcal{F}$ maps $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ into $L^2(\mathbb{R}^n)$. Moreover, we have that it is a *bounded* as an operator from $L^2(\mathbb{R}^n)$ to itself, since

$$\left\| \hat{f} \right\|_{L^2} \leqslant (2\pi)^{\frac{n}{2}} \|f\|_{L^2} .$$

This means that $\mathcal{F}$ is a bounded linear map defined on a dense subset of a Banach space. A general result tells us that the map extends uniquely to a bounded linear map on the entire space. Rather than invoke an abstract result, we can show this directly.

[2] You should check that you understand why this is valid.

Corollary 4.7. *There is a unique continuous linear operator* $\overline{\mathcal{F}} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ *such that:*

$$\overline{\mathcal{F}}[f] = \mathcal{F}[f], \qquad \textit{for all } \ f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n). \tag{4.5}$$

We say that $\overline{\mathcal{F}}$ *is the extension of the Fourier transform to* $L^2(\mathbb{R}^n)$*. It is sometimes known as the* Fourier-Plancherel *transform.*

Proof. For any $f \in L^2(\mathbb{R}^n)$, we can take a sequence $\{f_j\}_{j=1}^\infty \subset L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ with $f_j \to f$ in $L^2(\mathbb{R}^n)$ (for example by approximating f with smooth functions of compact support). By Theorem 4.6 we have that:

$$\left\| \hat{f}_j - \hat{f}_k \right\|_{L^2} = (2\pi)^{\frac{n}{2}} \left\| f_j - f_k \right\|_{L^2}. \tag{4.6}$$

Now, since f_j converges in $L^2(\mathbb{R}^n)$, it is in particular a Cauchy sequence in $L^2(\mathbb{R}^n)$. Equation (4.6) shows that $\hat{f}_j$ is also a Cauchy sequence in $L^2(\mathbb{R}^n)$, hence has a limit, say $F \in L^2(\mathbb{R}^n)$ by the completeness of $L^2(\mathbb{R}^n)$. Suppose $\{f'_j\}_{j=1}^\infty \subset L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ is another sequence with $f'_j \to f$, and suppose $\hat{f}_j \to F'$. Then we have:

$$\left\| F - F' \right\|_{L^2} = \lim_{j\to\infty} \left\| \hat{f}_j - \hat{f}'_j \right\|_{L^2} = \lim_{j\to\infty} (2\pi)^{\frac{n}{2}} \left\| f_j - f'_j \right\|_{L^2} = 0$$

since both f_j and f'_j tend to f. Thus F depends only f, and not on the sequence f_j which we chose to approximate f.

We define $\overline{\mathcal{F}}[f] = F$, i.e.:

$$\overline{\mathcal{F}}[f] = \lim_{j\to\infty} \mathcal{F}[f_j], \quad \text{where } \{f_j\}_{j=1}^\infty \subset L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n), \quad f_j \to f \text{ in } L^2(\mathbb{R}^n),$$

and the limit is to be understood to be in $L^2(\mathbb{R}^n)$. This certainly satisfies (4.5), since we can take our approximating sequence to be the constant sequence $f_j = f$ for all j when $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$. $\overline{\mathcal{F}}$ is clearly linear and moreover, we have that

$$\begin{aligned} \left\| \overline{\mathcal{F}}[f] \right\|_{L^2} &= \left\| \lim_{j\to\infty} \mathcal{F}[f_j] \right\|_{L^2} \\ &= \lim_{j\to\infty} \left\| \mathcal{F}[f_j] \right\|_{L^2} \\ &= \lim_{j\to\infty} (2\pi)^{\frac{n}{2}} \left\| f_j \right\|_{L^2} = (2\pi)^{\frac{n}{2}} \left\| f \right\|_{L^2}, \end{aligned}$$

so $\overline{\mathcal{F}}$ is bounded and hence continuous[3]. It remains to show that $\overline{\mathcal{F}}$ is unique. Suppose that $\overline{\mathcal{F}}'$ is another continuous linear operator satisfying (4.5). For any $f \in L^2(\mathbb{R}^n)$, take a sequence $\{f_j\}_{j=1}^\infty \subset L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ with $f_j \to f$ in $L^2(\mathbb{R}^n)$. We have:

$$\overline{\mathcal{F}}'[f] = \lim_{j\to\infty} \overline{\mathcal{F}}'[f_j] = \lim_{j\to\infty} \overline{\mathcal{F}}[f_j] = \overline{\mathcal{F}}[f]$$

so that $\overline{\mathcal{F}}' = \overline{\mathcal{F}}$. □

[3]If $\{f_j\}_{j=1}^\infty \subset L^2(\mathbb{R}^n)$ is a sequence with $f_j \to f$ in $L^2(\mathbb{R}^n)$, then

$$\left\| \overline{\mathcal{F}}[f_j] - \overline{\mathcal{F}}[f] \right\|_{L^2} = \left\| \overline{\mathcal{F}}[f_j - f] \right\|_{L^2} = (2\pi)^{\frac{n}{2}} \left\| f_j - f \right\|_{L^2} \to 0$$

so $\overline{\mathcal{F}}[f_j] \to \overline{\mathcal{F}}[f]$ in $L^2(\mathbb{R}^n)$.

Exercise($*$). (*) Suppose that $f, g \in L^2(\mathbb{R}^n)$, and denote the Fourier-Plancherel transform by $\overline{\mathcal{F}}$. You may assume any results already established for the Fourier transform.

a) Show that
$$(f, g) = \frac{1}{(2\pi)^n}\left(\overline{\mathcal{F}}[f], \overline{\mathcal{F}}[g]\right).$$

b) Recall that $\check{f}(y) = f(-y)$. Show that:
$$\overline{\mathcal{F}}\left[\overline{\mathcal{F}}[f]\right] = (2\pi)^n \check{f}.$$
Hence, or otherwise, deduce that $\overline{\mathcal{F}} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a bijection, and that $\overline{\mathcal{F}}^{-1} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a bounded linear map.

c) Show that:
$$\overline{\mathcal{F}}[f](\xi) = \lim_{R\to\infty} \int_{B_R(0)} f(x) e^{-ix\cdot\xi} dx$$
with convergence in the sense of $L^2(\mathbb{R}^n)$.

d) Suppose that $f \in C^1(\mathbb{R}^n)$ and $f, D_j f \in L^2(\mathbb{R}^n)$. Show that $\xi_j \overline{\mathcal{F}}[f](\xi) \in L^2(\mathbb{R}^n)$ and:
$$\overline{\mathcal{F}}[D_j f](\xi) = i\xi_j \overline{\mathcal{F}}[f](\xi)$$

e) For $x \in \mathbb{R}$ let:
$$f(x) = \frac{\sin x}{x}$$
   i) Show that $f \in L^2(\mathbb{R})$.
   ii) Show that:
$$\overline{\mathcal{F}}[f](\xi) = \begin{cases} \pi & -1 < \xi < 1, \\ 0 & |\xi| \geqslant 1. \end{cases}$$

f) i) Show that for all $x \in \mathbb{R}^n$:
$$|f \star g(x)| \leqslant \|f\|_{L^2} \|g\|_{L^2}.$$
   ii) Show that $f \star g \in C^0(\mathbb{R}^n)$ and:
$$f \star g = \mathcal{F}^{-1}\left[\overline{\mathcal{F}}[f] \cdot \overline{\mathcal{F}}[g]\right]$$
   where:
$$\mathcal{F}^{-1}[\hat{f}](x) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \hat{f}(\xi) e^{i\xi\cdot x} d\xi.$$

[Hint for parts a), b), d), f): approximate by Schwartz functions]

Exercise(∗). Work in $\mathbb{R}^3$. For $k > 0$, define the function:

$$G(x) = \frac{e^{-k|x|}}{4\pi |x|}$$

a) Show that $G \in L^1(\mathbb{R}^3)$.

b) Show that:

$$\hat{G}(\xi) = \frac{1}{|\xi|^2 + k^2}$$

[Hint: use Exercise 4.1, part c)]

Exercise 3.10. Suppose $f \in \mathscr{S}(\mathbb{R}^n)$. By observing that

$$\|f\|_{L^2}^2 = \int_{\mathbb{R}^n} \frac{1}{n}(\text{div } x) |f(x)|^2 dx,$$

or otherwise, show that:

$$(2\pi)^{\frac{n}{2}} \|f\|_{L^2}^2 \leqslant \frac{2}{n} \big\| |x| f(x) \big\|_{L^2} \big\| |\xi| \hat{f}(\xi) \big\|_{L^2}$$

with equality if and only if $f(x) = ae^{-\lambda |x|^2}$ for some $a \in \mathbb{C}, \lambda > 0$. Deduce that if $x_0, \xi_0 \in \mathbb{R}^n$:

$$(2\pi)^{\frac{n}{2}} \|f\|_{L^2}^2 \leqslant \frac{2}{n} \big\| |x - x_0| f(x) \big\|_{L^2} \big\| |\xi - \xi_0| \hat{f}(\xi) \big\|_{L^2}.$$

Explain how this shows that a function $f \in L^2(\mathbb{R}^n)$ cannot be sharply localised in both physical and Fourier space simultaneously. This is the *uncertainty principle.*

Usually one does not labour the distinction between the Fourier transform acting on $L^1(\mathbb{R}^n)$ and the Fourier-Plancherel transform acting on $L^2(\mathbb{R}^n)$. From now on we shall use the same notation for both, so that for $f \in L^2(\mathbb{R}^n)$ we write $\overline{\mathcal{F}}[f] = \mathcal{F}[f] = \hat{f}$. Since the two transforms agree wherever both are defined, there is no ambiguity in this. The majority of the results that we have already established for the Fourier transform extend to the Fourier-Plancherel transform in a straightforward way, see Exercise 4.2.

4.3 The Fourier transform on $\mathscr{S}'$

We are now going to extend the Fourier transform in a slightly different way, such that it acts on distributions. Suppose that $f \in L^1(\mathbb{R}^n)$, and $\phi \in \mathscr{S}$. Then since $\hat{f} \in C^0(\mathbb{R}^n)$ and $\hat{f}$ decays towards infinity, we have that $T_{\hat{f}} \in \mathscr{S}'$. By Fubini we have:

$$\begin{aligned} T_{\hat{f}}[\phi] = \int_{\mathbb{R}^n} \hat{f}(x)\phi(x)dx &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(y) e^{-ix\cdot y} dy \right) \phi(x) dx \\ &= \int_{\mathbb{R}^n} f(y) \left(\int_{\mathbb{R}^n} \phi(x) e^{-ix\cdot y} dx \right) dy \\ &= \int_{\mathbb{R}^n} f(x)\hat{\phi}(x)dx. \end{aligned}$$

Thus for $f \in L^1(\mathbb{R}^n)$ we have that:

$$T_{\hat{f}}[\phi] = T_f[\hat{\phi}], \qquad \text{for all } \phi \in \mathscr{S}.$$

Motivated by this, we define:

Definition 4.1. *For a distribution $u \in \mathscr{S}'$, we define the Fourier transform of u, written $\hat{u} \in \mathscr{S}'$ to be the distribution satisfying:*

$$\hat{u}[\phi] = u[\hat{\phi}], \qquad \textit{for all } \phi \in \mathscr{S}.$$

Notice that the definition makes sense because the Fourier transform maps $\mathscr{S}$ to $\mathscr{S}$ continuously. If we tried to use the above definition but with $\phi \in \mathscr{D}(\mathbb{R}^n)$ and $u \in \mathscr{D}'(\mathbb{R}^n)$, we would run into difficulties because $\hat{\phi} \notin \mathscr{D}(\mathbb{R}^n)$.

Example 17. *a) For $\xi \in \mathbb{R}^n$ we have:*

$$\widehat{\delta_\xi} = T_{e_{-\xi}}$$

To see this, we use the definition. For $\phi \in \mathscr{S}$:

$$\widehat{\delta_\xi}[\phi] = \delta_\xi[\hat{\phi}] = \hat{\phi}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi}\phi(x)dx = T_{e_{-\xi}}[\phi]$$

Since ϕ was arbitrary, the distributions are equal.

b) For $x \in \mathbb{R}^n$ we have:

$$\widehat{T_{e_x}} = (2\pi)^n \delta_x.$$

To see this, we note for $\phi \in \mathscr{S}$:

$$\widehat{T_{e_x}}[\phi] = T_{e_x}[\hat{\phi}] = \int_{\mathbb{R}^n} e^{ix\cdot\xi}\hat{\phi}(\xi)d\xi = (2\pi)^n\phi(x) = (2\pi)^n\delta_x[\phi].$$

Again, as ϕ is arbitrary the distributions are equal. Note that a particular case is $\widehat{T_1} = (2\pi)^n\delta_0$.

c) For α a multi-index, denote by X^α the map

$$X^\alpha : x \mapsto x^\alpha.$$

Then we have:

$$\widehat{T_{X^\alpha}} = (2\pi)^n i^{|\alpha|} D^\alpha \delta_0$$

For $\phi \in \mathscr{S}$:

$$\begin{aligned}
\widehat{T_{X^\alpha}}[\phi] = T_{X^\alpha}[\hat{\phi}] &= \int_{\mathbb{R}^n} \xi^\alpha \hat{\phi}(\xi)d\xi \\
&= (-i)^{|\alpha|}\int_{\mathbb{R}^n} \widehat{D^\alpha\phi}(\xi)d\xi \\
&= (2\pi)^n(-i)^{|\alpha|}D^\alpha\phi(0) = (2\pi)^n i^{|\alpha|} \times (-1)^{|\alpha|}\delta_0\left[D^\alpha\phi\right] \\
&= (2\pi)^n i^{|\alpha|} D^\alpha \delta_0[\phi]
\end{aligned}$$

Most of the properties of the Fourier transform defined on $\mathscr{S}$ are inherited by the transform defined on $\mathscr{S}'$. We first need to define a couple of operations on $\mathscr{S}'$. Recall that if $\phi \in \mathscr{S}$, then $\tau_x \phi \in \mathscr{S}$ is the translate of ϕ, given by $\tau_x\phi(y) = \phi(y-x)$, and $\check{\phi} \in \mathscr{S}$ is given by $\check{\phi}(y) = \phi(-y)$. For $u \in \mathscr{S}$, we define:

$$\tau_x u[\phi] = u[\tau_{-x}\phi], \qquad \check{u}[\phi] = u[\check{\phi}]$$

Notice also that if $f \in C^\infty(\mathbb{R}^n)$ is a function of tempered growth, i.e., if for each α and there exists a constant C_α and integer N_α such that:

$$|D^\alpha f(x)| \leqslant C_\alpha (1+|x|)^{N_\alpha}, \qquad \forall x \in \mathbb{R}^n.$$

then $\phi f \in \mathscr{S}$ when $\phi \in \mathscr{S}$ and we can define $fu \in \mathscr{S}'$ by

$$fu[\phi] = u[f\phi]$$

Exercise($*$). Verify that if $f \in L^1_{loc.}$ is such that $T_f \in \mathscr{S}'$, then:

$$\tau_x T_f = T_{\tau_x f}, \quad \text{and} \quad \check{T}_f = T_{\check{f}}$$

Lemma 4.8. *Suppose $u \in \mathscr{S}'$ is a tempered distribution. Then:*

$$\widehat{e_x u} = \tau_x \hat{u}, \qquad \widehat{\tau_x u} = e_{-x}\hat{u}, \qquad \widehat{D^\alpha u} = i^{|\alpha|} X^\alpha \hat{u} \qquad D^\alpha \hat{u} = (-i)^{|\alpha|}\widehat{X^\alpha u}$$

Moreover:

$$\hat{\hat{u}} = (2\pi)^n \check{u},$$

so that the Fourier transform on $\mathscr{S}'$ is invertible.

Proof. These are all calculations using the corresponding results for $\mathscr{S}$. Take $\phi \in \mathscr{S}$. We have:

$$\widehat{e_x u}[\phi] = e_x u[\hat{\phi}] = u[e_x\hat{\phi}] = u\left[\widehat{\tau_{-x}\phi}\right] = \hat{u}[\tau_{-x}\phi] = \tau_x u[\phi].$$

Since ϕ was arbitrary, we have $\widehat{e_x u} = \tau_x \hat{u}$. Similarly, we calculate:

$$\widehat{\tau_x u}[\phi] = \tau_x u[\hat{\phi}] = u[\tau_{-x}\hat{\phi}] = u\left[\widehat{e_{-x}\phi}\right] = \hat{u}[e_{-x}\phi] = e_{-x}u[\phi].$$

Next we have

$$\begin{aligned}\widehat{D^\alpha u}[\phi] &= D^\alpha u\left[\hat{\phi}\right] = (-1)^{|\alpha|} u\left[D^\alpha \hat{\phi}\right] \\ &= (-1)^{|\alpha|} u\left[(-i)^{|\alpha|}\widehat{X^\alpha \phi}\right] = i^{|\alpha|} u\left[\widehat{X^\alpha\phi}\right] \\ &= i^{|\alpha|}\hat{u}\left[X^\alpha\phi\right] = \left(i^{|\alpha|}X^\alpha\hat{u}\right)[\phi]\end{aligned}$$

similarly:

$$\begin{aligned}D^\alpha\hat{u}[\phi] &= (-1)^{|\alpha|}\hat{u}[D^\alpha\phi] = (-1)^{|\alpha|}u[\widehat{D^\alpha\phi}] \\ &= (-1)^{|\alpha|}u[i^{|\alpha|}X^\alpha\hat{\phi}] = (-i)^{|\alpha|}X^\alpha u[\hat{\phi}] \\ &= \left((-i)^{|\alpha|}\widehat{X^\alpha u}\right)[\phi].\end{aligned}$$

Finally, we have

$$\hat{\hat{u}}[\phi] = \hat{u}[\hat{\phi}] = u[\hat{\hat{\phi}}] = u[(2\pi)^n \check{\phi}] = (2\pi)^n \check{u}[\phi]$$

Since $\check{\check{u}} = u$, we have that the Fourier transform is invertible. □

Importantly, the Fourier transform is also a continuous linear map $\mathscr{S}' \to \mathscr{S}'$.

Lemma 4.9. *The map:*

$$\begin{array}{rcccc} \mathcal{F} & : & \mathscr{S}' & \to & \mathscr{S}' \\ & & u & \mapsto & \hat{u} \end{array}$$

is a linear homeomorphism.

Proof. We already have that $\mathcal{F}$ is linear. From the definition of the weak-$\star$ topology, a sequence $\{u_j\}_{j=1}^{\infty} \subset \mathscr{S}'$ converges to u if

$$u_j[\phi] \to u[\phi]$$

for all $\phi \in \mathscr{S}$. Suppose that we have such a convergent sequence in $\mathscr{S}'$. We calculate:

$$\hat{u}_j[\phi] = u_j[\hat{\phi}] \to u[\hat{\phi}] = \hat{u}[\phi].$$

Thus if $u_j \to u$ we have $\mathcal{F}(u_j) \to \mathcal{F}(u)$. Thus $\mathcal{F}$ is continuous. Since $\mathcal{F}^4 = (2\pi)^{2n}\iota$, we have that $\mathcal{F}$ is invertible and the inverse is also continuous. □

Remark. *Strictly, we have only established that $\mathcal{F}$ is sequentially continuous with respect to the weak-$\star$ topology induced on $\mathscr{S}'$ by $\mathscr{S}$. Establishing genuine continuity is not difficult, but requires the full description of the weak-$\star$ topology, and we leave this as an exercise.*

Exercise 3.11. Let $f : \mathbb{R} \to \mathbb{R}$ be the sign function

$$f(x) = \begin{cases} -1 & x < 0 \\ 1 & x \geqslant 0 \end{cases}$$

and define $f_R(x) = f(x)\mathbb{1}_{[-R,R]}(x)$.

a) Sketch $f_R(x)$, and show that $T_{f_R} \to T_f$ in $\mathscr{S}'(\mathbb{R})$ as $R \to \infty$.

b) Compute $\hat{f}_R(\xi)$, and show that for $\phi \in \mathscr{S}(\mathbb{R})$:

$$T_{\hat{f}_R}[\phi] = -2i \int_0^\infty \frac{\phi(x) - \phi(-x)}{x} dx + 2i \int_0^\infty \left(\frac{\phi(x) - \phi(-x)}{x} \right) \cos Rx dx.$$

Deduce $\widehat{T_f} = -2iP.V.\left(\frac{1}{x}\right)$, where we define the distribution $P.V.\left(\frac{1}{x}\right)$ by:

$$P.V.\left(\frac{1}{x}\right)[\phi] = \lim_{\epsilon \to 0} \int_{\mathbb{R}\setminus(-\epsilon,\epsilon)} \frac{\phi(x)}{x} dx, \qquad \phi \in \mathscr{S}(\mathbb{R}).$$

c) Write down $\widehat{T_H}$, where H is the Heaviside function:

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

By considering $e^{-\epsilon x}H(x)$, or otherwise, find an expression for the distribution u which acts on $\phi \in \mathscr{S}(\mathbb{R})$ by:

$$u[\phi] := \lim_{\epsilon \to 0^+} \int_{\mathbb{R}} \frac{\phi(x)}{x + i\epsilon} dx.$$

Exercise 3.12. Suppose $\phi \in C_c^\infty(\mathbb{R}^n \times \mathbb{R}^m)$. For each $y \in \mathbb{R}^m$ let $\phi_y : \mathbb{R}^n \to \mathbb{C}$ be given by $\phi_y(x) = \phi(x, y)$. Let $u \in \mathscr{D}'(\mathbb{R}^n)$.

a) Show that $\psi : y \mapsto u[\phi_y]$ is smooth and find an expression for $D^\alpha \psi$. Deduce that

$$\int_{\mathbb{R}^m} \psi(y) dy = u[\Psi], \quad \text{where} \quad \Psi(x) = \int_{\mathbb{R}^m} \phi(x, y) dy.$$

b) Show that there exists a sequence of smooth functions $f_n \in C_c^\infty(\mathbb{R}^n)$ such that $T_{f_n} \to u$ in the weak-$*$ topology of $\mathscr{D}'(\mathbb{R}^n)$.

4.3.1 Convolutions

We have generalised almost all of the properties of the Fourier transform to distributions. The final result that we shall establish concerns convolutions. Recall that if $u \in \mathscr{D}'(\mathbb{R}^n)$ and $\phi \in \mathscr{D}(\mathbb{R}^n)$ then $u \star \phi \in C^\infty(\mathbb{R}^n)$ is given by:

$$u \star \phi(x) = u\left[\tau_x \check{\phi}\right].$$

Notice that this definition continues to make sense for each x, provided $u \in \mathscr{S}'$ and $\phi \in \mathscr{S}$, although it is no longer clear that the resulting function is smooth. We have the following results concerning this convolution.

Theorem 4.10. *Suppose $u \in \mathscr{S}'$ and $\phi \in \mathscr{S}$ are given. Then the function:*

$$u \star \phi : \mathbb{R}^n \to \mathbb{C}$$

has the following properties

a) $u \star \phi \in C^\infty(\mathbb{R}^n)$ *with*

$$D^\alpha(u \star \phi) = D^\alpha u \star \phi = u \star D^\alpha \phi.$$

b) There exist constants $N \in \mathbb{N}$, $K > 0$ depending on u and ϕ such that:

$$|u \star \phi(x)| \leqslant K(1 + |x|)^N.$$

c) $T_{u\star\phi} \in \mathscr{S}'$ *and moreover:*

$$\widehat{T_{u\star\phi}} = \hat{\phi}\hat{u}.$$

d) For any $\psi \in \mathscr{S}$, we have:

$$(u \star \phi) \star \psi = u \star (\phi \star \psi)$$

e) We have:

$$T_{\hat{u} \star \hat{\phi}} = (2\pi)^n \widehat{\phi u}$$

Proof. a) The smoothness of $u \star \phi$ is proven exactly as in Lemma 3.10, *ii*). The only modification to the argument required is to note that for $\phi \in \mathscr{S}$, we have

$$\Delta_i^h \phi \to D_i \phi \quad \text{in} \quad \mathscr{S}, \quad \text{as} \quad h \to 0.$$

b) First, we note the following simple inequality which holds for all $x, y \in \mathbb{R}^n$:

$$1 + |x + y| \leqslant 1 + |x| + |y| \leqslant (1 + |x|)(1 + |y|).$$

Next, recall from Lemma 3.15 that there exist $N, k \in \mathbb{N}$ and $C > 0$ such that:

$$|u[\psi]| \leqslant C \sup_{y \in \mathbb{R}^n; |\alpha| \leqslant k} \left| (1 + |y|)^N D^\alpha \psi(y) \right|, \qquad \text{for all } \psi \in \mathscr{S}.$$

Applying this inequality with $\psi = \tau_x \check{\phi}$, we calculate:

$$\begin{aligned} |u \star \phi(x)| = \left| u \left[\tau_x \check{\phi} \right] \right| &\leqslant C \sup_{y \in \mathbb{R}^n; |\alpha| \leqslant k} \left| (1 + |y|)^N D^\alpha \phi(y - x) \right| \\ &= C \sup_{z \in \mathbb{R}^n; |\alpha| \leqslant k} \left| (1 + |z + x|)^N D^\alpha \phi(z) \right| \\ &\leqslant \left[C \sup_{z \in \mathbb{R}^n; |\alpha| \leqslant k} \left| (1 + |z|)^N D^\alpha \phi(z) \right| \right] (1 + |x|)^N \end{aligned}$$

which gives the result on setting:

$$K = C \sup_{y \in \mathbb{R}^n; |\alpha| \leqslant k} \left| (1 + |z|)^N D^\alpha \phi(z) \right|.$$

c) Combining the above two results, we have that $T_{u \star \phi} \in \mathscr{S}'$, since $u \star \phi \in L^1_{loc.}(\mathbb{R}^n)$ and $u \star \phi$ grows at most polynomially. It therefore makes sense to consider the Fourier transform. Suppose that $\psi \in \mathscr{D}(\mathbb{R}^n)$. We calculate:

$$\begin{aligned}
\widehat{T_{u\star\phi}}[\hat{\psi}] &= T_{u\star\phi}\left[\hat{\hat{\psi}}\right] = (2\pi)^n T_{u\star\phi}\left[\check{\psi}\right] && \text{Fourier Inversion Thm}\\
&= (2\pi)^n \int_{\mathbb{R}^n} u\star\phi(x)\psi(-x)dx && \text{Defn. of } T_f\\
&= (2\pi)^n \int_{\mathbb{R}^n} u\left[\tau_x\check{\phi}\right]\psi(-x)dx && \text{Defn. of } u\star\phi\\
&= (2\pi)^n \int_{\mathbb{R}^n} u\left[\psi(-x)\tau_x\check{\phi}\right]dx && \text{Linearity of } u\\
&= (2\pi)^n u\left[\int_{\mathbb{R}^n}\psi(-x)\tau_x\check{\phi}dx\right] && \text{(!!)}\\
&= (2\pi)^n u\left[(\phi\check{\star}\psi)\right] && \text{Defn. of } \phi\star\psi\\
&= u\left[\widehat{\widehat{\phi\star\psi}}\right] = \hat{u}\left[\widehat{\phi\star\psi}\right] && \text{Fourier Inversion Thm}\\
&= \hat{u}\left[\hat{\phi}\hat{\psi}\right] = \left(\hat{\phi}\hat{u}\right)[\hat{\psi}] && \text{F.T. of convolution}
\end{aligned}$$

Most of the manipulations here are relatively straightforward. We have used Theorems 4.2, 4.5 in addition to various definitions. The step marked (!!), in which we interchanged an integration and an application of u requires some justification. Crudely this is true because we can replace the integration with an appropriately convergent Riemann sum and use the linearity and continuity of u. We shall justify this step in Lemma 4.11. The conclusion of this calculation is that:

$$\widehat{T_{u\star\phi}}[\hat{\psi}] = \left(\hat{\phi}\hat{u}\right)[\hat{\psi}]$$

This holds for all $\psi \in \mathscr{D}(\mathbb{R}^n)$. Now, since $\mathscr{D}(\mathbb{R}^n)$ is dense in $\mathscr{S}$ and $\mathcal{F} : \mathscr{S} \to \mathscr{S}$ is a homeomorphism, we have that:

$$\mathcal{F}\left[\mathscr{D}(\mathbb{R}^n)\right] = \left\{\hat{\psi} : \psi \in \mathscr{D}(\mathbb{R}^n)\right\}$$

is dense in $\mathscr{S}$. Thus, by approximation,

$$\widehat{T_{u\star\phi}}[\chi] = \left(\hat{\phi}\hat{u}\right)[\chi]$$

holds for any $\chi \in \mathscr{S}$ and we're done.

d) Note that in the process of proving the previous part, we established that for any $\psi \in \mathscr{D}(\mathbb{R}^n)$:

$$\int_{\mathbb{R}^n} u\star\phi(x)\psi(-x)dx = u\left[(\phi\check{\star}\psi)\right]$$

which is equivalent to:

$$(u\star\phi)\star\psi(0) = u\star(\phi\star\psi)(0). \tag{4.7}$$

Now, note that:

$$u \star \tau_y \phi = \tau_y (u \star \phi), \qquad \phi \star \tau_y \psi = \tau_y (\phi \star \psi)$$

as can be easily seen from the definitions. Applying (4.7) with ψ replaced by $\tau_y \psi$, we conclude that:

$$(u \star \phi) \star \psi(y) = u \star (\phi \star \psi)(y).$$

Since this holds for any $\psi \in \mathscr{D}(\mathbb{R}^n)$ and $\mathscr{D}(\mathbb{R}^n)$ is dense in $\mathscr{S}$, we're done.

e) This result follows by applying part *c)* to $\hat{u} \star \hat{\phi}$ and repeatedly making use of the Fourier inversion theorem. We calculate:

$$\begin{aligned} \widehat{T_{\hat{u}\star\hat{\phi}}} &= \hat{\hat{\phi}}\hat{\hat{u}} = (2\pi)^{2n} \check{\phi}\check{u} \\ &= (2\pi)^{2n} (\check{\phi u}) = (2\pi)^n \widehat{\widehat{(\phi u)}} \end{aligned}$$

Since the Fourier transform is a bijection on $\mathscr{S}'$, the result follows.

□

In order to complete the proof of the above result, we need to justify the step marked (!!) in which a tempered distribution and an integration were interchanged. We will first prove a result concerning the convergence of Riemann sums, which will enable us to establish that the (!!) step was justified. Let us suppose that $\Omega \subset \mathbb{R}^n$ and $\Omega' \subset \mathbb{R}^m$ are open and that $f \in C^0(\Omega \times \Omega')$ is uniformly continuous. We will also assume that there is some $R > 0$ such that:

$$\operatorname{supp} f(\cdot, y) \subset [-R, R)^n \subset \Omega$$

for each $y \in \Omega'$.

Next, we define a dyadic family of partitions of $[-R, R)^n$ into cubes as follows:

$$\Pi_k = \left\{ \left[-\frac{R}{2^k} i_1, \frac{R}{2^k}(i_1 + 1) \right) \times \cdots \times \left[-\frac{R}{2^k} i_n, \frac{R}{2^k}(i_n + 1) \right) : i_l \in [-2^k, 2^k - 1] \cap \mathbb{Z} \right\}$$

where $k = 0, 1, \ldots$. The $(k+1)^{st}$ partition is obtained by chopping each cube in the k^{th} partition into cubes with half the side length. Clearly for each fixed k:

$$\bigcup \Pi_k = [-R, R)^n$$

For $\pi \in \Pi_k$, we define x_π to be the point at the centre of the cube π. We define the k^{th} Riemann sum with respect to this partition by:

$$S_k(y) = \sum_{\pi \in \Pi_k} f(x_\pi, y) \, |\pi| \, .$$

Lemma 4.11. *With the definitions as above,*

$$S_k(y) \to \int_\Omega f(x, y) dx$$

uniformly in $y \in \Omega'$.

Proof. First note that $x \mapsto f(x,y)$ is continuous and of compact support, hence Riemann integrable on Ω. Thus for each fixed y we have:

$$S_k(y) \to \int_\Omega f(x,y)dx$$

Next consider $k' \geqslant k$. We have that $\Pi_{k'}$ is a refinement of Π_k, i.e. if $\pi' \in \Pi_{k'}$, then there is a unique $\pi \in \Pi_k$ with $\pi' \subset \pi$. We calculate:

$$\begin{aligned} S_k(y) - S_{k'}(y) &= \sum_{\pi\in\Pi_k} f(x_\pi, y)\,|\pi| - \sum_{\pi\in\Pi_k} \sum_{\substack{\pi'\in\Pi_{k'} \\ \pi'\subset\pi}} f(x_{\pi'}, y)\,|\pi'| \\ &= \sum_{\pi\in\Pi_k} \sum_{\substack{\pi'\in\Pi_{k'} \\ \pi'\subset\pi}} \left(f(x_\pi, y) - f(x_{\pi'}, y)\right)|\pi'| \end{aligned}$$

here we have used that:

$$|\pi| = \sum_{\substack{\pi'\in\Pi_{k'} \\ \pi'\subset\pi}} |\pi'|$$

Now, since f is uniformly continuous, we know that for any $\epsilon > 0$ there exists a δ, independent of y, such that

$$\left|f(x,y) - f(x',y)\right| < \epsilon$$

for all $|x - x'| < \delta$. Notice that for $\pi' \subset \pi$ we have:

$$\left|x'_\pi - x_\pi\right| \leqslant \frac{R}{2^{k+1}}\sqrt{n}.$$

Thus given $\epsilon > 0$, there exists K such that for all $k \geqslant K$:

$$|f(x_\pi, y) - f(x_{\pi'}, y)| < \frac{\epsilon}{(2R)^n}.$$

Now suppose $k' \geqslant k \geqslant K$. We estimate:

$$\begin{aligned} |S_k(y) - S_{k'}(y)| &\leqslant \sum_{\pi\in\Pi_k} \sum_{\substack{\pi'\in\Pi_{k'} \\ \pi'\subset\pi}} |f(x_\pi, y) - f(x_{\pi'}, y)|\,|\pi'| \\ &\leqslant \frac{\epsilon}{(2R)^n} \sum_{\pi\in\Pi_k} \sum_{\substack{\pi'\in\Pi_{k'} \\ \pi'\subset\pi}} |\pi'| = \epsilon, \end{aligned}$$

since the sum over the partition simply gives us back the volume of the large cube. Sending k' to infinity, we have the result we require. □

This result allows us to establish the result we require:

Corollary 4.12. *Suppose $u \in \mathscr{S}'$, $\phi \in \mathscr{S}$ and $\psi \in \mathscr{D}(\mathbb{R}^n)$. Then:*

$$u\left[\int_{\mathbb{R}^n} \psi(-x)\tau_x\check{\phi}dx\right] = \int_{\mathbb{R}^n} u\left[\psi(-x)\tau_x\check{\phi}\right]dx$$

Proof. Fix Ω, $R > 0$ such that $\operatorname{supp} \check{\psi} \subset [-R, R)^n \subset \Omega$. Define the map:

$$\begin{array}{rcccl} f & : & \Omega \times \mathbb{R}^n & \to & \mathbb{C} \\ & & (x, y) & \mapsto & \psi(-x)\phi(y-x) \end{array}$$

Notice that $(1+|y|)^N D^\alpha_y f$ is uniformly continuous on $\Omega \times \mathbb{R}^n$ for any α, N. Thus applying Lemma 4.11 we deduce that:

$$S_k \to \int_{\mathbb{R}^n} \psi(-x)\tau_x\check{\phi}dx, \qquad \text{in } \mathscr{S}.$$

By the continuity of u, we deduce that:

$$u\left[\int_{\mathbb{R}^n} \psi(-x)\tau_x\check{\phi}dx\right] = u\left[\lim_{k\to\infty} S_k\right] = \lim_{k\to\infty} u\left[S_k\right]$$

By the linearity of u, we calculate:

$$u\left[S_k\right] = u\left[\sum_{\pi\in\Pi_k} f(x_\pi, \cdot)\,|\pi|\right] = \sum_{\pi\in\Pi_k} u\left[f(x_\pi, \cdot)\right]|\pi| = \sum_{\pi\in\Pi_k} u\left[\psi(-x_\pi)\tau_{x_\pi}\check{\phi}\right]|\pi|$$

But $x \mapsto u\left[\psi(-x)\tau_x\check{\phi}\right]$ is smooth, hence Riemann integrable, and we have that

$$\lim_{k\to\infty} \sum_{\pi\in\Pi_k} u\left[\psi(-x_\pi)\tau_{x_\pi}\check{\phi}\right]|\pi| = \int_{\mathbb{R}^n} u\left[\psi(-x)\tau_x\check{\phi}\right]dx.$$

□

4.4 The Fourier–Laplace transform on $\mathscr{E}'(\mathbb{R}^n)$

Recall that $\mathscr{E}'(\mathbb{R}^n) \subset \mathscr{S}'(\mathbb{R}^n)$ is a continuously embedded subspace, consisting of the distributions of *compact support*. These distributions are precisely those which extend to continuous linear maps from $\mathscr{E}(\mathbb{R}^n)$ to $\mathbb{C}$ (see Theorem 3.14). For these distributions, we can express the Fourier transform in a very clean fashion.

Theorem 4.13. *Suppose that $u \in \mathscr{E}'(\mathbb{R}^n)$. Then $\hat{u} = T_{\hat{v}}$ for some $\hat{v} \in C^\infty(\mathbb{R}^n)$ with:*

$$\hat{v}(\xi) = u[e_{-\xi}].$$

Proof. Suppose that $\operatorname{supp} u \subset B_R(0)$. Pick $\psi \in \mathscr{D}(\mathbb{R}^n)$ with $\psi = 1$ on $B_{R+1}(0)$, so that $\psi u = u$. We calculate:

$$\hat{u} = \widehat{\psi u} = \frac{1}{(2\pi)^n} T_{\hat{u}\star\hat{\psi}}.$$

By Theorem 4.10 e). Thus we have $\hat{u} = T_{\hat{v}}$ with $\hat{v} = (2\pi)^{-n}\hat{u}\star\hat{\psi} \in C^\infty(\mathbb{R}^n)$, by Theorem 4.10 a).

Now let $\phi \in \mathscr{S}$ be such that $\hat{\phi} = \psi$. We calculate:

$$\begin{aligned} \hat{v}(\xi) &= \frac{1}{(2\pi)^n}\hat{u}\star\hat{\psi}(\xi) = \hat{u}\star\check{\phi}(\xi) \\ &= \hat{u}\left[\tau_\xi\phi\right] = u\left[\widehat{\tau_\xi\phi}\right] = u[e_{-\xi}\psi] = (\psi u)[e_{-\xi}] \\ &= u[e_{-\xi}]. \end{aligned}$$

□

In practice, one does not distinguish between the distribution $\hat{u}$ and the function $\hat{v}$ and one uses the same letter to denote both. Notice that for $u \in \mathscr{E}'(\mathbb{R}^n)$, the expression $u[e_{-z}]$ makes sense for $z \in \mathbb{C}^n$. Moreover, this function is in fact *holomorphic* on $\mathbb{C}^n$. The analytic extension of a Fourier transform from $\mathbb{R}^n$ to $\mathbb{C}^n$ (or a subset thereof) is sometimes called the *Fourier-Laplace* transform.

4.5 Periodic distributions and Poisson's summation formula

Recall that the translate of a distribution $u \in \mathscr{D}'(\mathbb{R}^n)$ is defined by:

$$\tau_z u[\phi] = u\left[\tau_{-z}\phi\right], \qquad \text{for all } \phi \in \mathscr{D}(\mathbb{R}^n),$$

Bearing this in mind, we can make a very natural definition of what it means to be periodic:

Definition 4.2. *We say that a distribution $u \in \mathscr{D}'(\mathbb{R}^n)$ is* periodic *if for each $g \in \mathbb{Z}^n$ we have:*

$$\tau_g u = u.$$

Example 18. *a) The distribution $u = T_{e_{2\pi g}}$ is periodic for any $g \in \mathbb{Z}^n$. Suppose $g' \in \mathbb{Z}^n$. Then:*

$$\begin{aligned}\tau_{g'} T_{e_{2\pi g}}[\phi] &= T_{e_{2\pi g}}\left[\tau_{-g'}\phi\right] = \int_{\mathbb{R}^n} e^{2\pi i g\cdot y}\phi(y+g')dy \\ &= \int_{\mathbb{R}^n} e^{2\pi i g\cdot(z-g')}\phi(z)dz = e^{-2\pi i g\cdot g'}\int_{\mathbb{R}^n} e^{2\pi i g\cdot z}\phi(z)dz \\ &= T_{e_{2\pi g}}[\phi]\end{aligned}$$

b) Suppose $v \in \mathscr{E}'(\mathbb{R}^n)$. Then

$$u = \sum_{g\in\mathbb{Z}^n} \tau_g v$$

is periodic. It is straightforward to show that u defines a distribution (see Exercise below). To check periodicity, we have, for $g \in \mathbb{Z}^n$:

$$\tau_{g'} u[\phi] = u[\tau_{-g'}\phi] = \sum_{g\in\mathbb{Z}^n} \tau_g v\left[\tau_{-g'}\phi\right] = \sum_{g\in\mathbb{Z}^n} v\left[\tau_{-g-g'}\phi\right] = \sum_{g\in\mathbb{Z}^n} \tau_{g+g'} v\left[\phi\right] = u[\phi],$$

where we shift the dummy variable in the sum for the last step.

Exercise(∗). Suppose $v \in \mathscr{E}'(\mathbb{R}^n)$ and let:

$$u = \sum_{g\in\mathbb{Z}^n} \tau_g v.$$

Show that if $\phi \in \mathscr{D}(\mathbb{R}^n)$ with $\operatorname{supp}\phi \subset K$ for some compact $K \subset \mathbb{R}^n$ then

$$u[\phi] = \sum_{g\in A} \tau_g v[\phi],$$

for some finite set $A \subset \mathbb{Z}^n$ which depends only on K. Deduce that u defines a distribution.

When dealing with a periodic function $f \in C^\infty(\mathbb{R}^n)$, many quantities of interest are averages over a fundamental cell of the periodic lattice:

$$q = \left\{ x \in \mathbb{R}^n : -\frac{1}{2} \leqslant x_i < \frac{1}{2}, i = 1, \ldots, n \right\}$$

For example:

$$M(f) = \int_q f(x)dx$$

is the mean value that f attains. We'd like to extend this notion to makes sense for periodic functions, but we're presented with a difficulty. The obvious definition would be to set:

$$M(u) \stackrel{!}{=} u[\mathbb{1}_q]$$

but of course $\mathbb{1}_q \notin \mathscr{D}(\mathbb{R}^n)$ so we're not able to do this. Instead we will 'smear out' the function $\mathbb{1}_q$. To do this, notice that a crucial property of $\mathbb{1}_q$ is the following identity:

$$\sum_{g \in \mathbb{Z}^n} \tau_g \mathbb{1}_q = 1,$$

which tells us that $\mathbb{1}_q$ generates a partition of unity.

We shall construct a smooth 'partition of unity', which will allow us to localise various objects, and thus render them easier to deal with, and will enable us to define the mean of a periodic distribution. This is a slightly technical result, but the basic idea is important and crops up in many areas of analysis.

Lemma 4.14. *Let*

$$Q = \{x \in \mathbb{R}^n : |x_i| < 1, i = 1, \ldots, n\}$$

be the cube of side length 2 *centred at the origin. There exists a function* $\psi \in C^\infty(\mathbb{R}^n)$ *with* $\psi \geqslant 0$ *and* $\operatorname{supp} \psi \subset Q$ *such that:*

$$\sum_{g \in \mathbb{Z}^n} \tau_g \psi = 1.$$

Suppose that $u \in \mathscr{D}'(\mathbb{R}^n)$ *is periodic, and* ψ, ψ' *are both as above. Then:*

$$u[\psi] = u[\psi']$$

We then define:

$$M(u) := u[\psi]$$

Proof. Note

$$\overline{q} = \left\{ x \in \mathbb{R}^n : |x_i| \leqslant \frac{1}{2}, i = 1, \ldots, n \right\}.$$

By Lemma 1.14, there exists a function $\psi_0 \in C_0^\infty(Q)$, with $\psi_0(x) = 1$ for $x \in \overline{q}$ and $\psi_0 \geqslant 0$. Consider:

$$S(x) := \sum_{g \in \mathbb{Z}^n} \psi_0(x - g).$$

For any bounded open set Ω, we have that

$$A = \{g \in \mathbb{Z}^n : (\Omega - g) \cap Q \neq \emptyset\}$$

is finite. For $x \in \Omega$, we have:

$$S(x) = \sum_{g \in A} \psi_0(x - g),$$

so $S(x)$ is smooth. Moreover, for each $x \in \mathbb{R}^n$, there is at least one $g \in \mathbb{Z}^n$ with $x - g \in \overline{q}$. Thus $S(x) \geqslant 1$. We can thus take:

$$\psi(x) = \frac{\psi_0(x)}{S(x)}.$$

This is smooth, positive, supported in Q and moreover:

$$\sum_{g \in \mathbb{Z}^n} \tau_g \psi(x) = \frac{1}{S(x)} \sum_{g \in \mathbb{Z}^n} \psi_0(x - g) = 1.$$

Now suppose $u \in \mathscr{D}'(\mathbb{R}^n)$ is periodic and ψ, ψ' are both partitions of unity as above. We calculate:

$$\begin{aligned} u[\psi] &= u\left[\psi \sum_{g \in \mathbb{Z}^n} \tau_g \psi'\right] = \sum_{g \in \mathbb{Z}^n} u\left[\psi \tau_g \psi'\right] \\ &= \sum_{g \in \mathbb{Z}^n} \tau_{-g} u\left[\tau_{-g} \psi \psi'\right] = u\left[\psi' \sum_{g \in \mathbb{Z}^n} \tau_{-g} \psi\right] = u[\psi'] \end{aligned} \qquad \square$$

We thus have an acceptable definition of the mean of a periodic distribution. Notice that if $u = T_f$ for some locally integrable periodic function f, then by choosing a bounded sequence of ψ_j's such that $\psi_j \to \mathbb{1}_{\overline{q}}$ pointwise, we can show that:

$$M(T_f) = \int_{\overline{q}} f(x) dx,$$

justifying calling M the mean of the distribution.

To see why this technical lemma is useful, let us apply it to show that a periodic distribution is necessarily tempered, and in fact the periodic distributions we found by translating a compact distribution are indeed all of the periodic distributions.

Lemma 4.15. *Suppose $v \in \mathscr{E}'(\mathbb{R}^n)$ is a compact distribution. Then:*

$$u = \sum_{g \in \mathbb{Z}^n} \tau_g v \tag{4.8}$$

converges in $\mathscr{S}'$. Conversely, suppose that $u \in \mathscr{D}'(\mathbb{R}^n)$ is a periodic distribution. Then there exists $v \in \mathscr{E}'(\mathbb{R}^n)$ such that (4.8) holds and thus u extends uniquely to a tempered distribution $u \in \mathscr{S}'$.

Proof. Let $K = \operatorname{supp} v$. Since $v \in \mathscr{E}'(\mathbb{R}^n)$, by Lemma 3.13 there exists $C > 0, N$ such that:

$$|v[\phi]| \leqslant C \sup_{x\in K;|\alpha|\leqslant N} |D^\alpha \phi(x)|, \qquad \text{for all } \phi \in \mathscr{E}(\mathbb{R}^n).$$

Now suppose $\phi \in \mathscr{S} \subset \mathscr{E}(\mathbb{R}^n)$. We have:

$$|\tau_g v\,[\phi]| = |v\,[\tau_{-g}\phi]| \leqslant C \sup_{x\in K;|\alpha|\leqslant N} |D^\alpha \phi(x+g)|.$$

Since K is bounded, we have that $K \subset B_R(0)$ for some $R > 0$. We calculate:

$$1 + |g| = 1 + |x + g - x| \leqslant 1 + R + |x + g| \leqslant (1 + R)(1 + |x + g|)$$

for all $x \in K$, so that:

$$1 \leqslant (1+R)\frac{1+|x+g|}{1+|g|}.$$

We conclude that for any $M \geqslant 1$:

$$\begin{aligned} |\tau_g v\,[\phi]| &\leqslant \frac{C(1+R)^M}{(1+|g|)^M} \sup_{x\in K;|\alpha|\leqslant N} (1+|x+g|)^M |D^\alpha\phi(x+g)| \\ &\leqslant \frac{C(1+R)^M}{(1+|g|)^M} \sup_{y\in \mathbb{R}^n;|\alpha|\leqslant N} (1+|y|)^M |D^\alpha\phi(y)|. \end{aligned}$$

Since $\phi \in \mathscr{S}$, in particular we have:

$$|\tau_g v\,[\phi]| \leqslant \frac{C'}{(1+|g|)^{n+1}}$$

where C' depends on v, ϕ. Now, since:

$$\sum_{g\in\mathbb{Z}^n} \frac{1}{(1+|g|)^{n+1}} < \infty,$$

(see Exercise below) we deduce that for each $\phi \in \mathscr{S}$ the sum:

$$\sum_{g\in\mathbb{Z}^n} \tau_g v\,[\phi]$$

converges. This is precisely the statement that the sum in (4.8) converges in $\mathscr{S}'$.

Now suppose $u \in \mathscr{D}'(\mathbb{R}^n)$ is periodic, and take ψ as in Lemma 4.14. Suppose $\phi \in \mathscr{D}(\mathbb{R}^n)$ is arbitrary. We have:

$$u[\phi] = \left(\sum_{g\in\mathbb{Z}^n} \tau_g\psi\right) u[\phi] = \sum_{g\in\mathbb{Z}^n} u\,[\tau_g\psi\phi]. \tag{4.9}$$

Now, since u is periodic,:

$$u\,[\tau_g\psi\phi] = \tau_g u\,[\tau_g\psi\phi] = u\,[\psi\tau_{-g}\phi] = (\psi u)\,[\tau_{-g}\phi] = \tau_g(\psi u)[\phi]$$

Now ψu has compact support, so by Theorem 3.14 extends uniquely to $v \in \mathscr{E}'(\mathbb{R}^n)$. Thus we have:

$$u = \sum_{g \in \mathbb{Z}^n} \tau_g v$$

which by the first part of the proof converges in $\mathscr{S}'$, thus $u \in \mathscr{S}'$. □

Exercise(∗). Recall that for $x \in \mathbb{R}^n$:

$$\|x\|_1 := \sum_{i=1}^n |x_i| \, .$$

For $k \in \mathbb{N}$ set:

$$Q_k = \left\{ g \in \mathbb{Z}^n : k - \frac{1}{2} \leqslant \|g\|_1 < k + \frac{1}{2} \right\}$$

a) Show that:

$$\#Q_k = (2k+1)^n - (2k-1)^n$$

so that $\#Q_k \leqslant c(1+k)^{n-1}$ for some $c > 0$.

b) By applying the Cauchy-Schwarz identity to estimate $a \cdot b$ for $a = (1, \ldots, 1)$ and $b = (|g_1|, \ldots, |g_n|)$, deduce that:

$$\|g\|_1 \leqslant \sqrt{n} \, |g|$$

c) Show that there exists a constant $C > 0$, depending only on n such that:

$$\sum_{g \in \mathbb{Z}^n; \|g\|_1 \leqslant K} \frac{1}{(1+|g|)^{n+1}} \leqslant 1 + C \sum_{k=1}^K \frac{1}{k^2}$$

holds for all $K \in \mathbb{N}$. Deduce that:

$$\sum_{g \in \mathbb{Z}^n} \frac{1}{(1+|g|)^{n+1}} < \infty.$$

We have shown that a periodic distribution is necessarily tempered. It is therefore reasonable to ask what one can say about the Fourier transform of a periodic distribution. In fact, it will turn out that the Fourier transform of a periodic distribution has a very simple form: it consists of a sum over δ-distributions with support on the points of an integer lattice.

Before we establish this, we will first need a technical result regarding distributions.

Lemma 4.16. *Suppose that $u \in \mathscr{S}$ satisfies:*

$$\left(e_{-g'} - 1 \right) u = 0 \tag{4.10}$$

for all $g' \in \mathbb{Z}^n$. Then:

$$u = \sum_{g \in \mathbb{Z}^n} c_g \delta_{2\pi g},$$

where $c_g \in \mathbb{C}$ satisfy the bound

$$|c_g| \leqslant K(1 + |g|)^N$$

for some $K > 0$, $N \in \mathbb{Z}$, and the sum converges in $\mathscr{S}'$.

Proof. First, we claim that $\operatorname{supp} u \subset \Lambda$, with

$$\Lambda = \{2\pi g : g \in \mathbb{Z}^n\}.$$

Suppose $\phi \in \mathscr{D}(R^n)$ with $\operatorname{supp} \phi \subset \mathbb{R}^n \backslash \Lambda$. Then for each $g' \in \mathbb{Z}^n$, we have $\left(e_{-g'} - 1\right)^{-1} \phi \in \mathscr{S}$, since ϕ vanishes near any zeros of $e_{-g'} - 1$. Applying the condition (4.10), we deduce:

$$0 = \left(e_{-g'} - 1\right) u \left[\left(e_{-g'} - 1\right)^{-1} \phi\right] = u[\phi]$$

so u vanishes. Thus $\operatorname{supp} u \subset \Lambda$.

Now, let us take ψ as in Lemma 4.14, and define $\tilde{\psi}(x) = \psi\left(\frac{x}{2\pi}\right)$. It's straightforward to check that:

$$\sum_{g \in \mathbb{Z}^n} \tau_{2\pi g} \tilde{\psi} = 1, \qquad \operatorname{supp} \tilde{\psi} \subset \{x \in \mathbb{R}^n : |x_i| < 2\pi\}.$$

For $g \in \mathbb{Z}^n$, let us consider $v_g = (\tau_{2\pi g} \tilde{\psi}) u$. This distribution is supported at $2\pi g$, and by multiplying (4.10) by $\tau_{2\pi g} \tilde{\psi}$ we have:

$$\left(e_{-g'} - 1\right) v_g = 0$$

In particular, we have, taking $g' = l_j$ for $j = 1, \dots n$, where $\{l_j\}$ is the canonical basis for $\mathbb{R}^n$:

$$\left(e^{-i(x_j - 2\pi g_j)} - 1\right) v_g = 0.$$

Now,

$$\left(e^{-i(x_j - 2\pi g_j)} - 1\right) = (x_j - 2\pi g_j) \kappa(x_j)$$

where $\kappa(x_j)$ is non-zero on a neighbourhood of g_j. Thus we conclude that:

$$(x_j - 2\pi g_j) v_g = 0, \qquad j = 1, \dots n.$$

Now suppose $\phi \in \mathscr{S}$. We can write:

$$\phi(x) = \phi(2\pi g) + \sum_{j=1}^{n} (x_j - 2\pi g_j) \phi_j(x)$$

where $\phi_j(x) \in C^\infty(\mathbb{R}^n)$. Since v_g has compact support, it extends to smoothly to act on $\mathscr{E}(\mathbb{R}^n)$ and we calculate:

$$v_g[\phi] = v_g[\phi(2\pi g)] + \sum_{j=1}^{n} (x_j - 2\pi g_j) v_g[\phi_j] = v_g[\phi(2\pi g)]$$

Returning to the definition of v_g, we have:

$$(\tau_{2\pi g}\tilde{\psi})u[\phi] = (\tau_{2\pi g}\tilde{\psi})u[\phi(2\pi g)] = u[\tau_{2\pi g}\tilde{\psi}]\delta_{2\pi g}[\phi]$$

so that

$$(\tau_{2\pi g}\tilde{\psi})u = u[\tau_{2\pi g}\tilde{\psi}]\delta_{2\pi g}$$

Summing over $g \in \mathbb{Z}^n$, we recover:

$$\sum_{g\in\mathbb{Z}^n}(\tau_{2\pi g}\tilde{\psi})u = \left(\sum_{g\in\mathbb{Z}^n}(\tau_{2\pi g}\tilde{\psi})\right)u = u = \sum_{g\in\mathbb{Z}^n} c_g\delta_{2\pi g}$$

Where

$$c_g = u[\tau_{2\pi g}\tilde{\psi}].$$

To establish the estimate for c_g, we recall from Lemma 3.15 that there exist $N, k \in \mathbb{N}$ and $C > 0$ such that:

$$|u[\phi]| \leqslant C \sup_{x\in\mathbb{R}^n;|\alpha|\leqslant k}\left|(1+|x|)^N D^\alpha\phi(x)\right|, \qquad \text{for all } \phi \in \mathscr{S}.$$

Applying this to $\tau_{2\pi g}\tilde{\psi}$, we have:

$$\begin{aligned} |c_g| &\leqslant C \sup_{x\in\mathbb{R}^n;|\alpha|\leqslant k}\left|(1+|x|)^N D^\alpha\tilde{\psi}(x-2\pi g)\right| \\ &\leqslant C \sup_{x\in\mathbb{R}^n;|\alpha|\leqslant k}\left|(1+|x+2\pi g|)^N D^\alpha\tilde{\psi}(x)\right| \\ &\leqslant C' \sup_{x\in\mathbb{R}^n;|\alpha|\leqslant k}\left|(1+|x|)^N D^\alpha\tilde{\psi}(x)\right| \times (1+|g|)^N \\ &\leqslant K(1+|g|)^N \end{aligned}$$

With this bound, it is a straightforward exercise to verify that the sum converges in $\mathscr{S}'$. □

Exercise(∗). Show that if c_g satisfy:

$$|c_g| \leqslant K(1+|g|)^N$$

for some $K > 0$ and $N \in \mathbb{N}$, then:

$$\sum_{g\in\mathbb{Z}^n} c_g\delta_{2\pi g}$$

converges in $\mathscr{S}'$.

This now enables us to establish a result regarding the Fourier series of a periodic distribution.

Theorem 4.17. *Suppose $u \in \mathscr{D}'(\mathbb{R}^n)$ is a periodic distribution. Then there exist constants $c_g \in \mathbb{C}$ such that:*

$$u = \sum_{g \in \mathbb{Z}^n} c_g T_{e_{2\pi g}}.$$

with c_g are given by:

$$c_g = M(e_{-2\pi g} u).$$

and satisfy the bound:

$$|c_g| \leqslant K(1 + |g|)^N \tag{4.11}$$

for some $K > 0$, $N \in \mathbb{Z}$.

Proof. Since u is periodic, it is tempered by Lemma 4.15. Thus we may take the Fourier transform. Noting that:

$$\tau_{g'} u = u$$

for all $g' \in \mathbb{Z}^n$, we have that

$$e_{-g'} \hat{u} = \hat{u} \qquad \Longrightarrow \qquad (e_{-g'} - 1)\hat{u} = 0.$$

By Lemma 4.16, we deduce that:

$$\hat{u} = (2\pi)^n \sum_{g \in \mathbb{Z}^n} c_g \delta_{2\pi g},$$

for some c_g satisfying (4.11), where the sum converges in $\mathscr{S}'$. We can apply the inverse Fourier transform, making use of the fact that it is continuous on $\mathscr{S}'$ to deduce:

$$u = \sum_{g \in \mathbb{Z}^n} c_g T_{e_{2\pi g}},$$

with convergence again in $\mathscr{S}'$. To establish the formula for c_g, we make use of the comments after Lemma 4.14 to note that:

$$M(e_{-2\pi g} T_{e_{2\pi g'}}) = \int_{\overline{q}} e^{2\pi i (g - g')\cdot x} dx = \delta_{gg'}$$

Since $u \mapsto M(e_{-2\pi g} u)$ is a continuous map from $\mathscr{S}'$ to $\mathbb{C}$, we deduce that:

$$M(e_{-2\pi g} u) = \sum_{g' \in \mathbb{Z}^n} c_g M(e_{-2\pi g} T_{e_{2\pi g'}}) = c_{g'}.$$

□

Remark. *Usually one writes the Fourier series for u as:*

$$u = \sum_{g \in \mathbb{Z}^n} c_g e_{2\pi g},$$

ignoring the distinction between the function $e_{2\pi g}$ and the distribution it defines.

As a simple example, let us consider the distribution:

$$u = \sum_{g \in \mathbb{Z}^n} \delta_g.$$

By Lemma 4.15, this defines a periodic distribution, since $\delta_g = \tau_g \delta_0$ and $\delta_0 \in \mathscr{E}'(\mathbb{R}^n)$. Notice also that if ψ satisfies the conditions of Lemma 4.14, then since $\operatorname{supp} \psi \subset \{x \in \mathbb{R}^n : |x_j| < 1\}$, we have that $\tau_g \psi(0) = 0$ for $g \in \mathbb{Z}^n$ with $g \neq 0$. Thus, since $\sum_{g \in \mathbb{Z}^n} \tau_g \psi = 1$, we must have $\psi(0) = 1$. We can then calculate:

$$c_g = M(e_{-2\pi g} u) = u[\psi e_{-2\pi g}] = \psi(0) e^{-2\pi i g \cdot 0} = 1.$$

Thus we have established *Poisson's formula*:

$$\sum_{g \in \mathbb{Z}^n} \delta_g = \sum_{g \in \mathbb{Z}^n} T_{e_{2\pi g}},$$

where we understand both sums to converge in $\mathscr{S}'$. This is sometimes written, with an abuse of notation:

$$\sum_{g \in \mathbb{Z}^n} \delta(x - g) = \sum_{g \in \mathbb{Z}^n} e^{2\pi g \cdot x i}$$

We can specialise various results concerning Fourier transforms to the case of Fourier series.

Corollary 4.18. *i) Suppose $u \in \mathscr{D}'(\mathbb{R}^n)$ is periodic and may be written as:*

$$u = \sum_{g \in \mathbb{Z}^n} c_g T_{e_{2\pi g}}.$$

Then $D_j u \in \mathscr{D}'(R^n)$ is periodic and has Fourier series:

$$D_j u = \sum_{g \in \mathbb{Z}^n} (2\pi i g_j c_g) T_{e_{2\pi g}}.$$

ii) Suppose $f \in L^1_{loc.}(\mathbb{R}^n)$, then:

$$|c_g| \leqslant \|f\|_{L^1(q)},$$

and moreover, $c_g \to 0$ as $|g| \to \infty$.

iii) Suppose $f \in C^{n+1}(\mathbb{R}^n)$ is periodic. Then:

$$f(x) = \sum_{g \in \mathbb{Z}^n} c_g e^{2\pi i g \cdot x}$$

with the sum converging uniformly.

iv) Suppose $f, h \in L^2_{loc.}(\mathbb{R}^n)$ are periodic with Fourier coefficients f_g, h_g respectively. Then:

$$\int_q \overline{f}(x)h(x)dx = \sum_{g\in\mathbb{Z}^n} \overline{f}_g h_g.$$

This is the Fourier series version of Parseval's formula. Moreover,

$$f(x) = \sum_{g\in\mathbb{Z}^n} c_g e^{2\pi i g\cdot x}$$

holds, with the sum converging in $L^2(q)$.

Proof. i) Since the Fourier series for u converges in $\mathscr{S}'$, we may differentiate term by term (as differentiation is a continuous operation from $\mathscr{S}'$ to itself). Since

$$D_j T_{e_{2\pi g}} = (2\pi i g_j) T_{e_{2\pi g}},$$

the result follows.

ii) Note that if $f \in L^1_{loc.}(\mathbb{R}^n)$, then:

$$|c_g| = \left| \int_q e^{-2\pi i g\cdot x} f(x)dx \right| \leqslant \int_q |f(x)|\, dx = \|f\|_{L^1(q)}\,.$$

Now, given $\epsilon > 0$, we can approximate[4] f by a smooth periodic function f_ϵ, with Fourier coefficients c'_g, such that

$$\|f - f_\epsilon\|_{L^1(q)} < \frac{\epsilon}{2}.$$

Since $D_j D_j f_\epsilon \in L^1_{loc.}(\mathbb{R}^n)$, we have that $|g|^2 \left|c'_g\right| < C$, for each $j = 1, \dots, n$ so there exists $R > 0$ such that $\left|c'_g\right| < \frac{\epsilon}{2}$ for $|g| > R$. We have:

$$\left|c_g - c'_g\right| \leqslant \|f - f_\epsilon\|_{L^1(q)} < \frac{\epsilon}{2},$$

so we conclude that for $|g| > R$:

$$|c_g| = \left|c_g - c'_g + c'_g\right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus $c_g \to 0$ as $|g| \to \infty$.

iii) Since $f \in C^{k+1}(\mathbb{R}^n)$, we have that $D^\alpha f \in L^1_{loc.}(\mathbb{R}^n)$ for $|\alpha| < n + 1$. Applying the previous two results we conclude that $|c_g| \leqslant K(1 + |g|)^{-n+1}$ for some $K > 0$. Thus the partial sums:

$$F_n(x) = \sum_{g\in\mathbb{Z}^n, |g|\leqslant n} c_g e^{2\pi i g\cdot x}$$

[4] See Exercise 4.1

converge uniformly to some continuous function F by the Weierstrass $M-$test. We have:

$$T_f = \lim_{n\to\infty} \sum_{g\in\mathbb{Z}^n, |g|\leqslant n} c_g T_{e_{2\pi g}} = \lim_{n\to\infty} T_{F_n} = T_F$$

since uniform convergence implies convergence in $\mathscr{S}'$. By the injectivity of the mapping between continuous functions and distributions we conclude $f = F$.

iv) Suppose $f, h \in C^\infty(\mathbb{R}^n)$ are periodic. Then:

$$f(x) = \sum_{g\in\mathbb{Z}^n} f_g e^{2\pi i g\cdot x}, \qquad h(x) = \sum_{g\in\mathbb{Z}^n} h_g e^{2\pi i g\cdot x}$$

with $\sup_{g\in\mathbb{Z}^n}(1+|g|)^N |f_g| < \infty$ for all $N \in \mathbb{N}$, and similarly for h_g. We calculate:

$$\begin{aligned}\int_q \overline{f}(x)g(x)dx &= \int_q \left(\sum_{g\in\mathbb{Z}^n} \overline{f}_g e^{-2\pi i g\cdot x}\right)\left(\sum_{g'\in\mathbb{Z}^n} h_{g'} e^{2\pi i g'\cdot x}\right) dx \\ &= \sum_{g\in\mathbb{Z}^n}\sum_{g'\in\mathbb{Z}^n} \overline{f}_g h_{g'} \int_q e^{2\pi i (g'-g)\cdot x} dx \\ &= \sum_{g\in\mathbb{Z}^n}\sum_{g'\in\mathbb{Z}^n} \overline{f}_g h_{g'} \delta_{gg'} = \sum_{g\in\mathbb{Z}^n} \overline{f}_g h_g.\end{aligned}$$

In particular, we have that:

$$\|f\|_{L^2(q)} = \|f_g\|_{\ell^2(\mathbb{Z}^n)},$$

where for a sequence $\{a_g\}_{g\in\mathbb{Z}^n}$, we define:

$$\|a_g\|_{\ell^2(\mathbb{Z}^n)} = \left(\sum_{g\in\mathbb{Z}^n} |a_g|^2\right)^{\frac{1}{2}}.$$

Now suppose $f \in L^2_{loc.}(\mathbb{R}^n)$. Given $k > 0$, we can find $f^{(k)} \in C^\infty(\mathbb{R}^n)$ with Fourier coefficients $f_g^{(k)}$ such that:

$$\left\|f - f^{(k)}\right\|_{L^2(q)} < \frac{1}{k}.$$

Since by Cauchy-Schwarz we have:

$$\|f\|_{L^1(q)} = \int_q |f(x)|\, dx \leqslant \left(\int_q |f(x)|^2\, dx\right)^{\frac{1}{2}} \left(\int_q dx\right)^{\frac{1}{2}} = \|f\|_{L^2(q)}$$

we have that:

$$\left\|f_g - f_g^{(k)}\right\|_{\ell^\infty(\mathbb{Z}^n)} := \sup_{g\in\mathbb{Z}^n} \left|f_g - f_g^{(k)}\right| < \frac{1}{k},$$

Now, $f^{(k)}$ is a Cauchy sequence in $L^2(q)$, so $\{f_g^{(k)}\}$ is a Cauchy sequence in $\ell^2(\mathbb{Z}^n)$. We conclude that $f_g^{(k)}$ converges in $\ell^2(\mathbb{Z}^n)$, however we also know that $f^{(k)} \to f$

in $\ell^\infty(\mathbb{Z}^n)$, thus we must have $f^{(k)} \to f$ in $\ell^2(\mathbb{Z}^n)$. Taking a similar sequence of $h^{(k)} \in C^\infty(\mathbb{R}^n)$ approximating h with Fourier coefficients $h_g^{(k)}$, and recalling that:

$$\left(f^{(k)}, h^{(k)}\right)_{L^2(q)} = \left(f^{(k)}, h^{(k)}\right)_{\ell^2(\mathbb{Z}^n)},$$

the result follows on sending $k \to \infty$. The convergence of the Fourier series in $L^2(q)$ follows by showing that the partial sums form a Cauchy sequence in $L^2(q)$.

□

Exercise 4.1. Suppose $f \in L^p_{loc.}(\mathbb{R}^n)$ is a periodic function and let:

$$q = \left\{x \in \mathbb{R}^n : |x_j| < \frac{1}{2}, j = 1, \ldots, n\right\}.$$

Show that for any $\epsilon > 0$ there exists a smooth, periodic, function f_ϵ such that

$$\|f - f_\epsilon\|_{L^p(q)} < \epsilon.$$

Exercise 4.2. Let $u \in \mathscr{S}'(\mathbb{R})$ be the periodic distribution $u = \sum_{n=-\infty}^{\infty} \delta_n$, and suppose α is irrational. Let $w_N = \frac{1}{N}\sum_{n=1}^{N} \tau_{n\alpha} u$. By considering $\hat{w}_N$, or otherwise, show that w_N converges in $\mathscr{S}'(\mathbb{R})$ to a constant distribution. This is *Weyl's equidistribution theorem.*

Exercise(∗). Suppose that $f : \mathbb{R} \to \mathbb{R}$ is given by:

$$f(x) = x \quad \text{for } |x| < \frac{1}{2}, \qquad f(x+1) = f(x).$$

Show that:

$$f(x) = \sum_{n \in \mathbb{Z}, n \neq 0} \frac{i(-1)^n}{2\pi n} e^{2\pi i n x} = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n\pi} \sin(2\pi n x),$$

with convergence in $L^2_{loc.}(\mathbb{R})$.

Exercise(∗). Suppose $f : \mathbb{R} \to \mathbb{R}$ is given by:

$$f(x) = \begin{cases} -1 & -\frac{1}{2} < x \leqslant 0 \\ 1 & 0 < x \leqslant \frac{1}{2} \end{cases}, \qquad f(x+1) = f(x).$$

a) Show that:

$$f(x) = \frac{1}{\pi i} \sum_{n=-\infty}^{\infty} \frac{2}{2n+1} e^{2\pi i(2n+1)x} = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin\left[2\pi(2n+1)x\right]$$

With convergence in $L^2_{loc.}(\mathbb{R}^n)$.

Define the partial sum:

$$S_N(x) = 8\sum_{n=0}^{N-1} \frac{1}{2\pi(2n+1)} \sin\left[2\pi(2n+1)x\right].$$

b) Show that:

$$S_N(x) = 8\int_0^x \sum_{n=0}^{N-1} \cos\left[2\pi(2n+1)t\right] dt.$$

c) Show that:

$$\cos\left[2\pi(2n+1)t\right] \sin 2\pi t = \frac{1}{2}\left(\sin\left[2\pi(2n+2)t\right] - \sin\left[4\pi n t\right]\right)$$

And deduce:

$$S_N(x) = 8\int_0^x \frac{\sin 4\pi N t}{2\sin 2\pi t} dt.$$

d) Show that the first local maximum of S_N occurs at $x = \frac{1}{4N}$, and:

$$S_N\left(\frac{1}{4N}\right) \geqslant 8\int_0^{\frac{1}{4N}} \frac{\sin 4\pi N t}{4\pi t} dt = \frac{2}{\pi}\int_0^{\pi} \frac{\sin s}{s} ds \simeq 1.179\ldots$$

e) Conclude that the sum in part a) does not converge uniformly.

This lack of uniform convergence of a Fourier series at a point of discontinuity is known as Gibbs Phenomenon.

4.6 Sobolev spaces

4.6.1 The spaces $W^{k,p}(\Omega)$

Suppose $\Omega \subset \mathbb{R}^n$ is an open set. For $k \in \mathbb{Z}_{\geqslant 0}$ and $1 \leqslant p \leqslant \infty$, we say that $f \in L^p(\Omega)$ belongs to the Sobolev space $W^{k,p}(\Omega)$ if for any $|\alpha| \leqslant k$ there exists $f^\alpha \in L^p(\Omega)$ with:

$$D^\alpha T_f = T_{f^\alpha}.$$

We call f^α the weak, or distributional derivative of f and write $D^\alpha f := f^\alpha$. We can equip $W^{k,p}(\Omega)$ with the norm:

$$\|f\|_{W^{k,p}(\Omega)} := \sum_{|\alpha| \leqslant k} \|D^\alpha f\|_{L^p(\Omega)}.$$

With this norm, $W^{k,p}(\Omega)$ is complete, and hence a Banach space. The Sobolev spaces are particularly well suited to the study of PDE, and form the starting point for many modern PDE investigations.

We can think of k as telling us how *differentiable* our function is, while p tells us how *integrable* our function is. Roughly speaking spaces with larger k contain smoother functions, while spaces with larger p contain less 'spiky' functions. We shall see that (roughly speaking) one can trade smoothness for integrability: a function that belongs to $W^{k,p}(\mathbb{R}^n)$ belongs to certain $W^{l,q}(\mathbb{R}^n)$ where $l < k$ and $p > q$. If k and p are large enough we can even conclude that the function must be classically differentiable.

We will frame the result as concerning the embedding of $W^{k,p}(\mathbb{R}^n)$ spaces. Recall that a Banach space $(X, \|\cdot\|_X)$ is said to embed continuously into the Banach space $(Y, \|\cdot\|_Y)$ if $X \subset Y$ and there exists a constant C such that:

$$\|x\|_Y \leqslant C \|x\|_X, \qquad \text{for all } x \in X.$$

Theorem 4.19 (Sobolev embedding theorem)**.** *Suppose $k > l$ and $1 \leqslant p < q < \infty$ satisfy $(k-l)p < n$ and:*

$$\frac{1}{q} = \frac{1}{p} - \frac{k-l}{n}.$$

Then $W^{k,p}(\mathbb{R}^n)$ embeds continuously into $W^{l,q}(\mathbb{R}^n)$.

If $kp > n$, then $W^{k,p}(\mathbb{R}^n)$ embeds continuously into the Hölder space $C^{k-\left[\frac{n}{p}\right]-1,\gamma}(\mathbb{R}^n)$, where $[x]$ is the largest integer less than or equal to x, and

$$\gamma = \begin{cases} \left[\frac{n}{p}\right] + 1 - \frac{n}{p} & \frac{n}{p} \notin \mathbb{Z}, \\ \textit{any element of } (0,1) & \frac{n}{p} \in \mathbb{Z}. \end{cases}$$

Here we have introduced the Hölder space $C^{m,\kappa}(\mathbb{R}^n)$ which consists of $f \in C^m(\mathbb{R}^n)$ such that:

$$\|f\|_{C^{m,\kappa}(\mathbb{R}^n)} := \sum_{\alpha \leqslant m} \sup_{x \in \mathbb{R}^n} |D^\alpha f(x)| + \sum_{\alpha = m} \sup_{x,y \in \mathbb{R}^n} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x-y|^\kappa} < \infty.$$

We shan't attempt to prove the general Sobolev embedding theorems, but will establish a special case later on.

> **Exercise 4.3.** Suppose that $\Omega \subset \mathbb{R}^n$ is open and bounded, let $f \in C_c^\infty(\Omega)$, and suppose $0 < \epsilon < 1$.
>
> a) Show that $\int_\Omega (|f|^2 + \epsilon)^{\frac{p}{2}} dx \to \|f\|_{L^p}^p$ as $\epsilon \to 0$.
>
> b) By considering $\int_\Omega (|f|^2 + \epsilon)^{\frac{p}{2}} dx = \int_{\mathbb{R}^n} \left(\frac{1}{n} \text{div } x\right) (|f|^2 + \epsilon)^{\frac{p}{2}} dx$, or otherwise, show that there exists a constant C, depending on Ω, p but not on f, such that
> $$\|f\|_{L^p} \leq C \|Df\|_{L^p}.$$

4.6.2 The space $H^s(\mathbb{R}^n)$

We shall immediately specialise to the case $p = 2$ and $\Omega = \mathbb{R}^n$. This is an important special case for two reasons. Firstly, $W^{k,2}(\Omega)$ is a Hilbert space (in addition to being a Banach space), and so carries additional structure. Secondly, $W^{k,2}(\mathbb{R}^n)$ is very well adapted to the Fourier transform. To see this, we recall that if $f \in L^2(\mathbb{R}^n)$, then:

$$\widehat{T_f} = T_{\hat{f}}$$

where $\hat{f} \in L^2(\mathbb{R}^n)$ is the Fourier-Plancherel transform of f. We immediately obtain an alternative characterisation of the space $W^{k,2}(\mathbb{R}^n)$. A function $f \in L^2(\mathbb{R}^n)$ belongs to $W^{k,2}(\mathbb{R}^n)$ if and only if:

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

Notice that in this characterisation there is no need to restrict k to be an integer, nor in fact for f to belong to $L^2(\mathbb{R}^n)$. This motivates the following definition. For $s \in \mathbb{R}$ we say that $f \in \mathscr{S}'$ belongs to the space $H^s(\mathbb{R}^n)$ provided $\hat{f} \in L^2_{loc.}(\mathbb{R}^n)$ and:

$$\|f\|_{H^s(\mathbb{R}^n)} := \left(\int_{\mathbb{R}^n} \left(1 + |\xi|^2\right)^s \left|\hat{f}(\xi)\right|^2 d\xi\right)^{\frac{1}{2}} < \infty.$$

$H^s(\mathbb{R}^n)$ is complete, and moreover is a Hilbert space. We see that if $k \in \mathbb{Z}_{\geqslant 0}$ then $H^k(\mathbb{R}^n) = W^{k,2}(\mathbb{R}^n)$, where we make the canonical identification between a functions $f \in L^2(\mathbb{R}^n)$ and the distribution $T_f \in \mathscr{S}'(\mathbb{R}^n)$. From now on, we shall use f to mean both the function and the distribution.

Exercise 4.4. Let $s \in \mathbb{R}$.

a) Show that $\mathscr{S}$ is a dense subset of $H^s(\mathbb{R}^n)$.

b) Find a condition on s such that $\delta_x \in H^s(\mathbb{R}^n)$.

c) Show that $H^t(\mathbb{R}^n)$ is continuously embedded in $H^s(\mathbb{R}^n)$ for $s < t$.

d) Show that the derivative D^α is a bounded linear map from $H^{s+k}(\mathbb{R}^n)$ into $H^s(\mathbb{R}^n)$, where $k = |\alpha|$.

e) (*) Show that the pairing $\langle , \rangle : H^{-s}(\mathbb{R}^n) \times H^s(\mathbb{R}^n) \to \mathbb{C}$, which acts on $f \in H^{-s}(\mathbb{R}^n), g \in H^s(\mathbb{R}^n)$ by

$$\langle f, g\rangle = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\xi)\hat{g}(\xi) d\xi$$

is well defined, and show that the map $g \mapsto \langle f, g\rangle$ is a bounded linear operator on $H^s(\mathbb{R}^n)$. Deduce that $H^s(\mathbb{R}^n)'$ may be identified with $H^{-s}(\mathbb{R}^n)$. How does this relate to your answer to part b)?

4.6.3 Sobolev Embedding

An important feature of the Sobolev spaces $H^s(\mathbb{R}^n)$ is that for s sufficiently large, they embed into $C^k(\mathbb{R}^n)$. More previsely:

Theorem 4.20. *Fix $k \in \mathbb{Z}_{\geqslant 0}$. Suppose that $f \in H^s(\mathbb{R}^n)$ for some $s > k + \frac{n}{2}$, then (possibly after redefinition on a set of measure zero) $f \in C^k(\mathbb{R}^n)$. That is, we have:*

$$H^s(\mathbb{R}^n) \subset C^k(\mathbb{R}^n).$$

Proof. First suppose $f \in \mathscr{S}(\mathbb{R}^n)$. Then by the Fourier inversion theorem we have for $|\alpha| \leqslant k$:

$$D^\alpha f(x) = \frac{i^{|\alpha|}}{(2\pi)^n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \xi^\alpha \hat{f}(\xi) d\xi.$$

We estimate with the Cauchy-Schwarz inequality:

$$\begin{aligned} |D^\alpha f(x)| &= \frac{1}{(2\pi)^n} \left| \int_{\mathbb{R}^n} e^{ix\cdot\xi} \xi^\alpha \hat{f}(\xi) d\xi \right| \\ &\leqslant \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \left| \xi^\alpha \hat{f}(\xi) \right| d\xi \\ &\leqslant \frac{1}{(2\pi)^n} \left(\int_{\mathbb{R}^n} (1+|\xi|^2)^s \left| \hat{f}(\xi) \right|^2 d\xi \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} \frac{|\xi^\alpha|^2}{(1+|\xi|^2)^s} d\xi \right)^{\frac{1}{2}} \end{aligned}$$

Now, since $|\xi^\alpha|^2 \leqslant c_k (1+|\xi|^2)^k$ for some $c_k > 0$, we have that:

$$\frac{1}{(2\pi)^n} \left(\int_{\mathbb{R}^n} \frac{|\xi^\alpha|^2}{(1+|\xi|^2)^s} d\xi \right)^{\frac{1}{2}} \leqslant \frac{c_k}{(2\pi)^n} \left(\int_{\mathbb{R}^n} \frac{1}{(1+|\xi|^2)^{s-k}} d\xi \right)^{\frac{1}{2}} =: C_{n,k,s} < \infty$$

where we have used $s > k + \frac{n}{2}$ in order to ensure that the integral converges. We thus have that:

$$\sup_{|\alpha| \leqslant k, x \in \mathbb{R}^n} |D^\alpha f(x)| \leqslant C_{n,k,s} \|f\|_{H^s(\mathbb{R}^n)}. \tag{4.12}$$

Now suppose $f \in H^s(\mathbb{R}^n)$. We can approximate f by a sequence $(f_m)_{m=1}^\infty$ with $f_m \in \mathscr{S}(\mathbb{R}^n)$ and $f_m \to f$ in $H^s(\mathbb{R}^n)$ and pointwise almost everywhere. In particular, (f_m) is Cauchy in $H^s(\mathbb{R}^n)$, so by the estimate (4.12) applied to $f_m - f_l$ we have that (f_m) is Cauchy in $C^k(\mathbb{R}^n)$, thus there exists $f^* \in C^k(\mathbb{R}^n)$ such that $D^\alpha f_m \to D^\alpha f^*$ uniformly for all $|\alpha| \leqslant k$. Since $f_m \to f$ pointwise almost everywhere, we deduce that $f = f^*$ almost everywhere. □

Exercise 4.5. a) Suppose $s = \frac{n}{2} + \gamma$ for some $0 < \gamma < 1$. Show that there exists a constant $C_{n,\gamma} > 0$ such that for all $x, y \in \mathbb{R}^n$:

$$\int_{\mathbb{R}^n} \frac{\left|e^{ix\cdot\xi} - e^{iy\cdot\xi}\right|^2}{|\xi|^{2s}} d\xi \leqslant C_{n,\gamma} |x-y|^{2\gamma}$$

b) Show that if $s = \frac{n}{2} + k + \gamma$ for some $k \in \mathbb{Z}_{\geqslant 0}$, $0 < \gamma < 1$, then

$$H^s(\mathbb{R}^n) \subset C^{k,\gamma}(\mathbb{R}^n).$$

Exercise 4.6. Fix $s \in \mathbb{R}$, and suppose that $f \in H^s(\mathbb{R}^n)$.

a) Show that there exists a unique $u \in H^{s+4}(\mathbb{R}^n)$ which solves:

$$\Delta^2 u + u = f.$$

b) Show further that there exists $C > 0$ such that $\|u\|_{H^{s+4}} \leqslant C \|f\|_{H^s}$.

c) For what values of s does the equation hold in the sense of classical derivatives (possibly after redefining u, f on a set of measure zero)?

4.6.4 The trace theorem

We are often interested in the restriction of a function defined on $\mathbb{R}^n$, or some open subset, to some hypersurface $\Sigma \subset \mathbb{R}^n$. For example, when studying a PDE problem posed in some nice domain Ω we might wish to impose a boundary condition on $\partial\Omega$. If we work with functions in $H^s(\mathbb{R}^n)$ for $s > 0$, which are defined only almost everywhere, then this is a problem, since for nice domains $\partial\Omega$ will have Lebesgue measure zero. The trace theorem allows us to make sense of the restriction of a function in H^s to a hypersurface Σ, even when we don't have $f \in C^0$ by Sobolev embedding. We restrict to the problem of defining $f|_{\{x^n=0\}}$ when $f \in H^s(\mathbb{R}^n)$ is given, however by combining this result with coordinate transformations it is fairly easy to see how to generalise to the case of smoothly embedded submanifolds.

Theorem 4.21. *Let $s > \frac{1}{2}$. Then there is a bounded linear map $T : H^s(\mathbb{R}^n) \to H^{s-\frac{1}{2}}(\mathbb{R}^n)$ such that*

$$Tf = f|_{\{x^n=0\}}$$

for all $f \in H^s(\mathbb{R}^n) \cap C^0(\mathbb{R}^n)$.

Proof. See Exercise 4.7. □

Exercise 4.7. Assume $s > \frac{1}{2}$ and suppose $u \in \mathscr{S}(\mathbb{R}^n)$. Define $Tu \in \mathscr{S}(\mathbb{R}^{n-1})$ by:

$$Tu(x') = u(x', 0), \qquad x' \in \mathbb{R}^{n-1}.$$

a) Show that if $\xi' \in \mathbb{R}^{n-1}$:

$$\widehat{Tu}(\xi') = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{u}(\xi', \xi_n) d\xi_n.$$

b) Deduce that:

$$\left|\widehat{Tu}(\xi')\right|^2 \leqslant \frac{1}{(2\pi)^2} \left(\int_{\mathbb{R}} (1 + |\xi|^2)^s \left|\hat{u}(\xi', \xi_n)\right|^2 d\xi_n \right) \left(\int_{\mathbb{R}} \frac{d\xi_n}{\left(1 + |\xi|^2\right)^s} \right),$$

where $\xi = (\xi', \xi_n)$.

c) By changing variables in the second integral above to $\xi_n = t\sqrt{1 + |\xi'|^2}$, show that there exists a constant $C(s)$ such that:

$$\|Tu\|_{H^{s-\frac{1}{2}}(\mathbb{R}^{n-1})} \leqslant C(s) \|u\|_{H^s(\mathbb{R}^n)}.$$

d) Conclude that T extends to a bounded linear operator $T : H^s(\mathbb{R}^n) \to H^{s-\frac{1}{2}}(\mathbb{R}^{n-1})$.

e) (*) Suppose $v \in \mathscr{S}(\mathbb{R}^{n-1})$ and let $\phi \in C_c^\infty(\mathbb{R})$ satisfy $\int_{\mathbb{R}} \phi(t)dt = \sqrt{2\pi}$. Define u through its Fourier transform by:

$$\hat{u}(\xi', \xi_n) = \frac{\hat{v}(\xi')}{\sqrt{1+|\xi'|^2}} \phi \left(\frac{\xi_n}{\sqrt{1+|\xi'|^2}} \right).$$

Show that there exists a constant $C > 0$ such that:

$$\|u\|_{H^s(\mathbb{R}^n)} \leqslant C \|v\|_{H^{s-\frac{1}{2}}(\mathbb{R}^{n-1})}$$

and that $Tu = v$. Conclude that $T : H^s(\mathbb{R}^n) \to H^{s-\frac{1}{2}}(\mathbb{R}^{n-1})$ is surjective.

4.6.5 The space $H_0^1(\Omega)$

Suppose that Ω is an open subset of $\mathbb{R}^n$. For any function $f \in C_c^\infty(\Omega)$, we can trivially extend to an element of $C_c^\infty(\mathbb{R}^n)$ by $f(x) = 0$ for $x \in \Omega^c$, so can abuse notation slightly to denote by $C_c^\infty(\Omega)$ the space of smooth functions $f : \mathbb{R}^n \to \mathbb{C}$ with support in some compact $K \subset \Omega$. We define $H_0^1(\Omega)$ to be the completion of $C_c^\infty(\Omega)$ with respect to the $H^1(\mathbb{R}^n)$-norm. $H_0^1(\Omega)$ is a Hilbert space, equipped with the inner product:

$$(u, v)_{H^1} = \int_\Omega \left(\overline{Du(x)} \cdot Dv(x) + \overline{u(x)} v(x) \right) dx.$$

Let $u \in H_0^1(\Omega)$. Then by definition there exists a sequence $(\phi_n)_{n=1}^\infty$, with $\phi_n \in C_c^\infty(\Omega)$ and $\phi_n \to \phi$ in $H^1(\mathbb{R}^n)$. Since for any open $U \subset \mathbb{R}^n$ we have

$$\|f_n - f\|_{L^2(U)} \leqslant \|f_n - f\|_{L^2(\mathbb{R}^n)} \leqslant \|f_n - f\|_{H^1(\mathbb{R}^n)},$$

we deduce that $f_n|_U \to f|_U$ in L^2. If we choose $U = \Omega^c$, we conclude that if $f \in H_0^1(\Omega)$ then $f|_{\Omega^c} = 0$ almost everywhere.

If we assume the boundary of Ω is smooth, i.e. is an embedded smooth $(n-1)$-manifold, then we can make sense of the restriction of f to $\partial\Omega$ in the trace sense, and since the trace operator is a continuous map from $H^1(\mathbb{R}^n)$, we find that f vanishes on $\partial\Omega$ in the trace sense.

For many PDE problems, one wishes to solve some equation in an open set Ω, subject to the condition that the solution vanishes on the boundary of Ω. Seeking a solution in $H_0^1(\Omega)$ is often a convenient way to encode this boundary condition.

4.6.6 Rellich–Kondrachov

The Rellich–Kondrachov theorem is an important result concerning Sobolev spaces, with applications in PDE, calculus of variations and beyond. It concerns compact embedding for Sobolev spaces defined on a bounded domain. We shall prove a version of the result for the space $H_0^1(\Omega)$, where Ω is a bounded open set.

Theorem 4.22 (Rellich–Kondrachov)**.** *Suppose that Ω is a bounded open set and that $(u_i)_{i=1}^\infty$ is a bounded sequence in $H_0^1(\Omega)$. Then there exists $u \in H_0^1(\Omega)$ and a subsequence $(u_{i_j})_{j=1}^\infty$ such that:*

i) $u_{i_j} \rightharpoonup u$ *in* $H_0^1(\Omega)$*, and*

ii) $u_{i_j} \to u$ *in* $L^2(\Omega)$.

Proof. By assumption, we have that

$$\|u_i\|_{L^2(\Omega)} \leqslant \|u_i\|_{H_0^1(\Omega)} \leqslant K$$

so $(u_i)_{i=1}^\infty$ is bounded in both $H_0^1(\Omega)$, and $L^2(\Omega)$, and we immediately deduce from the Banach–Alaoglu theorem that there exists $u \in H_0^1(\Omega)$ and a subsequence $(u_{i_j})_{j=1}^\infty$ such that $u_{i_j} \rightharpoonup u$ in $H_0^1(\Omega)$, and $u_{i_j} \rightharpoonup u$ in $L^2(\Omega)$. For convenience, let us set $w_j = u_{i_j}$ so that $w_j \rightharpoonup u$ in $H_0^1(\Omega)$, and $w_j \rightharpoonup u$ in $L^2(\Omega)$. Thus our goal is to improve the weak-L^2 convergence of (w_j) to strong-L^2 convergence.

Fix $\epsilon > 0$. We make use of Parseval's Formula (Theorem 4.6) to give:

$$\begin{aligned}\|w_j - u\|_{L^2}^2 &= \frac{1}{(2\pi)^n} \|\hat{w}_j - \hat{u}\|_{L^2}^2 \\ &= \frac{1}{(2\pi)^n} \int_{|\xi|<R} |\hat{w}_j(\xi) - \hat{u}(\xi)|^2 \, d\xi + \frac{1}{(2\pi)^n} \int_{|\xi|\geqslant R} |\hat{w}_j(\xi) - \hat{u}(\xi)|^2 \, d\xi\end{aligned}$$

We deal with the two integrals on the final line separately. First we estimate:

$$\begin{aligned}\frac{1}{(2\pi)^n} \int_{|\xi|\geqslant R} |\hat{w}_j(\xi) - \hat{u}(\xi)|^2 \, d\xi &\leqslant \frac{2}{(2\pi)^n R^2} \int_{|\xi|\geqslant R} |\xi|^2 \left(|\hat{w}_j(\xi)|^2 + |\hat{u}(\xi)|^2\right) d\xi \\ &\leqslant \frac{2K^2}{(2\pi)^n R^2} < \epsilon,\end{aligned}$$

provided $R > 0$ is chosen sufficiently large.

Now consider the remaining integral that we need to bound. First, we note that

$$\hat{w}_j(\xi) = \int_\Omega w_j(x) e^{-ix\cdot\xi} dx = (w_i, e_{-\xi})_{L^2(\Omega)},$$

where we recall $e_y(x) = e^{ix\cdot y}$. Noting that $e_{-\xi} \in L^2(\Omega)$ since $|\Omega| < \infty$, and that $w_j \rightharpoonup u$ in $L^2(\Omega)$, we deduce that for each $\xi \in \mathbb{R}^n$:

$$\hat{w}_j(\xi) \to \hat{u}(\xi).$$

We can also estimate, for $|\xi| < R$:

$$\begin{aligned}|\hat{w}_j(\xi) - \hat{u}(\xi)|^2 &\leqslant 2\,|\hat{w}_j(\xi)|^2 + 2\,|\hat{u}(\xi)|^2 \leqslant 2\left(\|\hat{w}_j\|_{L^\infty}^2 + \|\hat{u}\|_{L^\infty}^2\right) \\ &\leqslant 2\left(\|w_j\|_{L^1(\Omega)}^2 + \|u\|_{L^1(\Omega)}^2\right) \leqslant 2|\Omega|\left(\|w_j\|_{L^2(\Omega)}^2 + \|u\|_{L^2(\Omega)}^2\right) \\ &\leqslant 4K^2|\Omega| \in L^1(B_R(0))\end{aligned}$$

So by the dominated convergence theorem we deduce that

$$\frac{1}{(2\pi)^n} \int_{|\xi|<R} |\hat{w}_j(\xi) - \hat{u}(\xi)|^2 \, d\xi \to 0$$

as $j \to \infty$, so that for j sufficiently large we have established:

$$\|w_j - u\|_{L^2}^2 < 2\epsilon$$

□

Corollary 4.23. *Let $\Omega \subset \mathbb{R}^n$ be open and bounded. Suppose that $L : L^2(\Omega) \to H_0^1(\Omega)$ is a bounded linear operator, then $L : L^2(\Omega) \to L^2(\Omega)$ is compact.*

Exercise 4.8. Suppose that $\lambda = \{\lambda_1, \ldots \lambda_n\}$ is a basis for $\mathbb{R}^n$. We define the lattice generated by λ to be $\Lambda = \left\{\sum_{j=1}^n z_j\lambda_j : z_j \in \mathbb{Z}\right\}$, and the the fundamental cell $q_\Lambda = \left\{\sum_{j=1}^n x_j\lambda_j : |x_j| < \frac{1}{2}\right\}$. We say that $u \in \mathscr{D}'(\mathbb{R}^n)$ is $\Lambda-$periodic if: $\tau_g u = u \qquad$ for all $g \in \Lambda$.

a) Show that there exists $\psi \in C_c^\infty(2q_\Lambda)$ such that $\psi \geqslant 0$ and $\sum_{g\in\Lambda} \tau_g\psi = 1$. If ψ, ψ' are two such functions and $u \in \mathscr{D}'(\mathbb{R}^n)$ is $\Lambda-$periodic, deduce

$$\frac{1}{|q_\Lambda|}u[\psi] = \frac{1}{|q_\Lambda|}u[\psi'] =: M(u).$$

b) Define the *dual lattice* by $\Lambda^* := \{x \in \mathbb{R}^n : g \cdot x \in 2\pi\mathbb{Z}, \ \forall g \in \Lambda\}$. Show that there exists a basis $\lambda^* = \{\lambda_1^*, \ldots \lambda_n^*\}$ such that $\lambda_j^* \cdot \lambda_k = 2\pi\delta_{jk}$, and Λ^* is the lattice generated by λ^*. Show that if $g \in \Lambda^*$ then e_g is $\Lambda-$periodic.

c) Show that if $u \in \mathscr{D}'(\mathbb{R}^n)$ is $\Lambda-$periodic, then $\hat{u} = \sum_{g\in\Lambda^*} c_g\delta_g$ for some $c_g \in \mathbb{C}$ satisfying $|c_g| \leqslant K(1 + |g|)^N$ for some $K > 0$, $N \in \mathbb{Z}$. Deduce that

$$u = \sum_{g\in\Lambda^*} d_g T_{e_g}$$

where $|d_g| \leqslant K(1 + |g|)^N$ for some $K > 0$, $N \in \mathbb{Z}$ are given by:

$$d_g = M(e_{-g}u)$$

Exercise 4.9. Suppose that $\Omega \subset \mathbb{R}^n$ is open and bounded. For $u \in H_0^1(\Omega)$, define the Dirichlet energy:

$$E[u] = \int_\Omega |Du|^2 \, dx.$$

a) Suppose that $(u_i)_{i=1}^\infty$ is a sequence with $u_i \in H_0^1(\Omega)$ such that $u_i \rightharpoonup u$. Show that $E[u] \leqslant \liminf_i E[u_i]$.

b) Consider the set

$$\mathcal{E}_1 = \{E[u] : u \in H_0^1(\Omega), \|u\|_{L^2} = 1\}$$

Let $\lambda_1 := \inf \mathcal{E}$. Show that there exists $w_1 \in H_0^1(\Omega)$ with $\|w_1\|_{L^2} = 1$ and $E[w_1] = \lambda_1$, and deduce $\lambda_1 > 0$.

c) Deduce that:

$$\lambda_1 \|u\|_{L^2}^2 \leqslant \int_\Omega |Du|^2 \, dx$$

holds for all $u \in H_0^1(\Omega)$, with equality for $u = w_1$. This is *Poincaré's inequality.*

d) By considering $u = w_1 + t\phi$ for $t \in \mathbb{R}$, $\phi \in \mathscr{D}(\Omega)$, or otherwise, show that w_1 satisfies

$$-\Delta w_1 = \lambda_1 w_1,$$

where we understand this equation as holding in $\mathscr{D}'(\Omega)$.

e) (*) Suppose $\chi \in C_c^\infty(\Omega)$, and let $v = \chi w_1$. Show that v satisfies $-\Delta v + v = f$, where we understand the equation as holding in $\mathscr{S}'(\mathbb{R}^n)$, where $f \in L^2(\mathbb{R}^n)$. Deduce that $v \in H^2(\mathbb{R}^n)$. By iterating this argument, deduce that $w_1 \in H_0^1(\Omega) \cap C^\infty(\Omega)$.

f) (*) By considering

$$\mathcal{E}_2 = \{E[u] : u \in H_0^1(\Omega), \|u\|_{L^2} = 1, (u, w_1)_{L^2} = 0\},$$

or otherwise, show that there exists $\lambda_2 \geqslant \lambda_1$ and $w_2 \in H_0^1(\Omega) \cap C^\infty(\Omega)$ with $w_2 \neq w_1$, $\|w_2\|_{L^2} = 1$ solving

$$-\Delta w_2 = \lambda_2 w_2.$$

4.7 PDE Examples

4.7.1 Elliptic equations on $\mathbb{R}^n$

Consider the following equation on $\mathbb{R}^n$, with $k > 0$:

$$-\Delta u + k^2 u = f,$$

where f is given and we wish to find u. Suppose that $f \in H^s(\mathbb{R}^n)$ for some $s \in \mathbb{R}$. We claim that there is a unique solution $u \in H^{s+2}(\mathbb{R}^n)$. Our assumptions on u, f permit us to take the Fourier transform of the equation so that:

$$(|\xi|^2 + k^2)\hat{u}(\xi) = \hat{f}(\xi)$$

holds pointwise almost everywhere. Since $|\xi|^2 + k^2 \geqslant C(1 + |\xi|^2) > 0$ for some C, we can divide through to find

$$\hat{u}(\xi) = \frac{\hat{f}(\xi)}{|\xi|^2 + k^2},$$

again using $|\xi|^2 + k^2 \geqslant C(1 + |\xi|^2) > 0$ we deduce:

$$\|u\|_{H^{s+2}(\mathbb{R}^n)} \leqslant C \|f\|_{H^s(\mathbb{R}^n)} .$$

Thus we indeed have that $u \in H^{s+2}(\mathbb{R}^n)$. Uniqueness follows from the injectivity of the Fourier transform. Note that if $s > \frac{n}{2}$ then $f \in C^0(\mathbb{R}^n)$ and $u \in C^2(\mathbb{R}^n)$, so that we in fact have a classical solution to the PDE. Note also that the solution is *more regular* than the data. This is an example of a phenomenon known as *elliptic regularity.*

4.7.2 Elliptic boundary value problems

Suppose that $\Omega \subset \mathbb{R}^n$ is open, assume $f : \Omega \to \mathbb{R}$ is given, and consider the equation:

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{4.13}$$

We wish to reformulate this so that we can solve it. In order to incorporate the boundary condition, we shall seek a solution $u \in H^1_0(\Omega)$. Since an element of $H^1_0(\Omega)$ only has weak derivatives in L^2 up to first order, we need to recast the equation in a form that makes sense. To do this, suppose we have a sufficiently regular solution, conjugate the equation and multiply it by $v \in C^\infty_c(\Omega)$ to deduce, after integrating by parts:

$$\int_\Omega \left(\overline{Du} \cdot Dv + \overline{u}v\right) dx = \int_\Omega \overline{f} v dx \tag{4.14}$$

holds for all $v \in C^\infty_c(\Omega)$. We realise that, if $f \in L^2(\Omega)$, we are seeking $u \in H^1_0(\Omega)$ such that:

$$(u, v)_{H^1} = (\overline{f}, v)_{L^2}$$

for all $v \in C^\infty_c(\Omega)$. We also notice that since $C^\infty_c(\Omega)$ is dense in $H^1_0(\Omega)$, this is equivalent to requiring the condition holds for $v \in H^1_0(\Omega)$. We say that $u \in H^1_0(\Omega)$ is a *weak solution* of (4.13) if (4.14) holds for all $v \in H^1_0(\Omega)$. Clearly, if u is a classical solution then it is a weak solution.

Now, for $f \in L^2(\Omega)$, the map $F : H^1_0(\Omega) \to \mathbb{C}$ given by $v \mapsto (\overline{f}, v)_{L^2}$ is a bounded linear operator, hence we can apply Riesz representation theorem for the Hilbert space $H^1_0(\Omega)$ to deduce that there exists a unique $\tilde{u} \in H^1_0(\Omega)$ such that $F(v) = (u, v)_{H^1}$ for all $v \in H^1_0(\Omega)$. This is precisely the solution we seek! In conclusion, then, we have shown:

Lemma 4.24. *Given $f \in L^2(\Omega)$ there exists a unique $u \in H^1_0(\Omega)$ solving (4.13) in the sense that (4.14) holds for all $v \in H^1_0(\Omega)$.*

We note that setting $v = u$ in (4.14), and using Cauchy-Schwarz we have:

$$\|u\|^2_{H^1} = (\overline{f}, u)_{L^2} \leqslant \|f\|_{L^2} \|u\|_{L^2} \leqslant \|f\|_{L^2} \|u\|_{H^1},$$

so that

$$\|u\|_{H^1} \leqslant \|f\|_{L^2}.$$

We will now show that we can improve the regularity of u, at least in the interior of Ω, provided we make some assumptions on f. For this, we introduce the space (here $k \in \mathbb{Z}_{\geqslant 0}$)

$$H^k_{loc.}(\Omega) = \left\{u : \Omega \to \mathbb{C} | \ \chi u \in H^k(\mathbb{R}^n), \text{ for all } \chi \in C^\infty_c(\Omega)\right\}$$

Fix a compact $K \subset \Omega$ and suppose that the real function $\chi \in C^\infty_c(\Omega)$ satisfies $\chi(x) = 1$ for $x \in K$. Let $\phi \in \mathscr{S}(\mathbb{R}^n)$, then since $\chi\phi \in C^\infty(\Omega)$, we can set $v = \chi\phi$ in (4.15):

$$\int_\Omega \overline{Du} \cdot D(\chi\phi) + \overline{u}\chi\phi dx = \int_\Omega \overline{f}\chi\phi dx$$

rearranging, we have:

$$\int_\Omega \overline{D(\chi u)} \cdot D\phi + \overline{Du} \cdot (D\chi)\phi - D\phi \cdot (D\chi)\overline{u} + \overline{u}\chi\phi dx = \int_\Omega \overline{f}\chi\phi dx$$

and hence:

$$\int_\Omega -\overline{(\chi u)}\Delta\phi + 2\overline{Du} \cdot (D\chi)\phi + \phi(\Delta\chi)\overline{u} + \overline{u}\chi\phi dx = \int_\Omega \overline{f}\chi\phi dx$$

So that $v = \chi u$ satisfies:

$$\int_{\mathbb{R}^n} \overline{v}(-\Delta\phi + 1)dx = \int_{\mathbb{R}^n} \overline{g}\phi dx,$$

where

$$g = -2Du \cdot (D\chi) - u\Delta\chi + f\chi \in L^2(\mathbb{R}^n).$$

We have deduced that $v \in H^1(\mathbb{R}^n)$ satisfies:

$$-\Delta v + v = g$$

in the sense of $\mathscr{S}'(\mathbb{R}^n)$. Now, by the results of the previous section, we deduce $v \in H^2(\mathbb{R}^n)$ with:

$$\|v\|_{H^2} \leqslant C \|g\|_{L^2}.$$

Further, $v(x) = u(x)$ for all $x \in K$. Suppose $\tilde{\chi} \in C_c^\infty(\Omega)$, then by applying the above argument with the compact set $K = \operatorname{supp} \tilde{\chi}$ we deduce that $\tilde{\chi}u = \tilde{\chi}\chi u \in H^2(\mathbb{R}^n)$. Thus $u \in H_0^1 \cap H^2_{loc.}(\Omega)$.

Now suppose that $f \in L^2 \cap H^1_{loc.}(\Omega)$. Repeating the above argument, we notice that $g \in H^1(\mathbb{R}^n)$, and so $v \in H^3(\mathbb{R}^n)$, and as a consequence we can conclude $u \in H_0^1 \cap H^3_{loc.}(\mathbb{R}^n)$. Iterating, we find:

Theorem 4.25. *Suppose $\Omega \subset \mathbb{R}^n$ is open and $f \in L^2 \cap H^k_{loc.}(\Omega)$. Then there exists a unique $u \in H_0^1 \cap H^{k+2}_{loc.}(\Omega)$ solving (4.13) in the weak sense. In particular, by Sobolev embedding if $f \in L^2 \cap C^\infty(\Omega)$, then $u \in H_0^1 \cap C^\infty(\Omega)$.*

Now, if $u \in C^\infty(\Omega)$, then we can see that the equation $-\Delta u + u = f$ must hold in Ω in the classical sense. If we assume more regularity of the boundary (and f), then we can also show that u extends to the boundary as a continuous function, and the boundary condition holds classically also. Discussing boundary regularity would take us beyond the remit of this course however.

We note, that our proof shows that the elliptic regularity phenomenon that we observed above for an equation on $\mathbb{R}^n$ is in fact localisable: if $(-\Delta u + u)$ is smooth in the interior of some open set, then u is smooth in that set. This is certainly not true for (for example) the wave operator $-\partial_t^2 + \Delta$. It is straightforward (try it!) to find a function that satisfies the wave equation in one dimension, hence $u_{tt} - u_{xx} = 0 \in C^\infty(\mathbb{R}^2)$, but for which $u \notin C^\infty(\mathbb{R}^2)$.

Spectral theory for elliptic boundary value problems

We now assume that $\Omega \subset \mathbb{R}^n$ is both open and *bounded.* Let us represent by A the map which takes $f \in L^2(\Omega)$ to the unique solution $u \in H^1_0(\Omega)$ to (4.13). We can check that A is linear, since if $u = Af$ and $w = Ag$ for some $f, g \in L^2(\Omega)$ and $a \in \mathbb{C}$, then for any $v \in H^1_0(\Omega)$ we have:

$$(u + aw, v)_{H^1} = (u, v)_{H^1} + \overline{a}(w, v)_{H^1} = (f, v)_{L^2} + \overline{a}(g, v)_{L^2} = (f + ag, v)_{L^2}$$

so that $Af + aAg = A(f + ag)$. Moreover, $A : L^2(\Omega) \to L^2(\Omega)$ is Hermitian. Suppose $u = Af$ and $w = Ag$ for some $f, g \in L^2(\Omega)$. Then

$$(f, Ag)_{L^2} = (f, w)_{L^2} = (u, w)_{H^1} = \overline{(w, u)_{H^1}} = \overline{(g, u)_{L^2}} = (Af, g)_{L^2}.$$

Finally, by Corollary 4.23 we have that $A : L^2(\Omega) \to L^2(\Omega)$ is compact. Thus by the spectral theorem for compact operators (see Linear Analysis), the spectrum of A takes the form $\sigma(A) = \{0, \mu_1, \mu_2, \ldots\}$, where $\mu_k \in \mathbb{R}$, $\mu_k \to 0$. Further, there exists an orthonormal basis for $L^2(\Omega)$ consisting of eigenvectors of A. An eigenvector of A satisfies $Aw = \mu w$ for $\mu \in \mathbb{R}$, and thus for $v \in H^1_0(\Omega)$:

$$(w, v)_{L^2} = (Aw, v)_{H^1} = \mu(w, v)_{H^1} \tag{4.15}$$

Setting $v = w$ we deduce $\mu > 0$, so in particular $\mu \neq 0$, and we deduce that w solves:

$$-\Delta w + w = \frac{1}{\mu} w$$

in the weak sense. This means that we can test the equation against elements of $H^1_0(\Omega)$ (alternatively, we can understand the equation as holding in $\mathscr{D}'(\Omega)$). Now, since $\mu^{-1} w \in H^1_0(\Omega)$, we conclude from our previous work that $w \in H^1_0 \cap H^3(\Omega)$. Hence $w \in H^1_0 \cap H^5(\Omega)$, etc. We conclude, after Sobolev embedding that $w \in C^\infty(\Omega)$.

Finally, noting that an eigenfunction of $(-\Delta + 1)$ is also an eigenfunction of $-\Delta$, we have shown:

Theorem 4.26. *Let $\Omega \subset \mathbb{R}^n$ be open and bounded. Then there exists an orthonormal basis $\{w_k\}_{k=1}^\infty$ for $L^2(\Omega)$ such that $w_k \in H^1_0 \cap C^\infty(\Omega)$ satisfy*

$$-\Delta w_k = \lambda_k w_k \qquad \text{in } \Omega,$$

where $\lambda_1 \leqslant \lambda_2 \leqslant \lambda_3 \leqslant \cdots$, and $\lambda_k \to \infty$. (In fact, by Exercise 4.9 we can show that $0 < \lambda_1$).

Exercise 4.10. Let H be the completion of $\mathscr{S}(\mathbb{R}^n)$ with respect to the norm

$$\|u\|_H := \left(\int_{\mathbb{R}^n} \left(|Du|^2 + |x|^2 |u|^2 \right) dx \right)^{\frac{1}{2}}$$

a) Show that H is a Hilbert space with the inner product:

$$(u, v)_H := \int_{\mathbb{R}^n} \left(\overline{Du} \cdot Dv + |x|^2 \overline{u} v \right) dx,$$

and show that if $u \in H, \chi \in C^\infty_c(B_R(0))$, then $\chi u \in H^1_0(B_R(0))$, with $\|\chi u\|_{H^1} \leqslant C_{R,\chi} \|u\|_H$ for some constant $C_{R,\chi} > 0$.

b) Show that H embeds compactly into $L^2(\mathbb{R}^n)$, that is $H \subset L^2(\mathbb{R}^n)$ and if $(u_n)_{n=1}^{\infty}$ is a bounded sequence in H then it admits a subsequence which converges in $L^2(\mathbb{R}^n)$.
[Hint: take a subsequence converging weakly in both H and $L^2(\mathbb{R}^n)$, and write $u_n = u_n\chi_R + u_n(1-\chi_R)$, where $\chi_R \in C_c^{\infty}(B_R(0))$ satisfies $\chi_R(x) = 1$ for $|x| < R-1$, where R is to be chosen.]

c) If $f \in L^2(\mathbb{R}^n)$, we say that $u \in H$ is a weak solution of:

$$-\Delta u + |x|^2 u = f \tag{$\dagger$}$$

if

$$(u, v)_H = (f, v)_{L^2} \text{ for all } v \in H. \tag{$\diamond$}$$

Show that if $u, f \in \mathscr{S}(\mathbb{R}^n)$ solve ($\dagger$), then u satisfies ($\diamond$). Show that for any $f \in L^2(\mathbb{R}^n)$, there exists a unique solution $u \in H$ to ($\diamond$).

d) Denote by Lf the unique solution $u \in H$ to ($\diamond$) for $f \in L^2(\mathbb{R}^n)$. Show that the map $f \mapsto Lf$ is a compact, symmetric, linear operator $L : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. Deduce that there exists an orthonormal basis $(w_k)_{k=1}^{\infty}$ for $L^2(\mathbb{R}^n)$ consisting of $w_k \in H$ satisfying:

$$(w_k, v)_H = \lambda_k (w_k, v)_{L^2} \text{ for all } v \in H, \tag{$\flat$}$$

where $0 < \lambda_1 \leqslant \lambda_2 \leqslant \lambda_3 \leqslant \cdots$, and $\lambda_k \to \infty$.

e) Show that if $w_k \in H$ satisfies ($\flat$), then in fact $w_k \in C^{\infty}(\mathbb{R}^n)$. Show further that $\hat{w}_k$ will also solve ($\flat$) with the same λ_k. Deduce that there exists an orthonormal basis for $L^2(\mathbb{R}^n)$, consisting of smooth functions, which diagonalises the Fourier–Plancherel transform.

f) (**) Show that $w \in H \cap C^{\infty}(\mathbb{R}^n)$ satisfies:

$$-\Delta w + |x|^2 w = \lambda w$$

for some $\lambda \in \mathbb{R}$ if and only if:

$$w(x) = cH_{k_1}(x_1) \cdots H_{k_n}(x_n) e^{-\frac{1}{2}|x|^2},$$

where $x = (x_1, \ldots, x_n)$, $c \in \mathbb{C}$, $H_k(t)$ are the Hermite polynomials, and $\lambda = n + 2k_1 + \ldots 2k_n$.
[Hint: treat the case $n = 1$ first. You may wish to look up the simple harmonic oscillator in a textbook on quantum mechanics.]

4.7.3 Spaces involving time

For certain PDE problems it's useful to separate out the time direction from the spatial directions. To do this, it's useful to introduce some new function spaces:

Definition 4.3. *Given a Banach space* $(X, \|\cdot\|_X)$*, and an interval* $I \subset \mathbb{R}$*, the space* $C^0(I; X)$ *is the space of continuous functions* $\boldsymbol{u} : I \to X$*.*

If I *is open, we define* $C^k(I; X)$ *for* $k \geqslant 0$ *inductively as follows. We say* $\boldsymbol{u} \in C^{k-1}(I; X)$ *belongs to* $C^k(I, X)$ *if there exists* $\boldsymbol{u}' \in C^{k-1}(I; X)$ *such that for each* $t \in I$*:*

$$\left\| \frac{\boldsymbol{u}(t+\epsilon) - \boldsymbol{u}(t)}{\epsilon} - \boldsymbol{u}'(t) \right\|_X \to 0, \qquad \text{as } \epsilon \to 0.$$

A typical example of X will be one of the space $H^s(\mathbb{R}^n)$ for $s > 0$.

4.7.4 The heat equation

Let us now give another example to show how powerful the Fourier transform can be for solving PDE problems. Let us consider the heat equation on $\mathbb{R}^n$. The problem we shall consider is, given $u_0 : \mathbb{R}^n \to \mathbb{R}$, determine $u : \mathbb{R}^n \times [0, T) \to \mathbb{R}$, such that

$$\begin{cases} u_t &= \Delta u \qquad \text{in } (0,T) \times \mathbb{R}^n, \\ u &= u_0 \qquad \text{on } \{0\} \times \mathbb{R}^n \end{cases} \tag{4.16}$$

We suppose that our solution is a continuous mapping from $(0, T)$ into $H^2(\mathbb{R}^n)$, i.e. for each fixed t we wish $u(t, \cdot) =: \boldsymbol{u}(t)$ to be an element of $H^2(\mathbb{R}^n)$. In terms of the function spaces above $\boldsymbol{u} \in C^0((0, T); H^2(\mathbb{R}^n))$. We will also suppose that u is continuously differentiable as a mapping from $(0, T)$ into $L^2(\mathbb{R}^n)$. In other words, $\boldsymbol{u} \in C^1((0, T); L^2(\mathbb{R}^n))$. Finally, we wish for the initial condition to make sense, so we also require $\boldsymbol{u} \in C^0([0, T); L^2(\mathbb{R}^n)$.

> **Exercise(∗).** Show that if $u \in C^0((0, T); H^2(\mathbb{R}^n)) \cap C^1((0, T); L^2(\mathbb{R}^n))$, then denoting by $\hat{u}$ the Fourier transform of u in the spatial variables:
>
> $$\hat{u}(t, \xi) = \lim_{R\to\infty} \int_{B_R(0)} u(t, x) e^{-ix\cdot\xi} dx,$$
>
> we have $\hat{u} \in C^0((0, T); L^2(\mathbb{R}^n)) \cap C^1((0, T); L^2(\mathbb{R}^n))$.

Let us, then, seek a solution of (4.16) such that

$$\boldsymbol{u} \in C^0([0, T); L^2(\mathbb{R}^n)) \cap C^0((0, T); H^2(\mathbb{R}^n)) \cap C^1((0, T); L^2(\mathbb{R}^n))$$

Under this assumption we can take the Fourier transform of (4.16) for $(t, x) \in (0, T) \times \mathbb{R}^n$ to get:

$$\begin{cases} \hat{u}_t(t, \xi) &= -|\xi|^2 \hat{u}(t, \xi) \qquad (t, \xi) \in (0, T) \times \mathbb{R}^n, \\ \hat{u}(0, \xi) &= \hat{u}_0(\xi) \qquad \xi \in \mathbb{R}^n \end{cases}$$

Now, the PDE has become an ODE for each fixed ξ! This ODE has a unique solution given for almost every $\xi \in \mathbb{R}^n$ by:

$$\hat{u}(t, \xi) = \hat{u}_0(\xi) e^{-t|\xi|^2}.$$

We note that if $u_0 \in L^2(\mathbb{R}^n)$, then $\hat{u}_0 \in L^2(\mathbb{R}^n)$ and thus $\hat{u} \in C^0([0, T); L^2(\mathbb{R}^n)) \cap C^1((0, T); L^2(\mathbb{R}^n))$. In fact, for $t > 0$, we have that $\hat{u}(t, \xi)$ and $\hat{u}_t(t, \xi)$ are rapidly decaying functions of ξ, in particular they belong to $H^s(\mathbb{R}^n)$ for any $s \geqslant 0$, so we have that $u(t,x)$ is smooth in x. Since u satisfies the equation $(\partial_t)^n u = (\Delta)^n u$, we have that u is smooth in both t and x. We can recover $u(t,x)$ via the inverse Fourier transform formula:

$$u(t,x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{u}_0(\xi) e^{-t|\xi|^2} e^{i\xi\cdot x} d\xi. \tag{4.17}$$

Summarising, we have the following result:

Lemma 4.27. *Suppose $u_0 \in L^2(\mathbb{R}^n)$. Then (4.16) admits a unique solution u such that*

$$\boldsymbol{u} \in C^0([0,T); L^2(\mathbb{R}^n)) \cap C^0((0,T); H^2(\mathbb{R}^n)) \cap C^1((0,T); L^2(\mathbb{R}^n))$$

given by (4.17). In fact,

$$u \in C^\infty((0,T) \times \mathbb{R}^n).$$

Even with very rough initial data, the heat equation instantaneously gives a smooth solution. This is an example of what is known as *parabolic regularity.*

Exercise 4.11. Suppose that $u_0 \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ and that $u(t,x)$ is the solution of the heat equation with initial data u_0. Explicitly, u is given by:

$$u(t,x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{u}_0(\xi) e^{-t|\xi|^2} e^{i\xi\cdot x} d\xi,$$

for $t > 0$.

a) Show that:

$$\|u(t,\cdot)\|_{L^2} \leqslant \|u_0\|_{L^2},$$

b) Show that:

$$u(t,x) = u_0 \star K_t(x)$$

where the *heat kernel* is given by:

$$K_t(x) = \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{|x|^2}{4t}}.$$

c) Suppose that $u_0 \geqslant 0$. Show that $u \geqslant 0$, and:

$$\|u(t,\cdot)\|_{L^1} = \|u_0\|_{L^1}.$$

Exercise 4.12. Consider the free Schrödinger equation:

$$\begin{cases} u_t = i\Delta u & \text{in } (0,T) \times \mathbb{R}^n, \\ u = u_0 & \text{on } \{0\} \times \mathbb{R}^n \end{cases} \tag{$*$}$$

Suppose $u_0 \in H^2(\mathbb{R}^n)$.

a) Show that $(*)$ admits a unique solution u such that

$$\boldsymbol{u} \in C^0([0,T); H^2(\mathbb{R}^n)) \cap C^1((0,T); L^2(\mathbb{R}^n)),$$

whose spatial Fourier-Plancherel transform is given by:

$$\hat{u}(t,\xi) = \hat{u}_0(\xi) e^{-it|\xi|^2}.$$

b) Show that:

$$\|u(t,\cdot)\|_{H^2(\mathbb{R}^n)} = \|u_0\|_{H^2(\mathbb{R}^n)}$$

*c) For $t > 0$, let $K_t \in L^1_{loc.}(\mathbb{R}^n)$ be given by:

$$K_t(x) = \frac{1}{(4\pi i t)^{\frac{n}{2}}} e^{\frac{i|x|^2}{4t}},$$

where for n odd we take the usual branch cut so that $i^{\frac{1}{2}} = e^{i\frac{\pi}{4}}$. For $\epsilon > 0$ set $K_t^\epsilon(x) = e^{-\epsilon|x|^2} K_t(x)$.

i) Show that $T_{K_t^\epsilon} \to T_{K_t}$ in $\mathscr{S}'$ as $\epsilon \to 0$.

ii) Show that if $\Re(\sigma) > 0$, then:

$$\int_{\mathbb{R}} e^{-\sigma x^2 - ix\xi} dx = \sqrt{\frac{\pi}{\sigma}} e^{-\frac{\xi^2}{4\sigma}}.$$

iii) Deduce that

$$\widehat{K_t^\epsilon}(\xi) = \left(\frac{1}{1+4it\epsilon}\right)^{\frac{n}{2}} e^{\frac{-it|\xi|^2}{1+4it\epsilon}}$$

iv) Conclude that:

$$\widehat{T_{K_t}} = T_{\tilde{K}_t},$$

where $\tilde{K}_t = e^{-it|\xi|^2}$.

*d) Suppose that $u \in \mathscr{S}(\mathbb{R}^n)$. Show that for $t > 0$:

$$u(t,x) = \int_{\mathbb{R}^n} u_0(y) K_t(x-y) dy,$$

and deduce that for $t > 0$:

$$\sup_{x \in \mathbb{R}^n} |u(t,x)| \leqslant \frac{1}{(4\pi t)^{\frac{n}{2}}} \|u_0\|_{L^1}.$$

This type of estimate which shows us that (locally) solutions to the Schrödinger equation decay in time is known as a *dispersive estimate*.

4.7.5 The wave equation

Now let us consider the wave equation on $\mathbb{R}^n$. The problem we shall consider is, given $u_0, u_1 : \mathbb{R}^n \to \mathbb{R}$, determine $u : \mathbb{R}^n \times (-T, T) \to \mathbb{R}$, such that

$$\begin{cases} u_{tt} &= \Delta u \quad \text{in } (-T,T) \times \mathbb{R}^n, \\ u &= u_0 \quad \text{on } \{0\} \times \mathbb{R}^n \\ u_t &= u_1 \quad \text{on } \{0\} \times \mathbb{R}^n \end{cases} \tag{4.18}$$

We will seek a solution in the space:

$$X_s := C^0((-T,T), H^{s+2}(\mathbb{R}^n)) \cap C^2((-T,T) \times H^s(\mathbb{R}^n)).$$

Fourier transforming in the spatial variable, we have:

$$\begin{cases} \hat{u}_{tt}(t,\xi) &= -|\xi|^2 \hat{u}(t,\xi) \quad (t,\xi) \in (-T,T) \times \mathbb{R}^n, \\ \hat{u}(0,\xi) &= \hat{u}_0(\xi) \quad \xi \in \mathbb{R}^n \\ \hat{u}_t(0,\xi) &= \hat{u}_1(\xi) \quad \xi \in \mathbb{R}^n \end{cases}$$

Again, this is an ODE for each fixed ξ, and we deduce:

$$\hat{u}(t,\xi) = \hat{u}_0(\xi) \cos(|\xi| t) + \hat{u}_1(\xi) \frac{\sin(|\xi| t)}{|\xi|}$$

Notice that if $u_0 \in H^{s+2}(\mathbb{R}^n)$ and $u_1 \in H^{s+1}(\mathbb{R}^n)$, then we conclude $\hat{u} \in X_s$. Thus (after taking the inverse Fourier transform) we have found the unique solution of the wave equation in X_s.

Let's specialise to $\mathbb{R}^3$. We'd like to write this solution as some sort of convolution, at least for initial data in the Schwarz class. For this we need to find the (inverse) Fourier transform of $\cos(|\xi| t)$ and $\frac{\sin(|\xi| t)}{|\xi|}$, where we have to understand these functions as tempered distributions. Let us define, for $t > 0$ the distribution:

$$U_t[\phi] = \frac{1}{4\pi t} \int_{\partial B_t(0)} \phi(y) d\sigma_y$$

for all $\phi \in \mathscr{S}'$, where $d\sigma_y$ is the surface measure on the sphere $\partial B_t(0)$. This is a distribution of compact support, so we can invoke Theorem 4.13 to find the Fourier transform:

$$\widehat{U_t} = T_{\hat{v}_t}$$

where:

$$\hat{v}_t(\xi) = U_t[e_{-\xi}] = \frac{1}{4\pi t} \int_{\partial B_t(0)} e^{-i\xi \cdot y} d\sigma_y$$

We can perform this integral by choosing spherical polar coordinates for y with the axis aligned with the vector ξ. Doing so, the integral becomes:

$$\begin{aligned} \hat{v}_t(\xi) &= \frac{1}{4\pi t} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} e^{-i|\xi| t \cos\theta} t^2 \sin\theta d\theta d\phi \\ &= \frac{t}{2} \int_{-1}^{1} e^{-i|\xi| t z} dz = \frac{t}{2} \left(\frac{e^{-i|\xi| t}}{-i|\xi| t} - \frac{e^{i|\xi| t}}{-i|\xi| t} \right) \\ &= \frac{\sin(|\xi| t)}{|\xi|}. \end{aligned}$$

Now, let us return to our expression for u:

$$\begin{aligned}\hat{u}(\xi) &= \hat{u}_0(\xi)\cos(|\xi|\,t) + \hat{u}_1(\xi)\frac{\sin(|\xi|\,t)}{|\xi|} \\ &= \frac{\partial}{\partial t}\left(\hat{u}_0(\xi)\frac{\sin(|\xi|\,t)}{|\xi|}\right) + \hat{u}_1(\xi)\frac{\sin(|\xi|\,t)}{|\xi|}\end{aligned}$$

Suppose $u_0, u_1 \in \mathscr{S}$. Then by Theorem 4.10, we have:

$$\begin{aligned}u(t,x) &= \frac{\partial}{\partial t}U_t \star u_0(x) + U_t \star u_1(x) \\ &= \frac{\partial}{\partial t}\left(\frac{1}{4\pi t}\int_{\partial B_t(0)} u_0(x-y)d\sigma_y\right) + \frac{1}{4\pi t}\int_{\partial B_t(0)} u_1(x-y)d\sigma_y \\ &= \frac{\partial}{\partial t}\left(\frac{1}{4\pi t}\int_{\partial B_t(x)} u_0(y)d\sigma_y\right) + \frac{1}{4\pi t}\int_{\partial B_t(x)} u_1(y)d\sigma_y \\ &= \frac{\partial}{\partial t}\left(t \fint_{\partial B_t(x)} u_0(y)d\sigma_y\right) + t\fint_{\partial B_t(x)} u_1(y)d\sigma_y \qquad (4.19)\end{aligned}$$

Where for a surface Σ with surface measure σ:

$$\fint_{\Sigma} d\sigma := \frac{1}{|\Sigma|}\oint_{\Sigma} d\sigma.$$

Expression (4.19) is known as Kirchoff's formula. While our derivation assumes $u_0, u_1 \in \mathscr{S}$, this assumption can be relaxed. This expression tells us some interesting facts about solutions to the wave equation. First note that the value of $u(x,t)$ depends only on the initial data on the sphere $\partial B_t(x)$. This is known as the strong Huygens principle. In particular this shows us that information is propagated at a *finite speed* by the wave equation. Secondly, note that the value of $u(x,t)$ depends on *derivatives* of u_0. This suggests that C^k-regularity is not propagated in wave evolution, although we have already seen that H^s-regularity is propagated.

Exercise($*$). Let $\mathbb{R}^3_* := \mathbb{R}^3 \setminus \{0\}$, $S_{*,T} := (-T,T)\times\mathbb{R}^3_*$ and $|x| = r$. You may assume the result that if $u = u(r,t)$ is radial, we have

$$\Delta u(|x|\,,t) = \Delta u(r,t) = \frac{\partial^2 u}{\partial r^2}(r,t) + \frac{2}{r}\frac{\partial u}{\partial r}(r,t)$$

a) Suppose $u(x,t) = \frac{1}{r}v(r,t)$ for some function v. Show that u solves the wave equation on $\mathbb{R}^3_* \times (0,T)$ if and only if v satisfies the one-dimensional wave equation

$$-\frac{\partial^2 v}{\partial t^2} + \frac{\partial^2 v}{\partial r^2} = 0$$

on $(0,\infty)\times(-T,T)$.

b) Suppose $f, g \in C^2_c(\mathbb{R})$. Deduce that

$$u(x,t) = \frac{f(r+t)}{r} + \frac{g(r-t)}{r}$$

is a solution of the wave equation on $S_{*,T}$ which vanishes for large $|x|$.

c) Show that if $f \in C^3_c(\mathbb{R})$ is an odd function (i.e. $f(s) = -f(-s)$ for all s) then

$$u(x,t) = \frac{f(r+t) + f(r-t)}{2r}$$

extends as a C^2 function which solves the wave equation on $S_T := (-T,T) \times \mathbb{R}^3$, with

$$u(0,t) = f'(t).$$

*d) By considering a suitable sequence of functions f, or otherwise, deduce that there exists no constant C independent of u such that the estimate

$$\sup_{S_T} (|u| + |u_t|) \leqslant C \sup_{\Sigma_0} (|u| + |u_t|)$$

holds for all solutions $u \in C^2(S_T)$ of the wave equation which vanish for large $|x|$.

Appendix A

Background Material: Functional Analysis

A.1 Topological vector spaces

This section is intended to recap some of the basic material in Linear Analysis, and to give a bit more detail on the functional analytical underpinnings of some of the more exotic spaces we consider in particular when constructing distributions. The material is not examinable, but is included here to justify various assertions earlier in the course.

In the Linear Analysis course, the principle objects of study are Hilbert or Banach spaces. These are vector spaces which are given the additional structure of an inner product or a norm respectively. This additional structure allows us to make sense of ideas such as convergence of a sequence, or continuity of a real valued map. Unfortunately, some of the vector spaces that we require for this course (for example $\mathscr{D}(U)$, $\mathscr{S}$ and $\mathscr{E}(U)$) are *not* Hilbert or Banach spaces. We need to add to the vector spaces some additional structure, which permits us to discuss the notions of convergence and continuity, but which is not as restrictive as assuming the presence of a norm or inner product. The extra structure that we shall require is of course a topology, but we shall require the topology to be in some sense *consistent* with the vector space structure. We are therefore led to the idea of *topological vector spaces.*

A.1.1 Vector spaces and normed spaces

In order to fix notation, let's recall a few standard definitions.

Definition A.1 (Field axioms)**.** *A field Φ is a set together with two operations, addition $+$ and multiplication $\cdot$ which satisfy the following axioms:*

i) Φ is closed *under addition and multiplication: for all $a, b \in \Phi$, we have $a + b \in \Phi$ and $a \cdot b \in \Phi$.*

ii) Both addition and multiplication are associative*: the following identities hold for all $a, b, c \in \Phi$:*

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c, \qquad a + (b + c) = (a + b) + c.$$

iii) Both addition and multiplication are commutative*: the following identities hold for all* $a, b \in \Phi$*:*

$$a \cdot b = b \cdot a, \qquad a + b = b + a.$$

iv) There exist unique, distinct, additive and multiplicative identity elements*: there exist* $0 \in \Phi$ *and* $1 \in \Phi$ *with* $0 \neq 1$ *such that for all* $a \in \Phi$ *we have:*

$$a \cdot 1 = a, \qquad a + 0 = a.$$

v) There exist additive and multiplicative inverses*. For every* $a \in \Phi$*, there exists an element* $(-a) \in \Phi$ *such that*

$$a + (-a) = 0.$$

Moreover, for every $a \in \Phi$ *with* $a \neq 0$*, there exists an element* a^{-1} *such that*

$$a \cdot a^{-1} = 1.$$

vi) The multiplication operation is distributive *over addition: the following identity holds for all* $a, b \in \Phi$*:*

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

> **Exercise(∗).** Show that $\mathbb{R}$, $\mathbb{C}$ and the integers modulo p, $\mathbb{Z}_p$ form fields with the usual definition of addition and multiplication.

The standard examples of fields that you should keep in mind for our purposes are $\mathbb{R}$ and $\mathbb{C}$. With the definition of a field in hand, we can now define a vector space.

Definition A.2 (Vector space axioms)**.** *Let* Φ *be a field, which we call the* scalar field*, and we call elements of* Φ scalars*. A vector space* X *over* Φ *is a set whose elements are called* vectors *together with two operations:*

i) ***Addition:*** *to every pair of vectors* $x, y \in X$ *is associated a unique vector* $x + y \in X$ *such that for all* $x, y, z \in X$*:*

$$x + y = y + x, \qquad \text{and} \qquad (x + y) + z = x + (y + z).$$

Moreover, there exists a unique element $0 \in X$ *such that for all* $x \in X$*:*

$$x + 0 = x.$$

Finally, for each $x \in X$*, there exists a unique vector* $(-x)$ *such that:*

$$x + (-x) = 0.$$

ii) ***Scalar multiplication:*** *to every pair* (a, x) *with* $a \in \Phi$ *and* $x \in X$ *is associate a unique vector* $ax \in X$ *in such a way that*

$$1x = x, \qquad a(bx) = (a \cdot b)x,$$

and such that the distributive laws:

$$a(x + y) = ax + ay, \qquad (a + b)x = ax + bx$$

hold for every $x, y \in X$ *and* $a, b \in \Phi$*.*

Note that the same symbols have different meanings in different contexts: $+$ can mean either scalar or vector multiplication, while 0 refers to the zero element of both the field and the vector space.

It is useful to extend the operations of vector addition and scalar multiplication to act on sets as follows. If $a \in X$, $\lambda \in \Phi$ $U_1, U_2 \subset X$, then we define:

$$\begin{aligned} a + U_1 &= \{a + x : x \in U\}, \\ U_1 + U_2 &= \{x + y : x \in U_1, y \in U_2\}, \\ \lambda U_1 &= \{\lambda x : x \in U_1\}. \end{aligned}$$

Note that $0 + U = U$, $1U = U$, $2U \subset U + U$, but that in general $2U \neq U + U$.

Definition A.3. *Suppose X is a vector space over Φ, where Φ is either $\mathbb{R}$ or $\mathbb{C}$. We say that a subset $U \subset X$ is* convex *if*

$$x, y \in U \quad \Longrightarrow \quad tx + (1-t)y \in U \text{ for all } t \in [0,1].$$

We say that U is balanced *if $\lambda U \subset U$ for all $\lambda \in \Phi$ with $|\lambda| \leqslant 1$.*

> **Exercise A.1.** Suppose that $\lambda_1 \lambda_2 \geqslant 0$ and that $U \subset X$ is a convex subset of a vector space X. Show that:
>
> $$\lambda_1 U + \lambda_2 U = (\lambda_1 + \lambda_2) U.$$

Finally, we shall define a norm on a vector space

Definition A.4. *A norm on a vector space X over Φ, where Φ is either $\mathbb{R}$ or $\mathbb{C}$ is a map:*

$$\|\cdot\| : X \to \mathbb{R}$$

such that

i) We have $\|x\| \geqslant 0$ for all $x \in X$, with equality if and only if $x = 0$.

ii) The triangle identity holds for all $x, y \in X$:

$$\|x + y\| \leqslant \|x\| + \|y\|.$$

iii) For any $a \in \Phi$ and $x \in X$ we have:

$$\|ax\| = |a| \, \|x\|.$$

A more general notion of distance than a norm is often useful. We define a metric space as follows:

Definition A.5. *A metric space (S, d) is a set S, together with a function $d : S \times S \to \mathbb{R}$, called the metric, which satisfies:*

i) The metric is symmetric*:*

$$d(x, y) = d(y, x), \qquad \textit{for all } x, y \in S.$$

ii) The metric is positive definite*:*

$$0 \leqslant d(x, y), \qquad \textit{for all } x, y \in S,$$

with equality if and only if $x = y$.

iii) The triangle inequality *holds:*

$$d(x, y) \leqslant d(x, z) + d(z, y), \qquad \textit{for all } x, y, z \in S.$$

To see that this is a more general notion than a normed space, we have the following result:

Lemma A.1. *If $(X, \|\cdot\|)$ is a normed vector space, then it is naturaly a metric space, with the metric:*

$$d(x, y) := \|x - y\|$$

Proof. We simply have to verify the three conditions on d. We find:

i) Noting that $|-1| = 1$, we have:

$$d(x, y) = \|x - y\| = \|(-1)(y - x)\| = |-1| \, \|y - x\| = \|y - x\| = d(y, x),$$

so the metric is symmetric.

ii) Since we know that $\|x\| \geqslant 0$, with equality if and only if $x = 0$, clearly

$$d(x, y) = \|x - y\| \geqslant 0$$

with equality if and only if $x - y = 0$, which holds if and only if $x = y$.

iii) Recall the triangle inequality for norms $\|x + y\| \leqslant \|x\| + \|y\|$. We calculate:

$$\begin{aligned} d(x, y) &= \|x - y\| = \|(x - z) - (y - z)\| \\ &\leqslant \|x - z\| + \|y - z\| \\ &= d(x, z) + d(z, y). \end{aligned}$$

Thus d satisfies the conditions to be a metric. □

A.1.2 Topological spaces

The definitions above are purely algebraic in nature. In particular, we have not introduced any notions of convergence, completeness or continuity for these spaces. The natural setting in which to do this is that of topology. Let us recall briefly a few definitions and facts.

Definition A.6 (Topology axioms)**.** *A topological space is a set S in which a collection of subsets τ (called* open sets*) has been specified, with the following properties:*

i) The empty set is open: $\emptyset \in \tau$

ii) The whole space is open: $S \in \tau$,

iii) If $U_1, U_2 \in \tau$ are open sets, then their intersection is open:

$$U_1 \cap U_2 \in \tau.$$

iv) If $\mathcal{U} \subset \tau$ is any *collection of open sets, then their union is open:*

$$\bigcup \mathcal{U} \in \tau.$$

Note that by repeatedly applying *iii*), we can easily see that any finite intersection of open sets is open. Let's recall some standard nomenclature associated with topological concepts. A set $E \subset S$ is closed if its complement $E^c = S \setminus E$ is open. The closure $\overline{E}$ of any set E is the intersection of all closed sets containing E. The interior E° of any set E is the union of all open sets contained in E. Note that the closure is always closed and the interior is always open. A *neighbourhood* of a point $p \in S$ is an open set containing p. A *limit point* of a set $E \subset S$ is a point $p \in S$ (not necessarily with $p \in E$) such that every neighbourhood of p intersects E in some point other than p itself.

Lemma A.2. *Suppose (S, τ) is a topological space. If $U \subset S$ is open then $U = U^\circ$. If $E \subset S$ is closed, then $\overline{E} = E$ and E contains all of its limit points.*

Proof. The fact that $U^\circ \subset U$ follows from the definition of the interior as U° is a union over sets contained in U. Since U is itself an open set contained in U, we also have $U \subset U^\circ$. Similarly, $E \subset \overline{E}$ from the definition of the closure. Since E is itself a closed set containing E, we have $\overline{E} \subset E$. Now suppose that p is a limit point, and assume for contradiction that $p \in E^c$. Then since E is closed, E^c is open and hence a neighbourhood of p. By the definition of a limit point we have $E \cap E^c$ is non-empty, a contradiction. Thus $p \in E$. □

A *base*, β for the topology τ is a collection of open sets, $\beta \subset \tau$ such that any open set in τ can be written as a union of elements of β. A collection γ of neighbourhoods of p is a *local base at p* if every neighbourhood of p contains a member of γ.

A set $K \subset S$ is compact if every open cover of K has a finite subcover. That is to say that from any collection $\{U_i\}_{i \in \mathcal{I}}$ of open sets such that $K \subset \cup_{i \in \mathcal{I}} U_i$, we can extract a finite collection $\{U_{i_k}\}_{k=1}^n$ such that $K \subset \cup_{k=1}^n U_{i_k}$. A topological space is *Hausdorff* if any two distinct points have disjoint neighbourhoods.

Lemma A.3. *Suppose (S, τ) is a Hausdorff topological space, and that $K \subset S$ is compact. Then K is closed.*

Proof. Let us fix $p \in K^c$, and consider an arbitrary $q \in K$. By the Hausdorff property of S, we know that there exist U_q, V_q open, with $q \in U_q$, $p \in V_q$ and $U_q \cap V_q = \emptyset$. Now, $\{U_q : q \in K\}$ is an open cover of K, hence by the compactness of K there is a finite subcover, i.e. $q_1, \ldots q_N$ such that $K \subset U = U_{q_1} \cup \ldots \cup U_{q_N}$. Consider $V = V_{q_1} \cap \ldots \cap V_{q_N}$. We have that $V_p \cup U = \emptyset$, so that $V \subset K^c$. Moreover, as a finite intersection of open sets V is open. Writing K^c as the union of the sets V for all $p \in K^c$, we see that K^c is open and thus K is closed. □

Lemma A.4. *Suppose (S, τ) is a Hausdorff topological space, and $E \subset S$. Then p is a limit point of E if and only if every neighbourhood of p contains infinitely many elements of E.*

Proof. If every neighbourhood of p contains infinitely many elements of E, it certainly intersects E in some point other than p, thus p is a limit point. Conversely, suppose that p is a limit point and suppose that U is some neighbourhood intersecting E in only finitely many points, say $\{x_1, \ldots, x_N\}$. by the Hausdorff property, we know that there exist open sets U_i, V_i such that $x_i \in U_i$, $p \in V_i$ and $U_i \cap V_i = \emptyset$. Then $\cap_{i=1}^N V_i \cap U$ is open, contains p and doesn't contain any other points of E. This contradicts the assumption that p is a limit point. □

A sequence $(x_n)_{n \in \mathbb{N}}$ in a Hausdorff space *converges* to a point x if every neighbourhood of x contains all but finitely many of the points x_n. If (S_1, τ_1) and (S_2, τ_2) are two topological spaces, then we say that $f : S_1 \to S_2$ is continuous if $f^{-1}(U) \in \tau_1$ for all $U \in \tau_2$. A *homeomorphism* $f : S_1 \to S_2$ is a bijective continuous map whose inverse is also continuous.

If τ_1, τ_2 are two different topologies on the same set S such that $\tau_1 \subset \tau_2$, then we say that τ_1 is a *coarser* topology than τ_2, or alternatively that τ_2 is a *finer* topology than τ_1. A finer topology has 'more open sets'. Note that if a sequence converges in τ_2 then it necessarily converges in τ_1 but that the converse does not hold. The coarsest topology on any set S is the trivial topology, whose only open sets are the empty set and S itself. The finest topology on any set is the discrete topology, for which any subset of S is declared to be open.

Exercise A.2. a) Suppose that (S, τ) is a topological space, and that β is a base for τ. Show that:

i) If $x \in S$, then there exists some $B \in \beta$ with $x \in B$.

ii) If B_1, $B_2 \in \beta$, then for every $x \in B_1 \cap B_2$ there exists $B \in \beta$ with:

$$x \in B \qquad B \subset B_1 \cap B_2.$$

b) Conversely, suppose that one is given a set S and a collection β of subsets of S satisfying $i)$, $ii)$ above. Define τ by:

$$U \in \tau \iff \text{for all } x \in U, \text{ there exists } B \in \beta \text{ such that } x \in B \text{ and } B \subset U.$$

i.e. τ is the set of all unions of elements of β. Show that (S, τ) is a topological space, with base β. We say that τ is the topology generated by β

c) Suppose that β, β' both satisfy conditions $i)$, $ii)$ above and generate topologies τ, τ' respectively. Moreover, suppose that if $B \in \beta$ then for every $x \in B$ there exists $B' \in \beta'$ satisfying

$$x \in B', \qquad \text{and} \qquad B' \subset B$$

Then $\tau \subset \tau'$.

If $E \subset S$ is any subset of a topological space (S, τ), then E inherits a topology, $\tau|_E$, called the subspace topology given by:

$$\tau|_E = \{E \cap U : U \in \tau\}.$$

If (S_1, τ_1) and (S_2, τ_2) are two topological spaces, then $S_1 \times S_2$ inherits a topology τ called the product topology, which is generated by the base

$$\beta = \{U_1 \times U_2 : U_i \in \tau_i, i = 1, 2\}$$

In other words, a set U is open in the product topology if it is the union of sets of the form $U_1 \times U_2$ with $U_i \in \tau_i$, $i = 1, 2$.

Exercise A.3. Suppose (S_1, τ_1), (S_2, τ_2) and (S_3, τ_3) are topological spaces, and that $f : S_1 \times S_2 \to S_3$ is a continuous map. Show that for each $a \in S_1$ and $b \in S_2$, the maps

$$\begin{aligned} f_a \quad &: \quad S_2 \to S_3, & \qquad f^b \quad &: \quad S_1 \to S_3, \\ & \quad\;\; y \mapsto f(a, y), & & \quad\;\; x \mapsto f(x, b), \end{aligned}$$

are continuous.

The condition that f is continuous with respect to the product topology is sometimes called *joint continuity*, while the continuity of f_a, f^b is called *separate continuity*. Thus joint continuity implies separate continuity. The converse is not true.

Theorem A.5. *Let (S_1, τ_1) and (S_2, τ_2) be two topological spaces, and let β_1 respectively β_2 be a base. Then the set*

$$\beta = \{B_1 \times B_2 : B_1 \in \beta_1, B_2 \in \beta_2\},$$

is a base for the product topology $(S_1 \times S_2, \tau)$.

Proof. Suppose $U \in \tau$, and let $x = (x_1, x_2) \in U$. By the definition of the product topology, there exist $U_1 \in \tau_1$ and $U_2 \in \tau_2$ with $x \in U_1 \times U_2$ and $U_1 \times U_2 \subset U$. Since β_1 is a base for (S_1, τ_1), and $U_1 \in \tau_1$, there exists $B_1 \in \beta_1$ with $x_1 \in B_1$ and $B_1 \subset U_1$. Similarly there exists $B_2 \in \beta_2$ such that $x_2 \in B_2$ and $B_2 \subset U_2$. Thus $x \in B_1 \times B_2$ and $B_1 \times B_2 \subset U_1 \times U_2 \subset U$. Considering these sets as x ranges over U, we see that U may be written as a union of elements of β and we're done. □

Example 19. *The real numbers* $\mathbb{R}$ *carry a topology, called the order topology, generated by the base:*

$$\beta_{\mathbb{R}} = \{(a,b) : a, b \in \mathbb{R},\ a < b\}\,.$$

This induces the product topology on $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$*. This is called the standard topology on* $\mathbb{R}^n$*.*

> **Exercise A.4.** Show that the base
>
> $$\beta_{\mathbb{Q}} = \{(p,q) : p, q \in \mathbb{Q},\ p < q\}\,,$$
>
> generates the standard topology on $\mathbb{R}$.

With the result of this exercise, we can establish the following very useful fact about open sets in $\mathbb{R}^n$.

Lemma A.6. *Suppose* $\Omega \subset \mathbb{R}^n$ *is open. Then there exists an* exhaustion of Ω by compact sets*. That is to say a family* $(K_i)_{i=1}^{\infty}$ *of compact sets* $K_i \subset \Omega$ *such that*

$$K_i \subset (K_{i+1})^{\circ}\,, \qquad \bigcup_{i=1}^{\infty} K_i = \Omega.$$

Proof.

1. Recall that by the definition of the product topology, a base for the standard topology of $\mathbb{R}^n$ is given by:

$$\beta = \{I_1 \times \cdots \times I_n : I_k \in \beta_{\mathbb{Q}}\}$$

For any $B \in \beta$ we have $B = \bigcup\{B' \in \beta : \overline{B'} \subset B\}$ since, for example

$$(p,q) = \bigcup_{n=N}^{\infty} \left(p + \frac{1}{n}, q - \frac{1}{n}\right)$$

for some $N > [2(q-p)]^{-1}$, and taking products of such sets the result follows.

2. Let

$$\beta' = \left\{B \in \beta : \overline{B} \subset \Omega\right\}.$$

Since β is a base, $\bigcup\{B \in \beta : B \subset \Omega\} = \Omega$, thus in view of the discussion above $\Omega = \bigcup \beta'$. Moreover, since β can be put into one-to-one correspondence with a subset of $\mathbb{Q}^{2n}$, we have that β and hence β' is countable.

3. Let us take an enumeration

$$\beta' = \{B_1, B_2, \ldots\}.$$

We define K_i inductively as follows. Pick $K_1 = \overline{B_1}$. This is a closed box in $\mathbb{R}^n$, so is compact. Now suppose that $K_1, \ldots K_n$ have been chosen. Since $K_n \subset \Omega$, β' is an open cover of K_n and so admits a finite subcover. Therefore there exists i_n such that $K_n \subset B_1 \cup \ldots \cup B_{i_n}$. We define $K_{n+1} = \overline{B_1} \cup \ldots \cup \overline{B_{i_n}}$. This is a union of closed boxes, hence is compact.

4. By construction we have $K_i \subset (K_{i+1})^\circ$. Moreover $K_{n+1} \not\subset B_1 \cup \ldots \cup B_{i_n}$, so $i_{n+1} > i_n$, and thus $i_n \to \infty$ as $n \to \infty$. Pick $x \in \Omega$. Then $x \in B_i$ for some i. Since $i_n \to \infty$, $x \in K_n$ for sufficiently large n, thus $\bigcup_{i=1}^{\infty} K_i = \Omega$.

□

For another example of a topological space, we return to the vector space setting.

Example 20. *Let (S, d) be a metric space. The open ball of radius $r > 0$ about $x \in S$ is defined to be:*

$$B_r(x) := \{y \in S : d(x, y) < r\}.$$

The metric topology *is the topology induced by the base:*

$$\beta = \{B_r(x) : x \in S, r \in \mathbb{R}_+\}.$$

We say that a general topological space (S, τ) is metrizable *if there exists some metric d on S such that the metric topoplogy of (S, d) coincides with τ.*

> **Exercise A.5.** Suppose that (S, d) is a metric space. Show that S is Hausdorff with respect to the metric topology.

An important feature of metric spaces is that the notions of compactness and *sequential compactness* are equivalent. We say that a topological space (S, τ) is sequentially compact if every sequence $(x_n)_{n=1}^{\infty}$ with $x_n \in S$ admits a subsequence $(x_{n_i})_{i=1}^{\infty}$ such that x_{n_i} converge to $x \in S$ as $i \to \infty$.

Theorem A.7. *Let (S, d) be a metric space endowed with the metric space topology. Then S is compact if and only if it is sequentially compact.*

Proof. 1. First suppose S is compact and consider the sequence $(x_n)_{n=1}^{\infty}$. We must exhibit a convergent subsequence. Let us consider the set $A = \{x_n\}_{n=1}^{\infty}$. If A is finite, then x_n must take at least one value an infinite number of times, so has a subsequence converging to that value, and we're done.

Now suppose A is infinite. We claim that A has a limit point. Suppose not. In particular, this means that each $y \in S$ has a neighbourhood U_y such that $U_y \cap A \subset \{y\}$. The collection $\{U_y : y \in S\}$ is an open cover of S, hence admits a finite subcover, say $\{U_{y_1}, \ldots U_{y_N}\}$. Note that we have

$$\begin{aligned} A = S \cap A &= (U_{y_1} \cup \ldots \cup U_{y_N}) \cap A \\ &= (U_{y_1} \cap A) \cup \ldots \cup (U_{y_N} \cap A) \subset \{y_1, \ldots, y_N\} \end{aligned}$$

Since A is infinite, this contradicts the assumption that A has no limit points.

Let x be a limit point of A. Since any metric space is Hausdorff, every neighbourhood of x must contain infinitely many points in A. Define a subsequence as follows. We pick n_1 such that $x_{n_1} \in B_1(x)$. Suppose we have $x_{n_{k-1}}$. We define n_k by requiring $x_{n_k} > x_{n_{k-1}}$ and $x_{n_k} \in B_{k^{-1}}(x)$. This can always be done. By construction the subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ converges to x.

2. Now suppose that (S, d) is sequentially compact. We first claim that if $\mathscr{U}$ is any open cover of S, then there exists $\delta > 0$ with the property that for each $x \in S$, there exists $U \in \mathscr{U}$ with $B_\delta(x) \subset U$. Note that while U will depend on x, δ does not.

   Suppose not. Then for each n, there exists $x_n \in S$ such that $B_{\frac{1}{n}}(x_n)$ is not contained in any element of $\mathscr{U}$. By the assumption of sequential compactness, we can choose a subsequence $x_{n_i} \to a$ for some $a \in S$. Now, since $\mathscr{U}$ is a open cover, there exists $U \in \mathscr{U}$ with $a \in U$. Since U is open, there exists $\epsilon > 0$ such that $B_\epsilon(a) \subset U$. Now pick i sufficiently large that $n_i^{-1} < \epsilon/2$ and $x_{n_i} \in B_{\epsilon/2}(a)$. Then we have $B_{1/n_i}(x_{n_i}) \subset B_\epsilon(a) \subset U$, a contradiction.

3. Next, we show that if (S, d) is sequentially compact, then for each $\epsilon > 0$ there exists a finite covering of S by balls of radius ϵ. Suppose not, then S cannot be covered by finitely many balls of radius ϵ. Construct a sequence as follows: take $x_1 \in S$ to be arbitrary. Given $x_1, \ldots x_n$, choose

$$x_{n+1} \in (B_\epsilon(x_1) \cup \ldots \cup B_\epsilon(x_n))^c$$

   which is always possible. Now, by construction (x_n) has no convergent subsequence, since $B_{\epsilon/2}(x)$ contains at most one element of (x_n), and we have a contradiction with the assumption of sequential compactness.

4. Finally we are ready to show that if (S, d) is sequentially compact, then it is compact. Let $\mathscr{U}$ be an open cover of S. Then by 2. above, there exists $\delta > 0$ such that for any $x \in S$, $B_\delta(x)$ is contained in an element of $\mathscr{U}$. By 3. we know that we can choose $x_1, \ldots x_N$ such that the sets $B_\delta(x_i)$ for $i = 1, \ldots, N$ cover S. Let $U_i \in \mathscr{U}$ be such that $B_\delta(x_i) \subset U_i$. Then we must have

$$S = \bigcup_{i=1}^{N} B_\delta(x_i) \subset \bigcup_{i=1}^{N} U_i,$$

   so by construction, $\{U_i\}_{i=1}^N$ is a finite subcover of $\mathscr{U}$. □

Exercise A.6. Let us take $X = \mathbb{R}^n$, thought of as a vector space over $\mathbb{R}$ and define:

$$\|(x_1, \ldots, x_n)\|_p = (|x_1|^p + \ldots |x_n|^p)^{\frac{1}{p}}, \qquad p \geq 1.$$

a) Show that $(\mathbb{R}^n, \|\cdot\|_p)$ is a normed vector space:

   i) First check that the positivity and homogeneity property are satisfied.

   ii) Establish the triangle inequality for the special case $p = 1$.

   iii) Next prove Young's inequality: if $a, b \in \mathbb{R}_+$ and $p, q > 1$ with $p^{-1} + q^{-1} = 1$ then:

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}$$

   Hint: set $t = p^{-1}$, *consider the function* $\log[ta^p + (1-t)b^q]$ *and use the concavity of the logarithm*

iv) With $p, q > 1$ such that $p^{-1} + q^{-1} = 1$, show that if $\|x\|_p = 1$ and $\|y\|_q = 1$ then

$$\sum_{i=1}^{n} |x_i y_i| \leqslant 1.$$

Deduce Hölder's inequality:

$$\sum_{i=1}^{n} |x_i y_i| \leqslant \|x\|_p \|y\|_q, \qquad \text{for all } x, y, \in \mathbb{R}^n.$$

v) Show that

$$\|x + y\|_p^p \leqslant \sum_{i=1}^{n} |x_i| |x_i + y_i|^{p-1} + \sum_{i=1}^{n} |y_i| |x_i + y_i|^{p-1}$$

vi) Apply Hölder's inequality to deduce:

$$\|x + y\|_p^p \leqslant \left(\|x\|_p + \|y\|_p\right) \|x + y\|_p^{p-1}$$

and conclude

$$\|x + y\|_p \leqslant \|x\|_p + \|y\|_p.$$

b) Show that the metric topology of $(\mathbb{R}^n, \|\cdot\|_p)$ agrees with the standard topology. *Hint: Use part c) of Exercise A.2*

Exercise A.7 ($\star$)**.** Let $X = C[0, 1]$, the set of continuous functions on the closed interval $[0, 1]$. For $f \in X$, $p \geqslant 0$ define:

$$\|f\|_p = \left(\int_0^1 |f(x)|^p \, dx\right)^{\frac{1}{p}}$$

a) Show that X is a vector space over $\mathbb{R}$, where scalar multiplication and vector addition are defined pointwise.

b) Establish Hölder's inequality:

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q$$

for $p, q > 1$ with $p^{-1} + q^{-1} = 1$.

c) Show that $(X, \|\cdot\|_p)$ is a normed space.

d) Suppose $p \leqslant p'$. Show that:

$$\|f\|_p \leqslant \|f\|_{p'}$$

e) Let τ_p be the metric topology of $(X, \|\cdot\|_p)$. Show that if $p \leqslant p'$:

$$\tau_p \subset \tau_{p'}.$$

f) Consider the sequence of functions:

$$f_n(x) = \begin{cases} n^{\gamma-1} & 0 \leqslant x < \frac{1}{n} \\ \frac{1}{n}x^{-\gamma} & \frac{1}{n} \leqslant x \leqslant 1 \end{cases}$$

where $n = 1, 2, \ldots$

i) Show that $f_n \in C[0,1]$ and

$$\lim_{n\to\infty} \|f_n\|_p = \begin{cases} 0 & \gamma < \frac{p+1}{p} \\ \left(\frac{p+1}{p}\right)^{\frac{1}{p}} & \gamma = \frac{p+1}{p} \\ \infty & \gamma > \frac{p+1}{p} \end{cases}$$

ii) By choosing γ carefully, show that if $p < p'$ then

$$\tau_{p'} \not\subset \tau_p.$$

Hint: in parts b), c) follow the same steps as for the finite dimensional case in Exercise A.6.

Exercise A.8. Verify that if (S, d) is a metric space, then the metric topology defines the same notions of convergence and continuity as the standard definitions for a metric space.

A.1.3 Topological vector spaces

Having briefly introduced the concept of vector spaces and topological spaces, we are now ready to define a topological vector space.

Definition A.7 (Topological vector space axioms)**.** *A vector space X over a field Φ, where Φ is either $\mathbb{R}$ or $\mathbb{C}$ is called a* topological vector space *if X is endowed with a topology τ such that:*

i) Every point of X is a closed set,

ii) The vector space operations are continuous with respect to τ.

To put a bit of flesh on the bones of this definition, the first condition implies that for any $x \in X$, the set $\{x\}$ is closed, or equivalently $X \setminus \{x\}$ should be open. The second condition should be understood as follows. We require firstly that the map:

$$\begin{aligned} + \quad : \quad & X \times X \to X \\ & (x, y) \mapsto x + y \end{aligned}$$

is continuous, where $X \times X$ inherits the product topology from X. Secondly, we require that the map:

$$\begin{aligned} \cdot \ : \ \ & \Phi \times X \to X \\ & (a, x) \mapsto ax \end{aligned}$$

Where $\Phi \times X$ is endowed with the product topology, and we take the topology on Φ to be the standard topology on $\mathbb{R}$ or $\mathbb{C} \simeq \mathbb{R}^2$ as appropriate.

We say that a subset, E, of a topological vector space is *bounded* if for every neighbourhood V of 0 we can find $s > 0$ such that $E \subset tV$ whenever $t > s$.

A useful source of topological vector spaces are the normed spaces that we previously introduced. We can verify that these indeed satisfy the topological vector space axioms:

Theorem A.8. *If $(X, \|\cdot\|)$ is normed vector space, endowed with the metric topology, then X is a topological vector space. A set $E \subset X$ is bounded if and only if $\sup_{x \in E} \|x\| < \infty$.*

Proof.

1. First we note that for each $x \in X$, the set $\{x\}$ is closed. To see this, suppose that $y \neq x$, and set $r = \frac{1}{2}\|y - x\|$. Then the open ball $B_r(y) = \{z \in X : \|z - y\| < r\}$ does not contain x, thus we have shown that $X \setminus \{x\}$ is open.

2. Now suppose U is an open set in X and let $x, y \in X$ be such that $z = x + y \in U$. By the openness of U in the norm topology, there exists $r > 0$ such that the set $B_r(z) \subset U$. Let $W = B_{\frac{r}{2}}(x) \times B_{\frac{r}{2}}(y)$, and suppose $(x', y') \in W$. Clearly

$$\left\|x' + y' - z\right\| = \left\|x' - x + y' - y\right\| \leqslant \left\|x' - x\right\| + \left\|y' - y\right\| < r$$

so that $x' + y' \in B_r(z) \subset U$. Thus the set $W \subset (+)^{-1}(U) \subset X \times X$. However, W is open in $X \times X$ by the definition of the product topology. Since $(x, y) \in (+)^{-1}(U)$ was arbitrary, we deduce that $(+)^{-1}(U)$ is open and so $+ : X \times X \to X$ is continuous.

3. Finally, suppose that U is an open set in X and let $x \in X$, $a \in \Phi$ be such that $z = ax \in U$. By the openness of U in the norm topology, there exists $r > 0$ such that the set $B_r(z) \subset U$. Let $W = \{b \in \Phi : |a - b| < r_1\} \times B_{r_2}(x)$, and suppose $(a', x') \in W$. Then we have:

$$\begin{aligned} \left\|a'x' - z\right\| &= \left\|a'x' - ax\right\| = \left\|(a' - a)x' - a(x - x')\right\| \\ &\leqslant \left\|(a' - a)x'\right\| + \left\|a(x - x')\right\| \\ &< r_1\left(\|x\| + r_2\right) + |a|\, r_2 \end{aligned}$$

Setting $r_1 = r_2 = \frac{\min\{r,1\}}{4(1 + \|x\| + |a|)}$, we have

$$r_1\left(\|x\| + r_2\right) + |a|\, r_2 \leqslant \frac{3r}{4},$$

and so $a'x' \in B_r(z) \subset U$. Thus the set $W \subset (\cdot)^{-1}(U) \subset \Phi \times X$. However, W is open in $\Phi \times X$ by the definition of the product topology. Since $(a, x) \in (\cdot)^{-1}(U)$ was arbitrary, we deduce that $(\cdot)^{-1}(U)$ is open and so $\cdot : \Phi \times X \to X$ is continuous.

4. Now suppose $E \subset X$ is bounded. Then in particular, since $B_1(0)$ is an open neighbourhood of 0, we have that $E \subset tB_1(0)$ for some $t > 0$. However, $tB_1(0) = B_t(0) = \{x \in X : \|x\| < t\}$, so necessarily we must have $\sup_{x \in E} \|x\| < t < \infty$. Conversely, suppose that $\sup_{x \in E} \|x\| = M < \infty$, and let V be any neighbourhood of 0. Since V is open in the metric topology, there exists $\epsilon > 0$ such that $B_\epsilon(0) \subset V$. Let $t = 2M\epsilon^{-1}$. We have $tB_\epsilon(0) = B_{2M}(0) = \{x \in X : \|x\| < 2M\}$, thus $E \subset tB_\epsilon(0) \subset tV$. □

Remark. *One has to be careful with various notions of boundedness for sets. While for a normed space the notion of boundedness introduced for topological vector spaces above is equivalent to the set having finite diameter, this is not true for general metric spaces. See the remark after Theorem A.18.*

We now prove a simple but useful consequence of the topological vector space definition.

Lemma A.9. *Let X be a topological vector space. For any $a \in X$ and $\lambda \in \Phi$ with $\lambda \neq 0$, define the maps:*

$$T_a \;:\; X \to X, \quad x \mapsto x + a. \qquad\qquad M_\lambda \;:\; X \to X, \quad x \mapsto \lambda x.$$

These are homeomorphisms of X to itself.

Proof. These maps are manifestly bijective, with inverses given by $(T_a)^{-1} = T_{-a}$ and $(M_\lambda)^{-1} = M_{\lambda^{-1}}$. All four maps are continuous by the definition of the topological vector space, since joint continuity implies separate continuity (see Exercise A.3). □

Lemma A.9 tells us that a set $E \subset X$ is open if and only if all of the translates $a + E$ are open. In particular this means that the topology of a topological vector space is determined by a *local* base at the origin.

Theorem A.10. *Suppose that (X, τ) is a topological vector space, and that $\dot\beta$ is a local base at 0. Then the collection*

$$\beta = \left\{ a + B : a \in X, B \in \dot\beta \right\}$$

is a base for τ.

Proof. Recall that a collection of open sets $\dot\beta$ is a local base at the origin if every neighbourhood of the origin contains a member of $\dot\beta$. First note that β is a collection of open sets, since translations of open sets are open. Now suppose that $U \in \tau$ is an open set and pick $x \in U$. We have that $(-x) + U$ is a neighbourhood of the origin, and so there exists $B \in \dot\beta$ such that $B \subset (-x) + U$. Since translation of sets preserves inclusions, we have $x + B \subset x + (-x) + U = U$. Thus for any $U \in \tau$ we have exhibited an element of β contained in U, so β is indeed a base for τ. □

Theorem A.11. *Suppose that X is a topological vector space. Then:*

a) If $U \subset X$ is a neighbourhood of 0 then U contains a balanced neighbourhood of 0.

b) If $U \subset X$ is a convex neighbourhood of 0, then U contains a convex balanced neighbourhood of 0.

Proof. 1. Since scalar multiplication is continuous, there exists $\delta > 0$ and V open such that $\alpha V \subset U$ for all $|\alpha| < \delta$. Let

$$W = \bigcup_{|\alpha|<\delta} \alpha V.$$

Then W is balanced and open, and $U' \subset U$, establishing a).

2. Now, suppose that U is convex, set

$$A = \bigcap_{|\alpha|=1} \alpha U$$

and choose W as in the previous paragraph. Since W is balanced, $\alpha^{-1}W = W$ whenever $|\alpha| = 1$, so $W \subset \alpha U$ for all $|\alpha| = 1$ and thus $W \subset A$. Thus A° is a neighbourhood of the origin. Clearly $A^\circ \subset U$. Since U is convex, so is αU for any α and thus A is an intersection of convex sets hence convex. The interior of a convex set is convex, thus A° is convex. Next I claim that A is balanced. Suppose $0 \leqslant r \leqslant 1$ and $|\beta| = 1$. To show A is balanced, it suffices to show that $r\beta A \subset A$. Note

$$r\beta A = \bigcap_{|\alpha|=1} r\beta\alpha U = \bigcap_{|\alpha|=1} r\alpha U.$$

However, since αU is convex and contains 0, we have $r\alpha U \subset \alpha U$, and it follows that A is balanced. It follows that A° is balanced, convex, open, contains 0 and is a subset of U.

□

Lemma A.12. *Suppose (X, τ) is a topological vector space. Then:*

a) τ is Hausdorff.

b) The set $\{x\}$ is bounded for any $x \in X$.

c) If E_1, E_2 are bounded, then so is $E_1 + E_2$. In particular, $x + E_1$ is bounded for any $x \in X$.

d) If $(x_n)_{n=1}^\infty$ is a sequence in X such that $\{x_n\}_{n=1}^\infty$ is bounded and $(a_n)_{n=1}^\infty$ is a sequence of scalars with $a_n \to 0$, then $a_n x_n \to 0$.

Proof. 1. We first show that every neighbourhood, W, of 0 contains a balanced open set U satisfying $U + U \subset W$. To see this, note that $0 + 0 = 0$, so by the continuity of 0, there exist neighbourhoods U_1, U_2 of 0 such that $U_1 + U_2 \subset W$. We let

$$U' = U_1 \cap U_2$$

which satisfies $U' + U' \subset W$. By Theorem A.11, U' has a balanced subset U, and $U + U \subset U' + U' \subset W$.

2. Now consider $x, y \in X$ with $x \neq y$. Since $\{y\}$ is closed and $x \in \{y\}^c$, there exists W, a neighbourhood of x with $y \notin W$. Since $-x + W$ is a neighbourhood of 0, there exists a balanced U with $U + U \subset -x + W$. Thus $x + U + U \subset W$ and in particular $y \notin x + U + U$. I claim $(x + U) \cap (y + U) = \emptyset$. Suppose not, then there exists $a, b \in U$ such that $x + a = y + b$, which implies $y = x + a - b$. But $a, -b \in U$, so $y \in x + U + U$, a contradiction. We have constructed sets $x + U$ and $y + U$ which are open, and contain x, y respectively, thus X is Hausdorff and we have established a).

3. Fix $x \in X$ and consider the map $f_x : \mathbb{R} \to X$ given by $f_x(\lambda) = \lambda x$. This is a continuous map, so ${f_x}^{-1}(W)$ is open in $\mathbb{R}$. Since $0 \in W$, we have $0 \in {f_x}^{-1}(W)$, and thus from the definition of an open set in $\mathbb{R}$, the interval $(-\epsilon, \epsilon) \in {f_x}^{-1}(W)$ for some $\epsilon > 0$. Thus $\lambda x \in W$ for $\lambda \in (0, \epsilon)$, or equivalently $x \in tW$ for $t > \epsilon^{-1}$. Thus we have established b).

4. Let W be any neighbourhood of 0. By paragraph 1. there exists U a neighbourhood or 0 such that $U + U \subset W$. Since E_1, E_2 are both bounded, there exists $s \in \mathbb{R}$ such that $t^{-1}E_i \subset U$ for $t > s$ and $i = 1, 2$. Thus for $t > s$,
$$t^{-1}(E_1 + E_2) = t^{-1}E_1 + t^{-1}E_2 \subset U + U \subset W,$$
or equivalently $E_1 + E_2 \subset tW$ and hence $E_1 + E_2$ is bounded, which is the first part of c). The final part of c) follows by applying the result from b).

5. For part d), suppose W is any neighbourhood of the origin in X. As in part 1., we can take U balanced and open with $U \subset W$. Then there exists $s > 0$ such that $x_n \in tU$ for all $n = 1, 2, \ldots$ and any $t > s$. Since $a_n \to 0$, there exists N such that $|a_n| < s^{-1}$ for all $n \geqslant N$. Since U is balanced, and we have $x_n \in tU$ and $|ta_n| < 1$ for $n \geqslant N$, we deduce that $a_n x_n \in U \subset W$ for all $n \geqslant N$ and we're done. □

Suppose that X is a vector space equipped with a metric d. We say that d is *invariant* if
$$d(x + z, y + z) = d(x, y),$$
for all $x, y, z \in X$. We have the following useful result

Lemma A.13. *Suppose that X is a vector space, equipped with an invariant norm d, and let τ be the induced metric topology. Given a sequence $(x_n)_{n=1}^{\infty}$ with $x_n \to 0$, there exist scalars $\alpha_n \to \infty$ such that $\alpha_n x_n \to 0$.*

Proof. 1. First note that if d is invariant, then
$$d(nx, 0) \leqslant nd(x, 0).$$
This is clearly true if $n = 1$. Suppose it holds for $n = 1, \ldots k - 1$ Then
$$\begin{aligned} d(kx, 0) &\leqslant d(kx, x) + d(x, 0) \\ &= d((k-1)x, 0) + d(x, 0) \\ &\leqslant (k-1)d(x, 0) + d(x, 0) = kd(x, 0) \end{aligned}$$
and we're done by induction.

2. Now note that since $x_n \to 0$, for any $m \in \mathbb{N}$ there exists N_m such that

$$d(x_n, 0) < \frac{1}{m^2} \qquad n \geqslant N_m$$

where we can assume that $N_m < N_{m+1}$. We define $\alpha_n = m$ for $N_m \leqslant n < N_{m+1}$. Suppose that $N_m \leqslant n < N_{m+1}$. Then

$$d(a_n x_n, 0) \leqslant m d(x_n, 0) < \frac{1}{m}.$$

Thus as $n \to \infty$, we have that $a_n x_n \to 0$, however $a_n \to \infty$.

□

A crucially important concept which you may have come across when studying metric spaces is the idea of a Cauchy sequence.

Definition A.8. *i) Suppose (S, d) is a metric space. We say that a sequence $(x_n)_{n=1}^{\infty}$ is d-Cauchy if for every $\epsilon > 0$ we can find an integer N such that*

$$d(x_n, x_m) < \epsilon, \qquad \textit{for all } n, m \geqslant N.$$

A metric space is called complete *if every d-Cauchy sequence converges in S.*

ii) Suppose (X, τ) is a topological vector space. We say that a sequence $(x_n)_{n=1}^{\infty}$ is τ-Cauchy if for every neighbourhood, U, of the origin we can find an integer N such that

$$x_n - x_m \in U, \qquad \textit{for all } n, m \geqslant N.$$

Exercise A.9. Let (X, τ) be a topological vector space

a) Show that if $(x_n)_{n=1}^{\infty}$ is a τ-Cauchy sequence, then $\{x_n\}_{n=1}^{\infty}$ is bounded.

b) Fix a local base $\dot{\beta}$. Show that a sequence $(x_n)_{n=1}^{\infty}$ is τ-Cauchy if and only if for any $B \in \dot{\beta}$ we can find an integer N such that

$$x_n - x_m \in B, \qquad \text{for all } n, m \geqslant N.$$

Lemma A.14. *Suppose that X is a vector space, equipped with an invariant norm d, and let τ be the induced metric topology. Then a sequence $(x_n)_{n=1}^{\infty}$ is d-Cauchy if and only if it is τ-Cauchy.*

Proof. Suppose that (x_n) is τ-Cauchy. Then for any $\epsilon > 0$, there exists N such that for all $n, m > N$ we have $x_n - x_m \in B_\epsilon(0)$, i.e.

$$\epsilon > d(0, x_n - x_m) = d(x_n, x_m),$$

thus (x_n) is d-Cauchy.

Now suppose (x_n) is d-Cauchy. Let V be any neighbourhood of 0. Since V is open, there exists $\epsilon > 0$ such that $B_\epsilon(0) \subset V$. Since (x_n) is d-Cauchy, there exists N such that $d(x_n, x_m) < \epsilon$ for all $n, m > N$. Thus

$$\epsilon > d(x_n, x_m) = d(0, x_n - x_m),$$

so $x_n - x_m \in B_\epsilon(0) \subset V$ and (x_n) is τ-Cauchy. □

We're now in a position to distinguish various useful classes of topological vector space. There is some difference of opinion on the definitions below, we follow here the conventions of Rudin. Here (X, τ) always refers to a topological vector space:

i) X is a *locally convex* topological vector space if there is a local base $\dot{\beta}$ whose members are convex.

ii) X is *locally bounded* if 0 has a bounded neighbourhood.

iii) X is *locally compact* if 0 has a neighbourhood whose closure is compact.

iv) X is *metrizable* if there exists some metric d on X such that τ is the metric topology induced by d.

v) X is an *F-space* if its topology τ is induced by a complete invariant metric.

vi) X is a *Fréchet space* if it is a locally convex F-space.

vii) X is *normable* if a norm exists on X such that the metric topology of the norm agrees with τ.

viii) A normed space $(X, \|\cdot\|)$, with the metric topology, is *Banach* if the metric induced by the norm is complete.

ix) A space X has the *Heine-Borel property* if every closed and bounded subset of X is compact.

The space $\mathbb{R}^n$ with the norm $\|\cdot\|_p$ introduced in the exercises is an example of a topological vector space which belongs to all of these classes. The spaces that you studied in Functional Analysis were mostly Banach spaces, although not all. For example if X is an infinite dimensional Banach space, then the weak-* topology of X^* is locally convex, but not metrisable.

We note that the converse of the Heine-Borel property is always true for a topological vector space:

Lemma A.15. *Suppose (X, τ) is a topological vector space and that $K \subset X$ is compact. Then K is closed and bounded.*

Proof. 1. The fact that K is closed follows immediately from Lemmas A.3, A.12.

2. Next, suppose U is a neighbourhood of 0. By the continuity of scalar multiplication, there exists $\delta > 0$ and a neighbourhood V of the origin in X such that $\alpha V \subset U$ for any $|\alpha| < \delta$. Define W to be the union of these sets as α varies over $\{|\alpha| < \delta\}$. Then $W \subset V$ is an open, balanced, neighbourhood of 0.

3. Now I claim that
$$\bigcup_{n=1}^{\infty} nW = X.$$

To see this, fix $x \in X$. Since the map $\alpha \mapsto \alpha x$ is continuous, the set of all α with $\alpha x \in W$ is open and contains 0, hence contains n^{-1} for sufficiently large n. Thus $n^{-1}x \in W$, or $x \in nW$ for large enough n. Note that since W is balanced, in particular $sW \subset tW$ for $s < t$.

4. Finally, since $\mathscr{U} = \{nW\}_{n=1}^{\infty}$ is an open cover of X, it is also an open cover of K. Thus there exist $n_1, \ldots, n_N$ such that

$$K \subset \bigcup_{i=1}^{N} n_i W = n_N W \subset n_N U.$$

Thus K is bounded.

□

A.2 Locally convex spaces

We shall now specialise somewhat, to the case of *locally convex* topological vector spaces. These can be given a nice description in terms of a family of semi-norms. When that family is countable, the topology is equivalent to that induced by an invariant metric, which if it is complete gives a Fréchet space. These can be thought of as generalisations of the Banach spaces that you may be familiar with from functional analysis. The canonical example of a Fréchet space that is not a Banach space is $C^{\infty}(\Omega)$, the space of smooth functions on an open set Ω. The finite regularity spaces on a compact set $C^k(K)$ are Banach spaces in a natural way, but this is not true of $C^{\infty}(\Omega)$.

A.2.1 Semi-norms

A very useful way to construct the topology for a locally convex topological vector space is via a family of semi-norms.

Definition A.9. *A* seminorm *on a vector space X over Φ (with Φ being either $\mathbb{R}$ or $\mathbb{C}$) is a map $p : X \to \mathbb{R}$ satisfying:*

i) p is subadditive. *For all $x, y \in X$ we have:*

$$p(x + y) \leqslant p(x) + p(y)$$

ii) For all $\lambda \in \Phi$ and $x \in X$ we have:

$$p(\lambda x) = |\lambda|\, p(x)$$

A family of seminorms $\mathscr{P}$ is said to be separating *if for every $x \in X$ with $x \neq 0$, there is at least one $p \in \mathscr{P}$ with $p(x) \neq 0$.*

From the definition we can immediately deduce some useful properties:

Lemma A.16. *Let X be a vector field over $\mathbb{R}$ or $\mathbb{C}$, and let $p : X \to \mathbb{R}$ be a seminorm. Then:*

a) $p(0) = 0$

b) $|p(x) - p(y)| \leqslant p(x - y)$

c) $p(x) \geqslant 0$

d) $\{x : p(x) = 0\}$ *is a vector subspace of* X*.*

e) The set $B = \{x : p(x) < 1\}$ *is convex and balanced.*

Proof. a) Applying property *ii)* from the definition of a seminorm with $\lambda = 0$ we immediately have $p(0) = 0$.

b) From the subadditivity property we have

$$p(x) = p(x - y + y) \leqslant p(x - y) + p(y)$$

so $p(x) - p(y) \leqslant p(x - y)$. Similarly $p(y) - p(x) \leqslant p(y - x)$, but $p(x - y) = p(y - x)$ and the result follows.

c) Applying $a), b)$ with $y = 0$ gives $|p(x)| \leqslant p(x)$ which implies $p(x) \geqslant 0$.

d) Suppose $p(x) = p(y) = 0$ and $\lambda, \mu \in \Phi$. Applying $c)$ we have:

$$0 \leqslant p(\lambda x + \mu y) \leqslant |\lambda|\, p(x) + |\mu|\, p(y) = 0,$$

so that $p(\lambda x + \mu y) = 0$ and thus $\{x : p(x) = 0\}$ is a vector subspace.

e) It is clear that B is balanced by property *ii)*. To see that B is convex, suppose that $x, y \in B$ and $0 < t < 1$. Then

$$p(tx + (1 - t)y) \leqslant tp(x) + (1 - t)p(y) < 1,$$

so $tx + (1 - t)y \in B$ and B is convex. □

Note that these results are already enough to show that a seminorm p with the property that $p(x) \neq 0$ whenever $x \neq 0$, is in fact a norm.

We are now ready to prove an important result that shows that a family of seminorms specifies a locally convex topology on a vector space. The proof is quite long, and you may wish to omit it on a first read through. The argument is similar to the proof of Theorem A.8, which in fact could be understood as a corollary of this result.

Theorem A.17. *Suppose that* $\mathscr{P}$ *is a separating family of seminorms on a vector space* X*. Associate to each* $p \in \mathscr{P}$ *and* $n \in \mathbb{N}$ *the set:*

$$V(p, n) = \left\{ x \in X : p(x) < \frac{1}{n} \right\}.$$

Let $\dot{\beta}$ *be the collection of all finite intersections of the sets* $V(p, n)$*. Then* $\dot{\beta}$ *is a convex, balanced, local base for a topology* τ *on* X*, which turns* X *into a locally convex topological vector space such that:*

a) every $p \in \mathscr{P}$ *is continuous*

b) a set $E \subset X$ *is bounded if and only if every* $p \in \mathscr{P}$ *is bounded on* E.

Proof. 1. Let us define β by

$$\beta = \left\{ x + B : x \in X, B \in \dot{\beta} \right\}.$$

Since $0 \in B$ for any $B \in \dot{\beta}$, we immediately have that for any $x \in X$ there is an element of β containing x. Now suppose that $B_1, B_2 \in \beta$. We may write

$$B_1 = y + \bigcap_{i=1}^{N} V(p_i, n_i), \qquad B_2 = z + \bigcap_{j=1}^{M} V(q_i, m_j),$$

for $y, z \in X$, $p_i, q_j \in \mathscr{P}$ and $N, M, n_i, m_j \in \mathbb{N}$. Fix $x \in B_1 \cap B_2$. Clearly $x \in B$ and $B \in \beta$. I claim that $B \subset B_1 \cap B_2$ for

$$B = x + \left(\bigcap_{i=1}^{N} V(p_i, n_i') \right) \cap \left(\bigcap_{j=1}^{M} V(q_j, m_j') \right),$$

provided that n_i', m_i' are chosen sufficiently large. Since $x \in B_1 \cap B_2$, we have:

$$p_i(x-y) < \frac{1}{n_i'}, \quad q_i(x-z) < \frac{1}{m_j'}, \qquad \text{for all } i = 1, \dots, N, \quad j = 1, \dots, M$$

For each i, j, pick n_i', m_j' sufficiently large that

$$p_i(x-y) + \frac{1}{n_i'} < \frac{1}{n_i}, \quad q_j(x-z) + \frac{1}{m_j'} < \frac{1}{m_j}$$

Now suppose $w \in B$. Then we have that:

$$p_i(w-x) < \frac{1}{n_i'}, \quad q_i(w-x) < \frac{1}{m_j'}, \qquad \text{for all } i = 1, \dots, N, \quad j = 1, \dots, M$$

Using the subadditivity of p_i we have that for each i:

$$p_i(w-y) \leqslant p_i(w-x) + p_i(x-y) < \frac{1}{n_i'} + p_i(x-y) < \frac{1}{n_i},$$

thus $w \in B_1$. Similarly, we have for each j that:

$$q_j(w-z) \leqslant q_j(w-x) + q_j(x-z) < \frac{1}{m_j'} + q_j(x-y) < \frac{1}{m_i},$$

Thus the collection β satisfies the conditions of Exercise A.2 and thus defines a topology τ on X. Moreover, $\dot{\beta}$ is a local base for τ and each element of $\dot{\beta}$ is convex and balanced.

2. Suppose that $x, y \in X$ with $x \neq y$. Then since $x - y \neq 0$ and $\mathscr{P}$ is separating, there exists $p \in \mathscr{P}$ such that $p(x-y) > 0$. Thus, there exists $n \in \mathbb{N}$ such that $np(x-y) > 1$. For this n we have that $x \notin (y + V(p,n))$. Thus we may write $\{x\}^c$ as a union of sets which are open in τ, hence $\{x\}$ is closed.

3. Next we must show that addition is continuous. Suppose U is an open set in X, and pick $z \in U$. Then
$$\bigcap_{i=1}^{N} V(p_i, n_i) \subset -z + U$$
for some $p_i \in \mathscr{P}$, $n_i \in \mathbb{N}$. Suppose that $(x, y) \in (+)^{-1}(z)$, i.e. $x + y = z$. Let
$$V_1 = x + \bigcap_{i=1}^{N} V(p_i, 2n_i), \qquad V_2 = y + \bigcap_{i=1}^{N} V(p_i, 2n_i)$$
and suppose $(w_1, w_2) \in V_1 \times V_2$. Then for all i we have
$$p_i(w_1 + w_2 - z) = p_i(w_1 - x + w_2 - y) \leqslant p_i(w_1 - x) + p_i(w_2 - y) < \frac{1}{2n_i} + \frac{1}{2n_i} < \frac{1}{n_i}$$
so that $V_1 + V_2 \subset U$, or alternatively $V_1 \times V_2 \subset (+)^{-1}(U)$. Thus we can write $(+)^{-1}(U)$ as a union of sets which are open in the product topology. This proves that addition is continuous.

4. Next we must show that scalar multiplication is continuous. Suppose U is an open set in X, and pick $z \in U$. Then
$$\bigcap_{i=1}^{N} V(p_i, n_i) \subset -z + U$$
for some $p_i \in \mathscr{P}$, $n_i \in \mathbb{N}$. Suppose that $(\alpha, x) \in (\cdot)^{-1}(z)$, i.e. $\alpha y = z$. Let
$$V = x + \bigcap_{i=1}^{N} V(p_i, n_i'), \qquad D = \{\beta \in \Phi : |\alpha - \beta| < \epsilon\}$$
Suppose $(\beta, y) \in D \times V$ Then for each i we have:
$$\begin{aligned} p_i(\beta y - \alpha x) &= p_i(\beta(y - x) - (\alpha - \beta)x) \\ &< \frac{|\beta|}{n_i'} + \epsilon p_i(x) \\ &\leqslant \frac{|\alpha| + \epsilon}{n_i'} + \epsilon p_i(x). \end{aligned}$$
Taking $\epsilon < (2n_i p_i(x))^{-1}$ and $n_i' > 2(|\alpha| + (2n_i p_i(x))^{-1})$ for each i, we conclude that
$$(\beta y - z) \in \bigcap_{i=1}^{N} V(p_i, n_i),$$
which implies that $D \times V \in (\cdot)^{-1}(U)$. Thus scalar multiplication is continuous, and we have established that (X, τ) is a locally convex topological space.

5. To see that $p \in \mathscr{P}$ is continuous, we must show that $p^{-1}(a, b)$ is open, where $a < b$. Suppose $x \in p^{-1}(a, b)$, so that $a < p(x) < b$. Consider

$$U = x + V(p, n)$$

Suppose $y \in V$. Then

$$|p(y) - p(x)| \leqslant p(y - x) < \frac{1}{n},$$

by part $c)$ of Lemma A.16. For n sufficiently large, we have $p(y) \in (a, b)$ so that $V \subset p^{-1}(a, b)$ and we're done.

6. It remains to show that $E \subset X$ is bounded if and only if every $p \in \mathscr{P}$ is bounded on E. First suppose E is bounded and fix $p \in \mathscr{P}$. Since $V(p, 1)$ is a neighbourhood of the origin, from the definition of boundedness we have that $E \subset kV(p, 1)$ for some $k < \infty$. But $x \in kV(p, 1)$ implies $p(x) < k$, so that p is bounded on E.

   Now suppose that every $p \in \mathscr{P}$ is bounded on E. Let U be a neighbourhood of the origin. Then

$$\bigcap_{i=1}^{N} V(p_i, n_i) \subset U$$

   for some $p_i \in \mathscr{P}$, $n_i \in \mathbb{N}$. By our assumption, there exist $M_i < \infty$ such that $p_i < M_i$ on E of $1 \leqslant i \leqslant N$. If $n > M_i n_i$ for all i, then $E \subset nU$, since if $p_i(x) < M_i$, we have

$$p_i\,(x) < M_i < \frac{n}{n_i} \qquad i = 1, \ldots, N$$

   so that

$$p_i\left(\frac{1}{n}x\right) < \frac{1}{n_i} \qquad i = 1, \ldots, N$$

   and $n^{-1}x \in U$. □

Thus we have seen that a separating family of seminorms gives rise to a locally convex topological space. In fact, the converse is true: given a locally convex topological space, we can find a (not necessarily unique) separating family of seminorms which generates the topology in the manner of the previous theorem.

In the case where the separating family of seminorms $\mathscr{P}$ is *countable*, we have an alternative means of describing the topology.

Theorem A.18. *Let*

$$\mathscr{P} = \{p_i\}_{i=1}^{\infty}$$

be a countable separating family of seminorms on a vector space X, and let τ be the topology induced by this family as described in Theorem A.17. Then the locally convex topological vector space (X, τ) is metrizable, *and the topology τ agrees with that induced by the invariant metric:*

$$d(x, y) = \sum_{i=1}^{\infty} 2^{-i} \frac{p_i(x - y)}{1 + p_i(x - y)}$$

Proof. 1. We first verify that d indeed defines an invariant metric. It is clearly symmetric since $p(-x) = p(x)$ for a seminorm. We note that the map

$$F : t \mapsto \frac{t}{1+t}$$

is smooth, monotone increasing, concave and takes $[0, \infty)$ to $[0, 1)$. Thus $d(x, y)$ is a sum of non-negative terms, so $d(x, y) \geqslant 0$. Equality occurs if and only if $p_i(x - y) = 0$ for all i, which by the fact that $\mathscr{P}$ is separating implies $x = y$. Next we claim that F is subadditive. To see this, we note that by the convexity of F, together with $F(0) = 0$ we have for $t \geqslant 0$ and $0 < \lambda < 1$:

$$F(\lambda t) = F(\lambda t + (1 - \lambda)0) \geqslant \lambda F(t) + (1 - \lambda)F(0) = \lambda F(t).$$

Then for $t, s \geqslant 0$:

$$\begin{aligned} F(t) + F(s) &= F\left((t+s)\frac{t}{t+s}\right) + F\left((t+s)\frac{s}{t+s}\right) \\ &\geqslant \frac{t}{t+s}F(t+s) + \frac{s}{t+s}F(t+s) = F(t+s). \end{aligned}$$

Now, since p is a seminorm we have

$$\begin{aligned} & p(x-y) \leqslant p(x-z) + p(z-y) & \\ \Longrightarrow\quad & F[p(x-y)] \leqslant F\left[p(x-z) + p(z-y)\right] & \text{(Monotonicity of } F) \\ \Longrightarrow\quad & F[p(x-y)] \leqslant F\left[p(x-z)\right] + F[p(z-y)] & \text{(Subadditivity of } F) \end{aligned}$$

Thus we conclude

$$\begin{aligned} d(x, y) &= \sum_{i=1}^{\infty} 2^{-i} \frac{p_i(x-y)}{1 + p_i(x-y)} \\ &\leqslant \sum_{i=1}^{\infty} 2^{-i} \left(\frac{p_i(x-z)}{1 + p_i(x-z)} + \frac{p_i(z-y)}{1 + p_i(z-y)} \right) \\ &= d(x, z) + d(z, y), \end{aligned}$$

so d is indeed a metric on X. It is manifestly invariant.

2. Now we need to show that the topology induced by d, which we denote τ_d, agrees with the topology τ induced by the family of seminorms $\mathscr{P}$. Recall that the open sets of τ_d are precisely those sets which can be written as a union of the open balls $B_r(x) = \{y \in X : d(x, y) < r\}$.

3. From the definition of τ, we have that each p_i is τ-continuous on X. Since $|F(t)| < 1$, we conclude by the Weierstrass M-test that the sum in $d(x, y)$ converges uniformly. Hence d is continuous as a real valued function on $X \times X$ with the product topology coming from τ. In particular, the map $d_x : X \to \mathbb{R}$ given by $d_x(y) = d(x, y)$ is continuous, and thus $d_x^{-1}(-r, r) = B_r(x)$ is an open set in the τ topology. Thus $\tau_d \subset \tau$.

4. Now suppose that W is an open set of τ, and that $x \in W$. By the definition of τ, there exists B such that $x + B \subset W$ and B has the form

$$B = \bigcap_{k=1}^{N} V\left(p_{i_k}, n_k\right)$$

for some $p_{i_k} \in \mathscr{P}$ and $n_k \in \mathbb{N}$, where we recall $V(p,n) = \{y \in X : p(y) < n^{-1}\}$. Now suppose $d(x,y) < \epsilon 2^{-M}$. Then in particular, for $0 \leqslant i \leqslant M$ we have

$$\frac{p_i(x-y)}{1+p_i(x-y)} \leqslant \epsilon,$$

so that if $\epsilon < \frac{1}{2}$:

$$p_i(x-y) \leqslant \frac{\epsilon}{1-\epsilon} \leqslant 2\epsilon.$$

Thus if we take $M \geqslant i_k$ and $\epsilon < (n_k)^{-1}$ for all $k = 1, \ldots, N$ we deduce that if $y \in B_{\epsilon 2^{-M}}(x)$ then $y - x \in B$ and thus $y \in W$. Since x was arbitrary we can write W as a union of open balls for the metric d, and thus $\tau \subset \tau_d$. □

Remark. 1. *Note that while a countable separating family of seminorms gives rise to a metrizable locally convex topology, it need not be the case that the metric balls $B_r(0)$ are themselves convex. The sets $V(p,n)$ however are.*

2. *It is straightforward to see that $d(x,y) < 1$ for any $x, y \in X$, so that* any *subset of X has finite diameter. On the other hand, it does not follow that all subsets of X are bounded in the sense introduced above for a topological vector space.*

A.3 The test function spaces

A.3.1 $\mathscr{E}(\Omega)$ and $\mathscr{D}_K$

Let $\Omega \subset \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$. Recall that for a function $f : \Omega \to \mathbb{C}$, we say $f \in C^\infty(\Omega)$ if $D^\alpha f$ is a continuous function in Ω for all multiindices α. Clearly $C^\infty(\Omega)$ is a vector space over $\mathbb{C}$, with addition and scalar multiplication defined pointwise: if $f, g \in C^\infty(\Omega)$, $\lambda \in \mathbb{C}$, we define the maps $f + g$, λf by

$$f+g \;:\; \Omega \to \mathbb{C}, \quad x \mapsto f(x)+g(x), \qquad \lambda f \;:\; \Omega \to \mathbb{C}, \quad x \mapsto \lambda f(x).$$

Then $f + g, \lambda g \in C^\infty(\Omega)$.

We shall endow $C^\infty(\Omega)$ with a topology which makes it into a Fréchet space with the Heine-Borel property. By the exhaustion lemma, Lemma A.6, we can find a sequence of compact sets $(K_i)_{i=0}^\infty$ such that $K_i \subset \Omega$, $K_i \subset (K_{i+1})^\circ$ and $\bigcup_i K_i = \Omega$. We define a family of seminorms by:

$$p_n(f) = \max\left\{|D^\alpha f(x)| : x \in K_n, |\alpha| \leqslant n\right\}. \tag{A.1}$$

The family $\mathscr{P} = \{p_n : n \in \mathbb{N}\}$ is separating. If $f \neq 0$, then $f(x) \neq 0$ at some point $x \in \Omega$. For n sufficiently large, $x \in K_n$, and thus $p_n(f) > 0$. Thus the family of seminorms $\mathscr{P}$ induces a topology, τ, on $C^\infty(\Omega)$ which is locally convex and metrizable by Theorem A.18. When $C^\infty(\Omega)$ is endowed with the topology τ, we use the notation $\mathscr{E}(\Omega)$. A local base is given by the sets

$$V_N = \left\{ f \in C^\infty(\Omega) : p_N(f) < \frac{1}{N} \right\}, \qquad N = 1, 2, \ldots.$$

It's useful to categorise convergence in this space in terms of more familiar concepts as follows:

Lemma A.19. *A sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathscr{E}(\Omega)$ converges to f if and only if $D^\alpha f_n \to D^\alpha f$ uniformly on compact sets for each multiindex α.*

Proof. By the translation invariance of the topology, we can assume w.l.o.g. that $f = 0$. The sequence (f_n) tends to 0 in $\mathscr{E}(\Omega)$ if and only if for each N there exists m_N such that $f_n \in V_N$ for all $n \geqslant m_N$.

First suppose $f_n \to 0$ in $\mathscr{E}(\Omega)$. Fix α and let $K \subset \Omega$ be any compact subset. For any ϵ, there exists N such that $N > \max\{|\alpha|, \epsilon^{-1}\}$, $K \subset K_N$. If $n \geqslant m_N$ then $f_n \in V_N$, which implies

$$\sup_K |D^\alpha f_n| \leqslant \sup_{K_N} |D^\alpha f_n| \leqslant p_N(f_n) \leqslant \frac{1}{N} < \epsilon,$$

Thus $D^\alpha f_n \to 0$ uniformly on K.

Conversely, suppose that for each multiindex α and compact set K we have that $D^\alpha f_n \to 0$ uniformly on K. Fix N. Then for each α with $|\alpha| \leqslant N$ we have $D^\alpha f_n \to 0$ uniformly on K_N. In particular, for each α there exists m_α such that if $n \geqslant m_\alpha$, we have that

$$\sup_{K_N} |D^\alpha f_n| \leqslant \frac{1}{N}.$$

Thus if $m = \max_{|\alpha|\leqslant N} m_\alpha$ then for all $n > m$ we have $f_n \in V_N$ and thus $f_n \to 0$ in $\mathscr{E}$. □

Theorem A.20. *The topological vector space $\mathscr{E}(\Omega)$ is a Fréchet space with the Heine-Borel property.*

Proof.
1. Since we already have that $\mathscr{E}(\Omega)$ is locally convex and inherits its topology from an invariant metric, in order to show that $\mathscr{E}(\Omega)$ is Fréchet, we simply have to show completeness. A sequence $(f_n)_{n\in\mathbb{N}}$ with $f_n \in \mathscr{E}(\Omega)$ is Cauchy if for any fixed N, there exists M such that for all $i, j \geqslant M$ we have $f_i - f_j \in V_N$. Thus

$$\sup_{K_N} |D^\alpha f_i - D^\alpha f_j| < \frac{1}{N}, \qquad \text{for all } |\alpha| \leqslant N.$$

   Since K_N exhaust Ω, this implies that there exist continuous functions g^α such that $D^\alpha f_n \to g^\alpha$ uniformly on compact subsets of Ω. By a standard result, this implies that there exists a smooth function f such that $f_n \to f$ and $D^\alpha f_n \to D^\alpha f$ uniformly on compact subsets of Ω. Thus every Cauchy sequence has a limit and $\mathscr{E}(\Omega)$ is complete.

2. Now suppose that $E \subset \mathscr{E}(\Omega)$ is closed and bounded. We need to show that E is compact. By Theorem A.7 it suffices to show that any sequence in E has a convergent subsequence. By Theorem A.17, the boundedness of E is equivalent to the existence of M_N such that $p_N(f) < M_N$ for all $f \in E$.

3. In particular, we have that

$$|D^\alpha f| < M_N \qquad \text{for } |\alpha| = N$$

holds for all $f \in E$ on K_N. This in particular implies that for each β with $|\beta| < N-1$ the set $\{D^\beta f : f \in E\}$ is equicontinuous on K_{N-1}, and it is trivially pointwise bounded by the condition $p_N(f) < M_N$.

4. Suppose that $(f_n)_{n\in\mathbb{N}}$ is any sequence in E. By Arzelà-Ascoli we can extract a subsequence $(f_{n_k^1})_{k\in\mathbb{N}}$ such that $f_{n_k^1}$ converges uniformly on K_0. Suppose now that an increasing sequence of integers n_k^N are given with the property that $(D^\beta f_{n_k^N})_{k\in\mathbb{N}}$ converges uniformly on K_{N-1} for all $|\beta| \leqslant N-1$. Consider the sequence $(f_{n_k^N})_{k\in\mathbb{N}}$. Since this is a sequence in E, we know that for each β with $|\beta| \leqslant N$ the set $\{D^\beta f_{n_k^N} : f \in E, k \in \mathbb{N}\}$ is equicontinuous and pointwise bounded on K_N. Thus we can extract a subsequence $(f_{n_k^{N+1}})_{k\in\mathbb{N}}$ such that $(D^\beta f_{n_k^{N+1}})_{k\in\mathbb{N}}$ converges uniformly on K_N for all $|\beta| \leqslant N$. Thus by induction, we can find n_k^N with the required property for all N.

5. Consider the sequence $(F_k)_{k\in\mathbb{N}}$ with $F_k = f_{n_k^k}$ where n_k^N are as constructed above. Since n_k^{N+1} is a subsequence of n_k^N, we conclude that $D^\alpha F_k$ converges uniformly on compact subsets for any α, and thus for any sequence in E we have exhibited a convergent subsequence.

□

If $K \subset \mathbb{R}^n$ is a compact set, we denote by $\mathscr{D}_K$ the space of all $f \in C^\infty(\mathbb{R}^n)$ whose support lies in K. If $K \subset \Omega$, then $\mathscr{D}_K$ may be identified with a vector subspace of $C^\infty(\Omega)$. In fact, this subspace is closed with respect to the $\mathscr{E}(\Omega)$ topology. To see this, note that the map $\delta_x : \mathscr{E}(\Omega) \to \mathbb{C}$ given by $f \mapsto f(x)$ is continuous. Thus the set $\delta_x^{-1}(\{0\})$ is a closed set in $\mathscr{E}(\Omega)$. Since we can write:

$$\mathscr{D}_K = \bigcap_{x\in\Omega\setminus K} \delta_x^{-1}(\{0\})$$

and arbitrary intersections of closed sets are closed, we deduce that $\mathscr{D}_K$ is closed (and hence complete in the subspace topology). Thus $\mathscr{D}_K$ is itself a Fréchet space, when equipped with the subspace topology, which we denote τ_K.

A.3.2 $\mathscr{D}(\Omega)$

We have described the spaces $\mathscr{D}_K$, which consist of smooth functions whose support is restricted to a given compact set $K \subset \Omega$. The set $\mathscr{D}(\Omega)$ of test functions is the union over the sets $\mathscr{D}_K$ with $K \subset \Omega$ compact:

$$\mathscr{D}(\Omega) = \bigcup_{K \subset \Omega} \mathscr{D}_K.$$

In other words, $f \in \mathscr{D}(\Omega)$ if f is smooth, and is supported in some compact subset of Ω. It is clear that $\mathscr{D}(\Omega)$ is closed under the natural operations of addition and multiplication by a complex number, and thus $\mathscr{D}(\Omega)$ is a vector space over $\mathbb{C}$. We would like to endow $\mathscr{D}(\Omega)$ with a topology which turns it into a complete, locally convex, topological vector space, such that the subspace topology induced on $\mathscr{D}_K$ agrees with the natural Fréchet topology, τ_K introduced above for each compact $K \subset \Omega$.

One natural possibility is to consider the norms

$$\|f\|_k = \max\{|D^\alpha f(x)| : x \in \Omega, |\alpha| \leqslant k\}.$$

The family $\mathscr{Q} = \{\|\cdot\|_k : k = 0, 1, \ldots\}$ is a countable separating family of seminorms and so defines a locally convex metrizable topology, $\tau_{\mathscr{Q}}$ on $\mathscr{D}(\Omega)$. A local base is given by:

$$W_N = \left\{ f \in C^\infty(\Omega) : \|f\|_N < \frac{1}{N} \right\}, \qquad N = 1, 2, \ldots.$$

The subspace topology induced on $\mathscr{D}_K$ by $\tau_{\mathscr{Q}}$ is indeed τ_K. To see this, recall that τ_K is defined by the family of seminorms introduced in (A.1). Note that for any fixed compact $K \subset \Omega$ there exists N_0 such that $K \subset K_{N_0}$. For $N \geqslant N_0$ we have $\|f\|_N = p_N(f)$ for all $f \in \mathscr{D}_K$. Clearly then:

$$V_N \cap \mathscr{D}_K = W_N \cap \mathscr{D}_K, \qquad N = N_0, N_0 + 1, \ldots.$$

Suppose U is an open set in the subspace topology induced on $\mathscr{D}_K$ by $\tau_{\mathscr{Q}}$ and pick $x \in U$. Then $-x + U$ is a neighbourhood of the origin, and thus there exists n such that

$$W_n \cap \mathscr{D}_K \subset -x + U$$

But $W_{m+1} \subset W_m$ for all m, so without loss of generality we may assume $n \geqslant N_0$. But then we conclude that

$$V_n \cap \mathscr{D}_K = W_n \cap \mathscr{D}_K \subset -x + U$$

and so U is open in τ_K. An identical argument shows the reverse inclusion: i.e. an open set in τ_K is open in the subspace topology induced on $\mathscr{D}_K$ by $\tau_{\mathscr{Q}}$.

Thus the topology $\tau_{\mathscr{Q}}$ is locally convex, and induces the right subspace topology on $\mathscr{D}_K$. However, it is not complete. To see this, consider $\Omega = \mathbb{R}$, and let ϕ be any non-zero function with support in $[0, 1]$. Consider the sequence of functions $(f_m)_{m \in \mathbb{N}}$ with:

$$\phi_m(x) = \phi(x-1) + \frac{1}{2}\phi(x-2) + \frac{1}{3}\phi(x-3) + \ldots + \frac{1}{m}\phi(x-m).$$

This is a Cauchy sequence with respect to the topology $\tau_{\mathscr{Q}}$: if $n < m$, then

$$\|\phi_n - \phi_m\|_k = \frac{1}{n}\|\phi\|_k,$$

thus for any N if $n, m > N \|\phi\|_N$ we have $\phi_n - \phi_m \in W_N$. On the other hand, the sequence has no limit in $\mathscr{D}(\Omega)$, since for sufficiently large m, ϕ_m has support outside any compact set. We thus are led to discard $\tau_{\mathscr{D}}$ as a prospective topology for $\mathscr{D}(\Omega)$.

In fact, the topology $\tau_{\mathscr{D}}$ is too *coarse*: in a sense, the notion of convergence is too loose. This suggests that we should seek a finer topology. The topology that we shall introduce for $\mathscr{D}(\Omega)$ will in fact be the *finest* locally convex topology such that the subspace topology induced on $\mathscr{D}_K$ agrees with the natural Fréchet topology for each compact $K \subset \Omega$.

Definition A.10. *Let Ω be a nonempty open subset of $\mathbb{R}^n$.*

a) For each compact $K \subset \Omega$, τ_K is the Fréchet space topology on $\mathscr{D}_K$ introduced above.

b) β is the collection of all convex balanced sets $W \subset \mathscr{D}(\Omega)$ such that $\mathscr{D}_K \cap W \in \tau_K$ for every compact $K \subset \Omega$.

c) τ is the collection of all (possibly empty) unions of sets of the form $\phi + W$ with $\phi \in \mathscr{D}(\Omega)$ and $W \in \beta$.

Theorem A.21. *a) τ is a topology for $\mathscr{D}(\Omega)$ and β is a local base for τ.*

b) τ makes $\mathscr{D}(\Omega)$ into a locally convex topological vector space.

Proof. 1. It is clear from the definition that $\mathscr{D}(\Omega), \emptyset \in \tau$ and that τ is closed under arbitrary unions. If we can show that for any $V_1, V_2 \in \tau$ and $\phi \in V_1 \cap V_2$, then

$$\phi + W \subset V_1 \cap V_2 \tag{A.2}$$

for some $W \in \beta$, then we can deduce that $V_1 \cap V_2$ is open and so τ is a topology. Moreover, setting $\phi = 0$ and $V_2 = \mathscr{D}(\Omega)$ in (A.2) we deduce that any neighbourhood of 0 contains an element of β and so β is a local base. To show *a)* then, it is enough to establish (A.2).

2. From the definition of τ, there exist $\phi_i \in \mathscr{D}(\Omega)$ and $W_i \in \beta$ such that $\phi \in \phi_i + W_i$ and $\phi_i + W_i \subset V_i$ for $i = 1, 2$. Choose a compact $K \subset \Omega$ such that $\phi, \phi_i \in \mathscr{D}_K$. Since $\mathscr{D}_K \cap W_i$ is open in $\mathscr{D}_K$, we have

$$\phi - \phi_i \in (1 - \delta_i) W_i \tag{A.3}$$

for some $\delta_i > 0$. To see this, recall that $\mathscr{D}_K$ is a topological vector space, so in particular scalar multiplication is continuous. Thus for any $\psi \in \mathscr{D}_K$, the map $F_\psi : \mathbb{R} \to \mathscr{D}_K$ given by $t \mapsto t\psi$ is continuous. Thus the set $A = (F_{\phi-\phi_i})^{-1}[W_i \cap \mathscr{D}_K]$ is open in $\mathbb{R}$. In particular, there exists ϵ_i such that $(1 - 2\epsilon_i, 1 + 2\epsilon_i) \subset A$, which is equivalent to $t(\phi - \phi_i) \in W_i \cap \mathscr{D}_K$ for $t \in (1 - 2\epsilon_i, 1 + 2\epsilon_i)$. But if $(1 + \epsilon_i)(\phi - \phi_i) \in W_i \cap \mathscr{D}_K$, then (A.3) must hold for some $\delta_i > 0$.

3. Since W_i is convex, we can use the result of Exercise A.1 to deduce that

$$\phi - \phi_i + \delta_i W_i \subset (1 - \delta_i) W_i + \delta_i W_i = W_i$$

whence we deduce that

$$\phi + \delta_i W_i \subset \phi_i + W_i \subset V_i.$$

Taking $W = \delta_1 \phi_1 \cap \delta_2 \phi_2$ we have established (A.2) and thus proven part a) of the theorem.

4. To show that τ makes $\mathscr{D}(\Omega)$ into a locally convex topological space it is enough to show that the topological vector space axioms are satisfied. Since β is a local base, and is convex by construction the result will follow. Suppose that $\phi_1, \phi_2 \in \mathscr{D}(\Omega)$ are distinct, and consider the set:

$$W = \{\phi \in \mathscr{D}(\Omega) : \|\phi\|_0 \leqslant \|\phi_1 - \phi_2\|_0\}$$

This is certainly convex as it is a metric ball. Moreover, since the sets $\{\phi \in \mathscr{D}_K : \|\phi\|_0 < r\}$ (and their translations) are open in τ_K for any r and all compact $K \subset \Omega$, we conclude that $W \in \beta$. Moreover, $\phi_1 \notin \phi_2 + W$. Thus the singleton set $\{\phi_1\}$ is closed in τ.

5. To establish the τ-continuity of addition, suppose $U \in \tau$ is any open set, and suppose that we have $\phi_1, \phi_2 \in \mathscr{D}(\Omega)$ with $\phi_1 + \phi_2 \in U$. Since β is a local base, $\phi_1 + \phi_2 + W \subset U$ for some $W \in \beta$. I claim $(\phi_1 + \frac{1}{2}W) \times (\phi_2 + \frac{1}{2}W) \subset (+)^{-1}(U)$. To see this, note that by the convexity of W:

$$(\phi_1 + \frac{1}{2}W) + (\phi_2 + \frac{1}{2}W) = \phi_1 + \phi_2 + W \subset U.$$

Thus $(+)^{-1}(U)$ is open in the product topology and addition is τ-continuous.

6. Finally, to show that scalar multiplication is continuous, suppose $U \in \tau$ is any open set, and suppose that we have $\alpha \in \Phi$, $\phi \in \mathscr{D}(\Omega)$ with $\alpha\phi \in U$. Since β is a local base, $\alpha\phi + W \subset U$ for some $W \in \beta$. I claim that for ϵ, δ sufficiently small, we have $\{\alpha' \in \Phi : |\alpha' - \alpha| < \delta\} \times (\phi + \epsilon W) \subset (\cdot)^{-1}(U)$. Note that

$$\alpha'\phi' - \alpha\phi = \alpha'(\phi' - \phi) + (\alpha' - \alpha)\phi$$

Now, by a similar argument to that in paragraph 2. above, the continuity of scalar multiplication restricted to $\mathscr{D}_K$ for a compact K which contains the support of ϕ ensures we can choose $\delta > 0$ such that $\delta\phi \in \frac{1}{2}W$. Let us set $\epsilon = (2(|\alpha| + \delta))^{-1}$. By the fact that W is balanced and convex, we deduce that

$$\alpha'\phi' - \alpha\phi \in \frac{1}{2}W + \frac{1}{2}W = W,$$

so that $\{\alpha' \in \Phi : |\alpha' - \alpha| < \delta\} \times (\phi + \epsilon W) \subset (\cdot)^{-1}(U)$ and scalar multiplication is indeed continuous. □

From now on, whenever we refer to $\mathscr{D}(\Omega)$, we shall assume that it is given the topology τ that has just been constructed. The main results of this section (indeed this chapter) are the following two results which characterise convergence and continuity in $\mathscr{D}(\Omega)$. These results justify the approach taken in lectures to disregard a close study of the topology of $\mathscr{D}(\Omega)$ and focus instead on sequential definitions of continuity.

Theorem A.22. *a) A convex balanced subset V of $\mathscr{D}(\Omega)$ is open if any only if $V \in \beta$.*

b) The Fréchet topology τ_K of any $\mathscr{D}_K \subset \mathscr{D}(\Omega)$ coincides with the subspace topology that $\mathscr{D}_K$ inherits from $\mathscr{D}(\Omega)$.

c) If E is a bounded subset of $\mathscr{D}(\Omega)$ then $E \subset \mathscr{D}_K$ for some $K \subset \Omega$, and there are real numbers $M_N < \infty$ such that every $\phi \in E$ satisfies the inequalities

$$\|\phi\|_N \leqslant M_N, \qquad N = 0, 1, \ldots$$

d) $\mathscr{D}(\Omega)$ has the Heine-Borel property.

e) If $(\phi_i)_{i\in\mathbb{N}}$ is a Cauchy sequence in $\mathscr{D}(\Omega)$, then $\{\phi_i\}_{i\in\mathbb{N}} \subset \mathscr{D}_K$ for some compact $K \subset \Omega$, and (ϕ_i) is Cauchy with respect to the norm $\|\cdot\|_N$ for each $N = 0, 1, \ldots$.

f) If $\phi_i \to 0$ in $\mathscr{D}(\Omega)$, then there is a compact $K \subset \Omega$ which contains the support of every ϕ_i , and $D^\alpha \phi_i \to 0$ uniformly, as $i \to \infty$ for every multiindex α.

g) In $\mathscr{D}(\Omega)$, every Cauchy sequence converges.

Proof. 1. Since β is a local base, clearly if $V \in \beta$ then it is open. Now suppose V is an arbitrary convex, balanced, open set. Let K be any compact subset of Ω and pick $\phi \in \mathscr{D}_K \cap V$. Since β is a local base, we have $\phi + W \subset V$ for some $W \in \beta$. Thus

$$\phi + (\mathscr{D}_K \cap W) \subset \mathscr{D}_K \cap V.$$

From the definition of β, we know that $\mathscr{D}_K \cap W \in \tau_K$, so we have shown that $\mathscr{D}_K \cap V$ is open in $\mathscr{D}_K$. Since β contains all convex, balanced sets whose intersection with each $\mathscr{D}_K$ is open, $V \in \beta$ and we have established *a*).

2. The previous paragraph shows that any element of $\tau|_{\mathscr{D}_K}$ also belongs to τ_K, i.e. any set which is open with respect to the subspace topology is open in the Fréchet topology. Suppose now that $E \in \tau_K$. To show $E \in \mathscr{D}_K \cap W \in \tau_K$, we have to show that $E = \mathscr{D}_K \cap U$ for some $U \in \tau$. Suppose $\phi \in E$. Then from the definition of the topology of τ_K, there exists N, δ such that:

$$\{\psi \in \mathscr{D}_K : \|\psi - \phi\|_N < \delta\} \subset E.$$

Let:

$$W_\phi = \{\psi \in \mathscr{D}(\Omega) : \|\psi\|_N < \delta\}.$$

Then $W_\phi \in \beta$, and moreover

$$\mathscr{D}_K \cap (\phi + W_\phi) = \phi + \mathscr{D}_K \cap W_\phi \subset E$$

Taking:

$$U = \bigcup_{\phi \in E} (\phi + W_\phi)$$

we have $U \in \tau$ and $E = \mathscr{D}_K \cap U$. This establishes *b*).

3. To show the first part of *c*), we prove the contrapositive. Suppose $E \subset \mathscr{D}(\Omega)$ does not lie in any $\mathscr{D}_K$. Let $\{K_m\}_{m\in\mathbb{N}}$ be an exhaustion of Ω. Since $E \not\subset \mathscr{D}_{K_m}$, for each m, we can find $\phi_m \in E$ with $\text{supp}\phi_m \not\subset K_m$. In particular, there exists $x_m \in \Omega \setminus K_m$ with $\phi_m(x_m) \neq 0$. Let

$$W = \{\phi \in \mathscr{D}(\Omega) : |\phi(x_m)| < m^{-1} |\phi_m(x_m)|\}.$$

A short calculation shows W is convex and balanced. Suppose $K \subset \Omega$ is compact, then there exists M such that $K \subset K_M$. In particular this implies $x_m \notin K$ for $m \geqslant M$. Thus $\mathscr{D}_K \cap W$ is an intersection of finitely many open sets, and so $W \in \beta$. However, $\phi_m \notin mW$ for any m, so E is not bounded. Thus any bounded set in $\mathscr{D}(\Omega)$ belongs to $\mathscr{D}_K$ for some K. By *b*), E is thus bounded in $\mathscr{D}_K$, and the final part of *c*) follows by Theorem A.17.

4. Statement *d*) follows from *c*), since $\mathscr{D}_K$ has the Heine-Borel property (Theorem A.20). Since Cauchy sequences are bounded (Exercise A.9), *c*) implies that every Cauchy sequence $(\phi_i)_{i\in\mathbb{N}}$ lies in some $\mathscr{D}_K$. By *b*), $(\phi_i)_{i\in\mathbb{N}}$ is Cauchy with respect to τ_K, and *e*) follows. Statement *f*) is a restatement of *e*). Finally, *g*) follows from *e*) together with *b*) and the completeness of $\mathscr{D}_K$.

□

The final major result of this section concerns linear maps from $\mathscr{D}(\Omega)$ into a locally convex space. Before we state the theorem, we introduce the notion of a bounded operator as one which takes bounded sets to bounded sets. That is to say if X, Y are topological vector spaces, then a linear map $\Lambda : X \to Y$ is bounded if $\Lambda(E)$ is bounded in Y whenever E is bounded in X.

Theorem A.23. *Let Y be a locally convex topological vector space. Suppose that $\Lambda : \mathscr{D}(\Omega) \to Y$ is a* linear *mapping. Then the following are equivalent:*

a) Λ is continuous.

b) Λ is bounded.

c) If $\phi_i \to 0$ in $\mathscr{D}(\Omega)$ then $\Lambda\phi_i \to 0$ in Y.

d) The restrictions of Λ to every $\mathscr{D}_K \subset \mathscr{D}(\Omega)$ are continuous.

a) ⇒ *b*) Let E be a bounded set in $\mathscr{D}(\Omega)$ and let W be a neighbourhood of 0 in Y. Since Λ is continuous, there exists a neighbourhood V of 0 in $\mathscr{D}(\Omega)$ such that $\Lambda(V) \subset W$. Since E is bounded, there exists $s > 0$ such that $E \subset tV$ for all $t > s$. Since Λ is linear, $\Lambda(E) \subset \Lambda(tV) = t\Lambda(V) \subset tW$, and hence $\Lambda(E)$ is bounded.

b) ⇒ *c*) By part *e*) of Theorem A.22, if $\phi_i \to 0$ in $\mathscr{D}(\Omega)$, then there exists $K \subset \Omega$ such that $\phi_i \to 0$ in $\mathscr{D}_K$. Since $\mathscr{D}_K$ is metrizable, there exist scalars $\alpha_i \to \infty$ such that $\alpha_i\phi_i \to 0$ in $\mathscr{D}_K$ and hence in $\mathscr{D}(\Omega)$ by part *b*) of Theorem A.22. By the linearity of Λ, we have

$$\Lambda\phi_i = \alpha_i^{-1}\Lambda(\alpha_i\phi_i).$$

Now $(\alpha_i \phi_i)$ is Cauchy in $\mathscr{D}(\Omega)$, and hence bounded. Since Λ is bounded by assumption, $\{\Lambda(\alpha_i \phi_i)\}$ is bounded. As $\alpha_i^{-1} \to 0$, by Lemma A.12, part *d*), $\Lambda\phi_i \to 0$.

c) $\Rightarrow$ *d*) By part *b*) of Theorem A.22, we have that *c*) implies that if $\phi_i \to 0$ in $\mathscr{D}_K$ then $\Lambda\phi_i \to 0$. We work by contradiction. Suppose that the restriction of Λ to $\mathscr{D}_K$ is not continuous. Then there exists a neighbourhood W of 0 in Y such that $\Lambda^{-1}(W)$ contains no neighbourhood of 0 in $\mathscr{D}_K$. Since $\mathscr{D}_K$ is metrisable, pick a metric d which generates τ_K and construct a sequence (x_n) by choosing x_n such that $d(x_n, 0) < n^{-1}$ and $x_n \notin \Lambda^{-1}(W)$. Then $x_n \to 0$ in $\mathscr{D}_K$ and hence in $\mathscr{D}(\Omega)$, but $\Lambda(x_n) \not\to 0$, contradicting *c*).

d) $\Rightarrow$ *a*) Suppose that U is a convex, balanced, neighbourhood of the origin in Y and set $V = \Lambda^{-1}(U)$. Then V is convex and balanced by the linearity of Λ. By part *a*) of Theorem A.22, V is open in $\mathscr{D}(\Omega)$ if $\mathscr{D}_K \cap V$ is open in $\mathscr{D}_K$ for every $\mathscr{D}_K \subset \mathscr{D}(\Omega)$, but if Λ is continuous when restricted to each $\mathscr{D}_K$, then $\mathscr{D}_K \cap V$ is open in $\mathscr{D}_K$ from the definition of continuity. Thus V is open. Now suppose that W is any open set in Y, and suppose $\phi \in \Lambda^{-1}(W)$. Since Y is locally convex, there exists U, a convex neighbourhood of 0 in U such that $\Lambda\phi + U \subset W$. By Theorem A.11, we may assume that U is balanced. Since Λ is linear, $\phi + \Lambda^{-1}(U) \subset \Lambda^{-1}(W)$, and $\phi + \Lambda^{-1}(U)$ is open in $\mathscr{D}(\Omega)$, so Λ is continuous.

Appendix B

Background Material: Measure Theory and integration

In this appendix we shall briefly review some of the basics of measure theory, including sigma algebras, measurable spaces, measures and the construction of the Lebesgue measure. These notes follow parts of the notes from Prof. Norris' version of the course *Probability and measure*, as well as the books *Real and complex analysis* by Rudin, *Real Analysis* by Stein and Shakarchi annd *Measure Theory and Integration* by M. Taylor.

B.1 Sigma algebras and measures

Given a set E, the basic goal of measure theory is to assign to certain subsets $A \subset E$ a value, $\mu(A)$ which represents in some appropriate sense the 'size' of A. For example, if E is finite or countable and A is any subset of E we might set $\mu(A)$ to be the number of elements in A (where $\mu(A)$ may be ∞ if A is not finite). In this case μ is defined on all of the power set 2^E. We call μ the *counting measure*.

For $E = \mathbb{R}$, it is natural to wish to define $\mu(A)$ to be the 'length' of A. This is unambiguous if A is some interval, but it turns out that we run into problems trying to define the 'length' of an arbitrary subset of $\mathbb{R}$. As a consequence, we will need to restrict our attention to a smaller collection of sets than the power set $2^{\mathbb{R}}$.

Definition B.1. *Let E be a set. A collection $\mathcal{E}$ of subsets of E is called a* σ-algebra[1] *if $\mathcal{E}$ contains $\emptyset$ and is closed under taking the complement and forming countable unions. That is if $A \in \mathcal{E}$ then*

$$A^c = \{x \in E | x \notin A\} \in \mathcal{E},$$

and if $(A_n)_{n=1}^{\infty}$ is a sequence with $A_n \in \mathcal{E}$, then

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{E}.$$

$(E, \mathcal{E})$ *is called a* measurable space.

[1] pronounced "sigma algebra"

A measure *on* $(E, \mathcal{E})$ *is a set function* $\mu : \mathcal{E} \to [0, \infty]$ *such that* $\mu(\emptyset) = 0$ *and* μ *is* countably additive. *That is for a sequence* $(A_n)_{n=1}^{\infty}$ *with* $A_n \in \mathcal{E}$ *disjoint, we have:*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

$(E, \mathcal{E}, \mu)$ *is called a* measure space.

Note that $(E, 2^E)$ is always a measurable space, since 2^E is always a σ-algebra. Suppose $(E, \mathcal{E}, \mu)$ is a measure space and that $A \in \mathcal{E}$. Then we can define a new measure space $(A, \mathcal{E}|_A, \mu|_A)$ by taking $\mathcal{E}|_A = \{B \in \mathcal{E} : B \subset A\}$ and defining $\mu|_A(B) = \mu(B)$ for all $B \in \mathcal{E}|_A$.

Exercise B.1. Let E be finite or countable and $\mathcal{E} = 2^E$.

a) Verify that if μ is the counting measure, then $(E, \mathcal{E}, \mu)$ is a measure space.

b) A mass function is a map $m : E \to [0, \infty]$. Define a set-function on $(E, \mathcal{E})$ by

$$\mu_m(A) = \sum_{x \in A} m(x).$$

Show that μ_m is a measure on $(E, \mathcal{E})$, and moreover if μ is any measure on $(E, \mathcal{E})$ then $\mu = \mu_m$ for some m.

For the examples in Exercise B.1, we can identify in a straightforward way both an appropriate σ-algebra and measure. In more general situations we may not be so lucky, so it is very helpful to be able to appeal to abstract results to construct measure spaces by starting with something simpler. We shall require the following Lemma, whose proof we defer to an exercise.

Lemma B.1. *Suppose that for each* $i \in I$*, where* I *is some (not necessarily countable) index set,* $\mathcal{E}_i$ *is a* σ*-algebra of the set* E*. Then the intersection* $\cap_{i \in I} \mathcal{E}_i$ *is a* σ*-algebra.*

With this fact in hand we can define the σ-algebra generated by a collection of sets.

Definition B.2. *If* $\mathcal{A}$ *is a collection of subsets of* E*, then the* σ-algebra generated by $\mathcal{A}$*, denoted* $\sigma(\mathcal{A})$*, is the intersection of all* σ*-algebras* $\mathcal{E}$ *on* E *such that* $\mathcal{A} \subset \mathcal{E}$*.*

Since 2^E is always a σ-algebra, and $\mathcal{A} \subset 2^E$, $\sigma(\mathcal{A})$ is always well defined. When (E, τ) is a topological space, with τ the collection of open sets, it is natural to introduce the *Borel algebra*[2] $\mathcal{B}(E) := \sigma(\tau)$. When $E = \mathbb{R}$ with it's standard topology, we often write $\mathcal{B} := \mathcal{B}(\mathbb{R})$. A measure defined on the measure space $(E, \mathcal{B}(E))$ is called a *Borel measure.* A Borel measure which is finite on compact sets is called a *Radon measure.*

Exercise B.2. a) Prove Lemma B.1.

b) Let $E = \{1, 2, 3\}$. Find $\sigma(\{1\})$, and show that $\sigma(\{1\}) \neq 2^E$.

[2] The notation $\mathcal{B}(E)$ assumes that the topology is obvious from context

Exercise B.3. a) Show that if $U \subset \mathbb{R}$ is open in the standard topology, then:

$$U = \bigcup_{n=1}^{\infty} I_n,$$

where each $I_n = (a_n, b_n)$ with $a_n < b_n$ is an open interval, and the I_n's are disjoint.

b) Show that $\mathcal{B} = \sigma(\mathcal{A})$ when $\mathcal{A}$ is given by:

i) $\mathcal{A} = \{(a,b)|a,b \in \mathbb{R}, a < b\}$, the collection of all open intervals in $\mathbb{R}$.
ii) $\mathcal{A} = \{[a,b]|a,b \in \mathbb{R}, a < b\}$, the collection of all closed intervals in $\mathbb{R}$.
iii) $\mathcal{A} = \{(a,b]|a,b \in \mathbb{R}, a < b\}$.
iv) $\mathcal{A} = \{[a,b)|a,b \in \mathbb{R}, a < b\}$.
v) $\mathcal{A} = \{(-\infty,b)|b \in \mathbb{R}\}$.
vi) $\mathcal{A} = \{(-\infty,b]|b \in \mathbb{R}\}$.
vii) $\mathcal{A} = \{(a,\infty)|a \in \mathbb{R}\}$.
viii) $\mathcal{A} = \{[a,\infty)|a \in \mathbb{R}, a\}$.
ix) $\mathcal{A} = \{(a,b)|a,b \in \mathbb{Q}, a < b\}$.

[Hint : reduce cases ii) − ix) to case i).]

We have a means of generating a σ-algebra from a smaller collection of sets, $\mathcal{A}$. We'd like to define a measure by how it acts on $\mathcal{A}$, and then 'extend' this measure to act on $\sigma(\mathcal{A})$ (or some larger σ-algebra containing $\mathcal{A}$. For this we need both an existence and a uniqueness result for the extension. We first introduce the idea of π-system and d-system and establish Dynkin's π-system Lemma, which will eventually furnish a proof of uniqueness for extensions of measures.

Definition B.3. *Let $\mathcal{A}$ be a collection of subsets of E. We say that*

i) $\mathcal{A}$ is a π-system if it contains the empty set and is closed under pairwise intersection, i.e.

- $\emptyset \in \mathcal{A}$,
- $A \cap B \in \mathcal{A}$ *for all* $A, B \in \mathcal{A}$.

ii) $\mathcal{A}$ is a d-system if it contains E and is closed under taking differences, and countable unions of increasing sets, i.e.

- $E \in \mathcal{A}$,
- $B \setminus A \in \mathcal{A}$ *for all* $A, B \in \mathcal{A}$ *with* $A \subset B$,
- $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ *for all sequences* $(A_n)_{n=1}^{\infty}$ *with* $A_n \in \mathcal{A}$ *and* $A_n \subset A_{n+1}$.

Exercise($*$). Show that if $\mathcal{A}$ is both a π-system and a d-system, then it is a σ-algebra.

Dynkin's π-system Lemma extends the previous exercise.

Lemma B.2 (Dynkin's π-system Lemma)**.** *Let $\mathcal{A}$ be a π-system. Then any d-system containing $\mathcal{A}$ also contains $\sigma(\mathcal{A})$.*

Proof. Let $\mathcal{D}$ denote the intersection of all d-systems containing $\mathcal{A}$. Then $\mathcal{D}$ is a d-system and it suffices to show that $\sigma(\mathcal{A}) \subset \mathcal{D}$. In order to do this, we show that $\mathcal{D}$ is a π-system, hence it is a σ-algebra containing $\mathcal{A}$ and thus it must contain $\sigma(\mathcal{A})$ from the definition of $\sigma(\mathcal{A})$.

We introduce

$$\mathcal{D}' = \{B \in \mathcal{D} | B \cap A \in \mathcal{D} \text{ for all } A \in \mathcal{A}\}.$$

Clearly $\mathcal{A} \subset \mathcal{D}'$ because $\mathcal{A}$ is a π-system. Next we claim that $\mathcal{D}'$ is a d-system. Clearly $E \in \mathcal{D}'$. Suppose $B_1, B_2 \in \mathcal{D}'$ with $B_1 \subset B_2$, then for $A \in \mathcal{A}$ we have:

$$(B_2 \setminus B_1) \cap A = (B_2 \cap A) \setminus (B_1 \cap A) \in \mathcal{D}$$

because $\mathcal{D}$ is a d-system and we conclude $B_2 \setminus B_1 \in \mathcal{D}'$. Now suppose $B_n \in \mathcal{D}'$ and $B_n \subset B_{n+1}$ and let $B = \bigcup_{n=1}^{\infty} B_n$. Then for any $A \in \mathcal{A}$, we have $C_n := B_n \cap A \in \mathcal{D}$, $C_n \subset C_{n+1}$ so $\bigcup_{n=1}^{\infty} C_n = B \cap A \in \mathcal{D}$ as $\mathcal{D}$ is a d-system. We deduce that $\mathcal{D}'$ is a d-system containing $\mathcal{A}$, hence $\mathcal{D}' = \mathcal{D}$ by the minimality of $\mathcal{D}$.

Now, we let

$$\mathcal{D}'' = \{B \in \mathcal{D} | B \cap A \in \mathcal{D} \text{ for all } A \in \mathcal{D}\}.$$

By the above, we have that $\mathcal{A} \subset \mathcal{D}''$, since $\mathcal{D}' = \mathcal{D}$. By the same arguments as above we can check that $\mathcal{D}''$ is a d-system, and so $\mathcal{D}'' = \mathcal{D}$ and $\mathcal{D}''$ is a π-system as required. □

B.1.1 Construction of measures

As described above, we are going to give a means of constructing a measure by specifying how it behaves on some suitable collection of sets. First we introduce some notation concerning *set functions*

Definition B.4. *Let $\mathcal{A}$ be a collection of subsets of E containing $\emptyset$. A* set function *is a function $\mu : \mathcal{A} \to [0, \infty]$ with $\mu(\emptyset) = 0$. We say that a set function μ is:*

- increasing *if*
$$\mu(A) \leqslant \mu(B), \quad \text{for all } A, B \in \mathcal{A}, \text{ with } A \subset B,$$
- additive *if, for all* ***disjoint*** *sets $A, B \in \mathcal{A}$ with $A \cup B \in \mathcal{A}$ we have:*
$$\mu(A \cup B) = \mu(A) + \mu(B),$$
- countably additive *if for all sequence of* ***disjoint*** *sets $(A_n)_{n=1}^{\infty}$ with $A_n \in \mathcal{A}$ and $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$ we have:*
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n),$$

- countably subadditive *if for all sequences* $(A_n)_{n=1}^{\infty}$ *with* $A_n \in \mathcal{A}$ *and* $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$ *we have:*
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leqslant \sum_{n=1}^{\infty} \mu(A_n),$$

We shall also need to define what it means for a collection of subsets to be a ring

Definition B.5. *Let* $\mathcal{A}$ *be a collection of subsets of* E*. We say* $\mathcal{A}$ *is a* ring *on* E *if* $\emptyset \in \mathcal{A}$ *and for all* $A, B \in \mathcal{A}$*:*
$$B \setminus A \in \mathcal{A}, \quad A \cup B \in \mathcal{A}.$$
We say $\mathcal{A}$ *is an* algebra *if* $\emptyset \in \mathcal{A}$ *and for all* $A, B \in \mathcal{A}$*:*
$$A^c \in \mathcal{A}, \quad A \cup B \in \mathcal{A}.$$

Let us suppose that $\mathcal{A}$ is a ring of subsets of E, together with a countably additive set function $\mu : \mathcal{A} \to [0, \infty]$. For any set $B \subset E$, we can introduce the *outer measure*
$$\mu^*(B) := \inf \sum_{n=1}^{\infty} \mu(A_n),$$
where the infimum is taken over all sequences $(A_n)_{n=1}^{\infty}$ of sets such that $A_n \in \mathcal{A}$ and $B \subset \bigcup_{n=1}^{\infty} A_n$. If no such sequence exists we set $\mu^*(B) = \infty$. We clearly have $\mu^*(\emptyset) = 0$, so we have a set function defined on 2^E and moreover, μ^* is increasing. In general, however, μ^* will not define a measure on the measure space $(E, 2^E)$, in order for μ^* to be a measure we must restrict to a smaller σ-algebra. We say that $A \subset E$ is μ^*-measurable if, for all $B \subset E$ we have:
$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c),$$
and we denote by $\mathcal{M}$ the collection of all μ^*-measurable sets. One of the fundamental results of measure theory is:

Theorem B.3 (Carathéodory's Theorem)**.** *Suppose* $\mathcal{A}$ *is a ring of subsets of* E*, and* $\mu : \mathcal{A} \to [0, \infty]$ *is a countably additive set function. Define* $\mu^*, \mathcal{M}$ *as above. The collection* $\mathcal{M}$ *is a* σ*-algebra which contains* $\mathcal{A}$*. The set function* $\mu^* : \mathcal{M} \to [0, \infty]$ *is a measure on* $(E, \mathcal{M})$*.*

We shall establish this result through several Lemmas. First, we establish countable subadditivity of μ^*.

Lemma B.4. *The set function* $\mu^* : 2^E \to [0, \infty]$ *is countably subadditive.*

Proof. Let $B = \cup_{n=1}^{\infty} B_n$. We wish to show
$$\mu^*(B) \leqslant \sum_{n=1}^{\infty} (B_n).$$

We can easily see that if $\mu^*(B_n) = \infty$ for some n, then necessarily $\mu^*(B) = \infty$, so we can focus on the case where $\mu^*(B_n) < \infty$ for all n. Fix $\epsilon > 0$. For each n we can find a sequence of sets $(A_{n,m})_{m=1}^\infty$ such that $A_{n,m} \in \mathcal{A}$ with $B_n \subset \cup_{m=1}^\infty A_{n,m}$ and

$$\sum_{m=1}^{\infty} \mu(A_{n,m}) \leqslant \mu^*(B_n) + \epsilon 2^{-n}.$$

Now, $B \subset \bigcup_{n=1}^\infty \bigcup_{m=1}^\infty A_{n,m}$, so we have:

$$\mu^*(B) \leqslant \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mu(A_{n,m}) \leqslant \sum_{n=1}^{\infty} \mu(B_n) + \epsilon.$$

Since ϵ was arbitrary, the result follows. □

Next we show that μ^* extends μ.

Lemma B.5. *Suppose $A \in \mathcal{A}$. Then $\mu^*(A) = \mu(A)$.*

Proof. It is obvious that $\mu^*(A) \leqslant \mu(A)$, by considering the sequence $A_1 = A$, $A_n = \emptyset$ for $n > 1$, so it suffices to show $\mu^*(A) \geqslant \mu(A)$. Since μ is countably additive, it is finitely additive (take all but finitely many elements of the sequence to be the empty set). Since $\mathcal{A}$ is a ring, if $A, B \in \mathcal{A}$ with $A \subset B$, then $B \setminus A \in \mathcal{A}$. By finite additivity of μ:

$$\mu(B) = \mu\left(A \cup (B \setminus A)\right) = \mu(A) + \mu(B \setminus A) \geqslant \mu(A)$$

so μ is increasing. Suppose $(A_n)_{n=1}^\infty$ is a sequence with $A_n \in \mathcal{A}$. Let $B_1 = A_1$ and

$$B_n = \bigcup_{k=1}^{n} A_k \setminus \bigcup_{k=1}^{n-1} A_k$$

for $n > 1$. Then $(B_n)_{n=1}^\infty$ is a disjoint sequence, $B_n \subset A_n$ and moreover each $B_n \in \mathcal{A}$ since $\mathcal{A}$ is a ring. We have:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \leqslant \sum_{n=1}^{\infty} \mu(A_n),$$

so μ is countably subadditive.

Now, suppose $A \in \mathcal{A}$ and take any sequence $(A_n)_{n=1}^\infty$ with $A_n \in \mathcal{A}$ and $A \subset \cup_{n=1}^\infty A_n$. Note that $A \cap A_n = A \setminus ((A \cup A_n) \setminus A)$, so $A \cap A_n \in \mathcal{A}$. We deduce:

$$\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} (A \cap A_n)\right) \leqslant \sum_{n=1}^{\infty} \mu(A \cap A_n) \leqslant \sum_{n=1}^{\infty} \mu(A_n)$$

Taking the infimum over all such sequences, we conclude $\mu(A) \leqslant \mu^*(A)$ and we're done. □

Lemma B.6. *$\mathcal{M}$ contains $\mathcal{A}$.*

Proof. Suppose $A \in \mathcal{A}$ and $B \subset E$. We need to show:

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

Since $B = (B \cap A) \cup (B \cap A^c)$ and using subadditivity of μ^*, it is immediate that $\mu^*(B) \leqslant \mu^*(B \cap A) + \mu^*(B \cap A^c)$, so it suffices to show

$$\mu^*(B) \geqslant \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

If $\mu^*(B) = \infty$ this is trivial, so we can focus on the case $\mu^*(B) < \infty$. Fix $\epsilon > 0$, then there exists a sequence $(A_n)_{n=1}^{\infty}$ with $A_n \in \mathcal{A}$, $B \subset \cup_{n=1}^{\infty} A_n$ and

$$\sum_{n=1}^{\infty} \mu(A_n) \leqslant \mu^*(B) + \epsilon.$$

We note that:

$$B \cap A \subset \bigcup_{n=1}^{\infty} (A_n \cap A), \qquad B \cap A^c \subset \bigcup_{n=1}^{\infty} (A_n \cap A^c).$$

Recalling that $A_n \cap A \in \mathcal{A}$ and noting that $A_n \cap A^c = (A \cup A_n) \setminus A \in \mathcal{A}$ we deduce:

$$\mu^*(B \cap A) + \mu^*(B \cap A^c) \leqslant \sum_{n=1}^{\infty} \mu(A_n \cap A) + \sum_{n=1}^{\infty} \mu(A_n \cap A^c) = \sum_{n=1}^{\infty} \mu(A_n) \leqslant \mu^*(B) + \epsilon.$$

Since ϵ was arbitrary, we're done. □

Lemma B.7. *$\mathcal{M}$ is an algebra.*

Proof. From the definition of $\mathcal{M}$ it is immediate that $E \in \mathcal{M}$ and that $A \in \mathcal{M}$ implies $A^c \in \mathcal{M}$. It remains to show that $\mathcal{M}$ is closed under pairwise union, or equivalently pairwise intersection (since $A \cup B = (A^c \cap B^c)^c$). Suppose that $A_1, A_2 \in \mathcal{M}$ and $B \subset E$. Then

$$\begin{aligned}
\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c) \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c)
\end{aligned}$$

so $A_1 \cap A_2 \in \mathcal{M}$. □

Finally, we are ready to prove Carathéodory's theorem:

Proof of Theorem B.3. We already know that $\mathcal{M}$ is an algebra containing $\mathcal{A}$, so it suffices to show that if $(A_n)_{n=1}^{\infty}$ is a sequence of disjoint sets with $A_n \in \mathcal{M}$, and $A = \bigcup_{n=1}^{\infty} A_n$, then we have:

$$A \in \mathcal{A}, \qquad \mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A_n).$$

so that $\mathcal{M}$ is closed under countable unions and hence is a σ-algebra, and μ^* is a countably additive set function on $\mathcal{M}$, hence a measure. Fix any $B \subset E$. Since the A_n are disjoint, we know $A_1 \cap A_2 = \emptyset$ and $A_1 \cap A_2^c = A_1$. We deduce:

$$\begin{aligned}
\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\
&= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\
&= \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\
&= \ldots = \sum_{k=1}^{n} \mu^*(B \cap A_k) + \mu^*(B \cap A_1^c \cap \cdots \cap A_n^c)
\end{aligned}$$

Now, since $B \cap A^c \subset B \cap A_1^c \cap \cdots \cap A_n^c$, by the fact that μ^* is increasing we know $\mu^*(B \cap A_1^c \cap \cdots \cap A_n^c) \geqslant \mu^*(B \cap A^c)$. Hence, letting $n \to \infty$ and using countable subadditivity we find:

$$\mu^*(B) \geqslant \sum_{k=1}^{\infty} \mu^*(B \cap A_k) + \mu^*(B \cap A^c) \geqslant \mu^*(B \cap A) + \mu^*(B \cap A^c) \tag{B.1}$$

The reverse inequality holds by subadditivity, and so we have

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

and thus $A \in \mathcal{M}$. Setting $B = A$ in (B.1) we deduce:

$$\mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A_n).$$ □

Carathéodory's theorem gives a way to extend a countably additive set function defined on a ring $\mathcal{A}$ to a measure on $\sigma(\mathcal{A})$, since we can restrict the outer measure to $\sigma(\mathcal{A})$. It is often useful to know whether this extension of μ is unique. We have the following result:

Theorem B.8. *Let μ_1, μ_2 be measures on $(E, \mathcal{E})$ with $\mu_1(E) = \mu_2(E) < \infty$. Suppose that $\mu_1 = \mu_2$ on $\mathcal{A}$, where $\mathcal{A}$ is a π-system which generates $\mathcal{E}$. Then $\mu_1 = \mu_2$ on $\mathcal{E}$.*

Proof. Let $\mathcal{D} = \{A \in \mathcal{E} | \mu_1(A) = \mu_2(A)\}$ be the collection of sets on which the measures agree. By hypothesis $E \in \mathcal{D}$ and $\mathcal{A} \subset \mathcal{D}$. We shall show that $\mathcal{D}$ is a d-system, so by Dynkin's π-system Lemma we have $\mathcal{E} = \sigma(\mathcal{A}) \subset \mathcal{D}$ and we're done.

Suppose $A, B \in \mathcal{E}$ with $A \subset B$, then by additivity of the measures, we have:

$$\mu_1(A) + \mu_1(B \setminus A) = \mu_1(B) < \infty, \quad \mu_2(A) + \mu_2(B \setminus A) = \mu_2(B) < \infty,$$

so that if $A, B \in \mathcal{D}$ then $B \setminus A \in \mathcal{D}$.

Now suppose that we have a sequence $(A_n)_{n=1}^{\infty}$ with $A_n \in \mathcal{D}$ and $A_n \subset A_{n+1}$ and $A = \cup_{n=1}^{\infty} A_n$. Then setting $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n > 1$, we can write $A = \cup_{n=1}^{\infty} B_n$, where the B_n are disjoint. Thus:

$$\mu_1(A) = \sum_{n=1}^{\infty} \mu_1(B_n) = \sum_{n=1}^{\infty} \mu_2(B_n) = \mu_2(A)$$

and hence $A \in \mathcal{D}$. Thus $\mathcal{D}$ is a d-system and so $\mathcal{E} = \mathcal{D}$ and we're done. □

This result requires that E has *finite* measure. For many of the situations we're interested in this is too restrictive an assumption. We can extend the result for measures which satisfy a weaker condition.

Corollary B.9. *Let μ_1, μ_2 be measures on $(E, \mathcal{E})$. Suppose that $\mu_1 = \mu_2$ on $\mathcal{A}$, where $\mathcal{A}$ is a π-system which generates $\mathcal{E}$. Suppose also that $E = \bigcup_{i=1}^{\infty} B_i$, where $B_i \in \mathcal{A}$ and the B_i's are disjoint with $\mu_1(B_i) = \mu_2(B_i) < \infty$. Then $\mu_1 = \mu_2$ on $\mathcal{E}$.*

Proof. For each i, and for any $A \in \mathcal{E}$, define $\mu_1^i(A) = \mu_1(A \cap B_i)$, $\mu_2^i(A) = \mu_2(A \cap B_i)$. By assumption we have $\mu_1^i(E) = \mu_2^i(E) < \infty$ and moreover $\mu_1^i(A) = \mu_2^i(A)$ for all $A \in \mathcal{A}$. Thus $\mu_1^i = \mu_2^i$ on $\mathcal{E}$. Further, if $A \in \mathcal{E}$ is any measurable set, then

$$
\begin{aligned}
\mu_1(A) = \mu_1\left(\bigcup_{i=1}^{\infty}(B_i \cap A)\right) &= \sum_{i=1}^{\infty} \mu_1(B_i \cap A) \\
&= \sum_{i=1}^{\infty} \mu_2(B_i \cap A) = \mu_2\left(\bigcup_{i=1}^{\infty}(B_i \cap A)\right) = \mu_2(A)
\end{aligned}
$$

□

Completeness of measures

A useful feature of the measures obtained from Carathéodory's theorem is that they have a property known as *completeness.*

Definition B.6. *Let $(E, \mathcal{E}, \mu)$ be a measure space. We say μ is* complete *if for any $A \in \mathcal{E}$ with $\mu(A) = 0$, each subset of A also belongs to $\mathcal{E}$.*

A subset of a set of measure zero is sometimes known as a *null set*, so a complete measure is one for which all null sets are measurable.

Lemma B.10. *Suppose $(E, \mathcal{M}, \mu)$ is a measure space obtained from Carathéodory's theorem. Then it is complete.*

Proof. Let μ^* be the outer measure on E whose restriction to $\mathcal{M}$ gives μ. Suppose $N \subset A$, where $A \in \mathcal{M}$ with $\mu(A) = 0$. Since μ^* is increasing we have $\mu^*(N) \leqslant \mu(A) = 0$, so $\mu^*(N) = 0$. For any set $B \subset E$ we have:

$$\mu^*(T \cap N) + \mu^*(T \cap N^c) \leqslant \mu^*(N) + \mu^*(T) = \mu^*(T)$$

again using the increasing property of μ^*. By Lemma B.4 we know μ^* is subadditive, hence

$$\mu^*(T) \leqslant \mu^*(T \cap N) + \mu^*(T \cap N^c),$$

and thus $N \in \mathcal{M}$. □

B.1.2 Lebesgue measure

We specialise now to (arguably) the most important measure, the Lebesgue measure. This measure gives us the standard notion of volume for sets in $\mathbb{R}^n$. We first introduce the *rectangles* in $\mathbb{R}^n$.

Definition B.7. *A* rectangle *in $\mathbb{R}^n$ is a set of the form:*

$$R = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n],$$

with $a_i < b_i$ for $i = 1, \dots, n$. We define $\mathcal{A}_R$ to be the collection of finite unions of disjoint rectangles.

Exercise(∗). Show that:

a) The collection of rectangles is a π-system.

b) $\mathcal{A}_R$ is a ring.

c) $\mathcal{A}_R$ generates $\mathcal{B}(\mathbb{R}^n)$.

The main result we will establish shows:

Theorem B.11. *There exists a unique Borel measure μ on $\mathbb{R}^n$ such that, for all rectangles $R = (a_1, b_1] \times \cdots \times (a_n, b_n]$ with $a_i < b_i$ for $i = 1, \dots, n$,*

$$\mu(R) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n).$$

The measure μ is called the Lebesgue measure on $\mathbb{R}^n$.

Proof. For any $A \in \mathcal{A}_R$ we can write $A = \cup_{i=1}^N R_i$ for disjoint rectangles $R_i := (a_1^i, b_1^i] \times \cdots \times (a_n^i, b_n^i]$. We define for such A:

$$\mu(A) := \sum_{i=1}^{n} (b_1^i - a_1^i)(b_2^i - a_2^i) \cdots (b_n^i - a_n^i).$$

Note that the decomposition of A into rectangles is not unique, however one can verify that this is well defined and additive. If we can show that μ is countable additive, then we can apply Carathéodory's theorem to establish the existence of the Lebesgue measure.

Suppose that $(A_n)_{n=1}^\infty$ is a sequence of disjoint sets with $A_n \in \mathcal{A}_R$, such that $A = \cup_{i=1}^\infty A_i \in \mathcal{A}_R$. We wish to show that

$$\sum_{i=1}^{\infty} \mu(A_i) = \mu(A)$$

Set $B_n = \cup_{i=n}^\infty A_i$, note $\cap_{i=1}^\infty B_i = \emptyset$ as the sets A_i are disjoint. Since $\mathcal{A}_R$ is a ring, $B_n = A \setminus \cup_{i=1}^{n-1} A_i \in \mathcal{A}_R$. By finite additivity of μ we have:

$$\mu(A) = \sum_{i=1}^{n-1} \mu(A_i) + \mu(B_n),$$

so it suffices to prove that $\mu(B_n) \to 0$ as $n \to \infty$. Suppose not, then there exists $\epsilon > 0$ such that $\mu(B_n) \geqslant 2\epsilon$ for all n. For each n we can find $C_n \in \mathcal{A}$ with $\overline{C}_n \subset B_n$ and $\mu(C_n \setminus B_n) \leqslant \epsilon 2^{-n}$. Then

$$\mu(B_n \setminus (C_1 \cap \cdots \cap C_n)) \leqslant \mu((B_1 \setminus C_1) \cup \ldots \cup (B_n \setminus C_n)) \leqslant \sum_{n=1}^{\infty} \epsilon 2^{-n} = \epsilon.$$

Since $\mu(B_n) \geqslant 2\epsilon$, we must have $\mu(C_1 \cap \cdots \cap C_n) \geqslant \epsilon$, so $C_1 \cap \cdots \cap C_n \neq \emptyset$ and so $K_n = \overline{C}_1 \cap \cdots \cap \overline{C}_n \neq \emptyset$. Now, K_n is a nested sequence of non-empty compact sets, and so $\emptyset \neq \cap_{i=1}^{\infty} K_i \subset \cap_{i=1}^{\infty} B_i$ which is a contradiction.

Thus, we conclude that a Borel measure μ exists on $\mathbb{R}^n$ with the required property acting on rectangles. In order to establish uniqueness, we can invoke Corollary B.9, after noting that the set of rectangles is a π-system and that moreover we can write $\mathbb{R}^n$ as a countable disjoint union of rectangles, for example by taking the rectangles of the form $z + (0,1]^n$, where $z \in \mathbb{Z}^n$. □

We note that the Lebesgue measure is translation invariant: $\mu(B + x) = \mu(B)$ for any $x \in \mathbb{R}^n$, $B \in \mathcal{B}(\mathbb{R}^n)$. To see this, for fixed $x \in \mathbb{R}^n$ let $\mu_x(B) = \mu(B)$. If B is a rectangle, then $\mu_x(R) = \mu(R)$ (since $b_1 - a_1 = (b_1 - x_1) - (a_1 - x_1)$, etc.) so by uniqueness $\mu_x = \mu$. We also note that Carathéodory's theorem actually shows us that the Lebesgue measure is actually defined on $\mathcal{M}$, a larger σ-algebra than $\mathcal{B}(\mathbb{R}^n)$. We call $\mathcal{M}$ the algebra of Lebesgue measurable sets. By construction, we have that the Lebesgue measure is complete when $\mathbb{R}^n$ is equipped with $\mathcal{M}$ as σ-algebra, however it is not complete on the Borel algebra. For any Lebesgue measurable subset $E \subset \mathbb{R}^n$ we can define the natural restriction of Lebesgue measure to E, which we also refer to as the Lebesgue measure.

Lemma B.12 (Borel regularity of Lebesgue measure)**.** *Suppose $A \in \mathcal{M}$ is Lebesgue measurable. Then for any $\epsilon > 0$ there exists an open set O and a closed set C such that $C \subset A \subset O$ and:*

$$\mu(O \setminus A) < \epsilon, \qquad \mu(A \setminus C) < \epsilon.$$

If $\mu(A) < \infty$, then we may take C to be compact.

Proof. First, let us assume $\mu(A) < \infty$. From the definition of Lebesgue measurability, we know that

$$\mu(A) = \mu^*(A) = \inf \sum_{n=1}^{\infty} \mu(A_n),$$

where the infimum is taken over all sequences $(A_n)_{n=1}^{\infty}$ of sets such that $A_n \in \mathcal{A}_R$ and $A \subset \cup_{i=1}^{n} A_n$. Since each $A_n \in \mathcal{A}_R$ is a finite disjoint union of rectangles, we may assume without loss of generality that each A_n is a rectangle. Fix $\epsilon > 0$. We can choose A_n such that:

$$\inf \sum_{n=1}^{\infty} \mu(A_n) < \mu(A) + \frac{\epsilon}{2}$$

For each rectangle A_n, we can find a rectangle $\tilde{A}_n$ with $A_n \subset \tilde{A}_n^\circ$ and $\mu(\tilde{A}_n) < \mu(A_n) + \frac{\epsilon}{2^{n+1}}$. Then let $O = \cup_{n=1}^\infty \tilde{A}_n^\circ$. By construction, $A \subset O$ and O is open. Moreover,

$$\mu(O) \leqslant \sum_{n=1}^\infty \mu(\tilde{A}_n) \leqslant \sum_{n=1}^\infty \mu(A_n) + \frac{\epsilon}{2} \sum_{n=1}^\infty 2^{-n} < \mu(A) + \epsilon.$$

We deduce that

$$\mu(O \setminus A) < \epsilon.$$

Now suppose $\mu(A) = \infty$. Set $A_k = A \cap \{|x| \leqslant k\}$, then $\mu(A_k) < \infty$, so we can find an open O_k with $\mu(O_k \setminus A_k) < \epsilon 2^{-k}$. We set $O = \cup_{k=1}^\infty O_k$. Then O is open and $A \subset O$. Moreover,

$$O \setminus A = (\cup_{k=1}^\infty O_k) \setminus A = \bigcup_{k=1}^\infty (O_k \setminus A) \subset \bigcup_{k=1}^\infty (O_k \setminus A_k)$$

so that

$$\mu(O \setminus A) \leqslant \sum_{k=1}^\infty \mu(O_k \setminus A_k) < \epsilon.$$

We have thus established the first part of the proof. For the second part, we note that if A is measurable, then so is A^c, and hence there exists an open O with $A^c \subset O$ and $\mu(O \setminus A^c) < \epsilon$. Set $C = O^c$. This is closed and $C \subset A$. Moreover, $A \setminus C = C^c \setminus A^c = O \setminus A^c$ so

$$\mu(A \setminus C) < \epsilon.$$

For the final observation, note that if $\mu(A) < \infty$, then since A_k is an increasing sequence with $\cup_k A_k = A$, we have that $\lim_{k\to\infty} \mu(A_k) = \mu(A) < \infty$, so there exists k such that $\mu(A \setminus A_k) = \mu(A) - \mu(A_k) < \frac{\epsilon}{2}$. Let $C \subset A_k$ be a closed set such that $\mu(A_k \setminus C) < \frac{\epsilon}{2}$. We have $\mu(A \setminus C) = \mu((A \setminus A_k) \cup (A_k \setminus C)) < \epsilon$, and moreover C is a subset of a bounded set, hence compact. □

We next show

Lemma B.13. *Let $A \subset \mathbb{R}^n$. Suppose that for any $\epsilon > 0$ there exists an open set O and a closed set C such that $C \subset A \subset O$ and:*

$$\mu(O \setminus C) < \epsilon.$$

Then $A = B_1 \cup N$, where $N \subset B_2$ where $B_1, B_2 \in \mathcal{B}(\mathbb{R}^n)$ with $\mu(B_2) = 0$.

Proof. For each i, we can find O_i open and C_i closed such that $C_i \subset A \subset O_i$ and

$$\mu(O_i \setminus C_i) < 2^{-i}.$$

We have that $B_1 = \cup_{i=1}^\infty C_i \in \mathcal{B}(\mathbb{R}^n)$ from the properties of σ-algebras. Furthermore, let $B_2 = \cap_{i=1}^\infty (O_i \setminus C_i)$. Again $B_2 \in \mathcal{B}(\mathbb{R}^n)$, and moreover:

$$\mu(B_2) \leqslant \mu\left(\cap_{i=1}^n (O_i \setminus C_i)\right) \leqslant 2^{-n+1}$$

for any n, so $\mu(B_2) = 0$. Since $A \setminus B_1 \subset B_2$ we are done. □

Now, noting that the union of a Borel set and a null set is Lebesgue measurable by the completeness of Lebesgue measure, we have established:

Theorem B.14. *Suppose $A \subset \mathbb{R}^n$. The following are equivalent:*

i) *A is Lebesgue measurable.*

ii) *For any $\epsilon > 0$ there exists an open set O and a closed set C such that $C \subset A \subset O$ and:*

$$\mu(O \setminus C) < \epsilon.$$

iii) *$A = B_1 \cup N$, where $N \subset B_2$ where $B_1, B_2 \in \mathcal{B}(\mathbb{R}^n)$ with $\mu(B_2) = 0$.*

B.2 Measurable functions

We next wish to introduce the idea of a *measurable function* between two measurable spaces. Suppose $(E, \mathcal{E})$ and $(G, \mathcal{G})$ are measurable spaces. We say $f : E \to G$ is *measurable* if $f^{-1}(A) \in \mathcal{E}$ whenever $A \in \mathcal{G}$. Note the similarity to the definition of continuous maps between topological spaces. If $(G, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$, then we simply refer to a measurable function on $(E, \mathcal{E})$. If[3] $(G, \mathcal{G}) = (\mathbb{R}, \mathcal{B}[0, \infty])$, we refer to a non-negative measurable function. While convenient, this nomenclature has the slightly unfortunate consequence that a non-negative measurable function need not be a measurable function. If E is a topological space and $\mathcal{E} = \mathcal{B}(E)$, then a measurable function on $(E, \mathcal{E})$ is called a Borel function on E.

Exercise B.4. a) Suppose $(G, \mathcal{G})$ is a measurable space and E is any set. Show that if $f : E \to G$ is any function, the collection:

$$f^{-1}(\mathcal{G}) = \{f^{-1}(A) : A \in \mathcal{G}\},$$

is a σ-algebra, known as the pull-back σ-algebra.

b) Suppose $(E, \mathcal{E})$ and $(G, \mathcal{G})$ are measurable spaces, with $\mathcal{G} = \sigma(\mathcal{A})$ for some collection $\mathcal{A}$. Further suppose that $f : E \to G$ has the property that $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{A}$. Show that

$$\{A \subset G : f^{-1}(A) \in \mathcal{E}\}$$

is a σ-algebra containing $\mathcal{A}$ and deduce that f is measurable.

c) Suppose $(E, \mathcal{E})$ is a measurable space. Show that $f : E \to \mathbb{R}$ is measurable if and only if

$$f^{-1}((-\infty, \lambda)) := \{x \in E : f(x) < \lambda\} \in \mathcal{E}, \quad \text{for all } \lambda \in \mathbb{R}.$$

and $f : [0, \infty]$ is measurable if and only if

$$f^{-1}([0, \lambda)) := \{x \in E : 0 \leqslant f(x) < \lambda\} \in \mathcal{E}, \quad \text{for all } 0 \leqslant \lambda < \infty.$$

[3] We give $[0, \infty]$ a topology by saying $U \subset [0, \infty]$ is open if and only if $\tan^{-1}(U)$ is open in the standard topology of $[0, \frac{\pi}{2}]$, where by convention $\tan(\pi/2) = +\infty$.

Exercise B.5. Suppose E, G are topological spaces equipped with their Borel σ-algebras.

a) Show that any continuous function $f : E \to G$ is measurable. Deduce that in particular any continuous function $f : E \to \mathbb{R}$ is a Borel function.

b) Show that if $g : G \to \mathbb{R}$ is continuous, and $f : E \to G$ is measurable then $g \circ f$ is measurable.

c) Let $G = \mathbb{R}^n$, with its canonical basis $(e_i)_{i=1}^n$. Show that $f : E \to G$ is measurable if and only if each component function $f_i = (f, e_i) : E \to \mathbb{R}$ is measurable.

An important feature of the class of measurable functions (and indeed a strong motivation for the development of the theory) is that it behaves well under limiting operations.

Theorem B.15. *Suppose $(E, \mathcal{E})$ is a measurable space and $(f_n)_{n=1}^\infty$ is a sequence of non-negative measurable functions. Then the functions $f_1 + af_2$ for $a \geqslant 0$ and $f_1 f_2$ are measurable, as are*

$$\inf_n f_n, \quad \sup_n f_n, \quad \liminf_n f_n, \quad \limsup_n f_n.$$

In particular, if $f_n(x) \to f(x)$, then f is measurable.

The same results hold for (not necessarily non-negative) measurable functions, provided the limiting functions are real valued (i.e. don't take the values $\pm\infty$).

Proof. By Exercise B.4 we know that $f_1^{-1}([0,\lambda))f_2^{-1}([0,\lambda)) \in \mathcal{E}$ for any $0 \leqslant \lambda < \infty$. Now, for any $0 \leqslant \lambda < \infty$:

$$(f_1 + af_2)^{-1}([0,\lambda]) = \bigcup_{r \in \mathbb{Q}, r \geqslant 0} [\{f < \lambda - ar\} \cap \{g < r\}] \in \mathcal{E}.$$

so $f_1 + af_2$ is measurable. We also note that f_1^2 is measurable, since $(f_1^2)^{-1}([0,\lambda)) = f_1^{-1}([0,\lambda^{\frac{1}{2}})) \in \mathcal{E}$. Combining these two results, and noting

$$f_1 f_2 = \frac{1}{4}\left((f_1 + f_2)^2 - (f_1 - f_2)^2\right)$$

we deduce that $f_1 f_2$ is measurable. Next, we note that

$$\{\inf_n f_n < \lambda\} = \bigcup_n \{f_n < \lambda\}$$

so $\inf_n f_n$ is measurable. Similarly,

$$\{\sup_n f_n < \lambda\} = \bigcup_{r \in \mathbb{Q}, r < \lambda} \left(\bigcap_n \{f_n < r\}\right)$$

so $\sup_n f_n$ is measurable. Finally, we note that $\limsup_n f_n = \inf_k g_k$, where $g_k = \sup_{n \geqslant k} f_n$ and $\liminf_n f_n = \sup_k h_k$, where $h_k = \inf_{n \geqslant k} f_n$. The last conclusion follows since if $f_n(x)$ converges, then $\lim_n f_n(x) = \limsup_n f_n(x) = \liminf_n f_n(x)$.

The proofs in the real valued case follow, mutatis mutandis. □

We establish one further result concerning measurable functions, before moving on to discuss the interaction between measurable functions and measures.

> **Exercise B.6.** Suppose $(E, \mathcal{E})$ is a measurable space and f is a measurable function on E. Show that the functions $f^+, f^-, |f|$ defined by:
>
> $$f^+(x) := \max\{f(x), 0\}, \qquad f^-(x) := \max\{-f(x), 0\}, \qquad |f| := f^+ + f^-,$$
>
> are non-negative measurable functions.

Theorem B.16 (Monotone Class Theorem)**.** *Let $(E, \mathcal{E})$ be a measurable space and let $\mathcal{A}$ be a π-system generating $\mathcal{E}$. Suppose V is a vector space of bounded functions $f : E \to \mathbb{R}$ such that:*

i) $1 \in V$ *and* $\mathbb{1}_A \in V$ *for all* $A \in \mathcal{A}$*;*

ii) *if* $f_n \in V$ *for all* n *and* f *is a bounded function such that* $0 \leqslant f_n \leqslant f_{n+1}$ *and* $f_n \to f$ *pointwise, then* $f \in V$*.*

Then V contains every bounded measurable function.

Proof. Let $\mathcal{D} = \{A \in \mathcal{E} : \mathbb{1}_A \in V\}$. Then the assumptions on V ensure $\mathcal{D}$ is a d-system containing $\mathcal{A}$, so $\mathcal{D} = \mathcal{E}$. Since V is a vector space, it must contain all finite linear combinations of indicator functions of measurable sets. If f is a bounded non-negative measurable function, then $f_n = 2^{-n}\lfloor 2^n f \rfloor$ is such a function, and moreover f_n is an increasing sequence which tends to f pointwise, so $f \in V$. Since any bounded measurable function can be written as the difference of two bounded non-negative measurable functions we're done. □

We shall now see how measurable functions interact with measures. Firstly, we note that a measurable function can be use to induce a measure on its image, given a measure on its domain. Suppose $(E, \mathcal{E})$ and $(G, \mathcal{G})$ are measurable spaces, μ is a measure on $(E, \mathcal{E})$ and $f : E \to G$ is a measurable function. We can define a measure on $(G, \mathcal{G})$, $f_*\mu$, called the push-forward or image measure by:

$$f_*\mu(A) = \mu(f^{-1}(A)), \qquad \text{for all } A \in \mathcal{G}.$$

Next we consider convergence in the context of a measurable space $(E, \mathcal{E}, \mu)$. Given some property P conditioned on a point $x \in E$, we say that P holds *almost everywhere* in E if

$$\mu\left(\{x \in E : P(x) \text{ is false}\}\right) = 0.$$

For example, we can consider $\mathbb{R}$ equipped with the Lebesgue measure, and introduce the Dirichlet function $f(x) = 1$ for $x \in \mathbb{Q}$, $f(x) = 0$ otherwise. Then we can say '$f = 0$ almost everywhere'. In circumstances where the choice of measure is ambiguous, one sometimes writes μ-almost everywhere. We often abbreviate almost everywhere to *a.e.*

If $(f_n)_{n=1}^{\infty}$ is a sequence of measurable functions on $(E, \mathcal{E}, \mu)$, we say $f_n \to f$ almost everywhere if

$$\mu(\{x \in E : f_n(x) \not\to f(x)\}) = 0.$$

Another notion of convergence that we can consider is *convergence in measure.* We say that $f_n \to f$ in measure if

$$\mu\left(\{x \in E : |f_n(x) - f(x)| > \epsilon\}\right) \to 0, \qquad \text{for all } \epsilon > 0.$$

The connection between these two notions is captured by

Theorem B.17. *Suppose $(E, \mathcal{E}, \mu)$ is a measure space, and $(f_n)_{n=1}^{\infty}$ is a sequence of measurable functions on E. Then:*

i) Suppose $\mu(E) < \infty$, then if $f_n \to f$ almost everywhere, then $f_n \to f$ in measure.

ii) If $f_n \to f$ in measure, then there exists a subsequence $(f_{n_k})_{k=1}^{\infty}$ such that $f_{n_k} \to f$ almost everywhere.

Proof. i) By considering $f_n - f$, assume wlog $f_n \to 0$ a.e.. Fix $\epsilon > 0$, then for any n:

$$\mu\left(\bigcap_{m \geqslant n} \{|f_m| \leqslant \epsilon\}\right) \leqslant \mu(\{|f_n| \leqslant \epsilon\})$$

Now, set $A_n = \bigcap_{m \geqslant n}\{|f_m| \leqslant \epsilon\}$. We have $A_n \subset A_{n+1}$ and

$$x \in \bigcup_n A_n \quad \Longleftrightarrow \quad \text{there exists } N \text{ such that } |f_n(x)| \leqslant \epsilon \text{ for all } n \geqslant N.$$

Thus as $n \to \infty$, we have:

$$\mu(A_n) \geqslant \mu\left(\{x : f_n(x) \to 0\}\right) = \mu(E).$$

ii) Again, wlog suppose $f_n \to 0$ in measure. Set $n_1 = 1$. For each $k > 1$ we can find $n_k > n_{k+1}$ such that

$$\mu(\{|f_{n_k}| > 1/k\}) \leqslant 2^{-k}.$$

Now, let

$$A_k = \bigcup_{m \geqslant k} \{x \in E : |f_{n_m}(x)| > 1/m\}.$$

we have that $x \in \bigcap_k A_k$ if and only if for any k there exists $m \geqslant k$ such that $|f_{n_m}(x)| > 1/m$. Thus $x \notin \bigcap_k A_k$ if and only if there exists k such that for any $m \geqslant k$ we have $|f_{n_m}(x)| \leqslant 1/m$ and we conclude $f_{n_k} \to 0$ for all $x \notin \bigcap_k A_k$. Now, $A_{k+1} \subset A_k$ so, for any m:

$$\mu\left(\bigcap_k A_k\right) \leqslant \mu\left(A_m\right) = \mu\left(\bigcup_{m \geqslant k} \{|f_{n_m}(x)| > 1/m\}\right)$$
$$\leqslant \sum_{m \geqslant k} \mu(\{|f_{n_m}(x)| > 1/m\}) \leqslant 2^{m-1}$$

we conclude that $\mu\left(\bigcap_k A_k\right) = 0$ and thus $f_{n_k} \to 0$ a.e..

□

Exercise B.7. Let $E = [0, 1]$ be equipped with the Lebesgue measure. Construct a sequence of functions $f_n : [0, 1] \to [0, 1]$ such that $f_n \to f$ in measure, but $(f_n(x))_{n=1}^{\infty}$ does not converge for any $x \in [0, 1]$.

A final result concerns the measurability of a function which equals a measurable function almost everywhere.

Lemma B.18. *Let $(E, \mathcal{E}, \mu)$ be a* complete *measure space, and let f be a measurable function on E. If $g : E \to \mathbb{R}$ is such that $f = g$ almost everywhere, then g is measurable.*

Proof. Under the assumptions, $N = \{f \neq g\}$ is null, hence measurable by the completeness hypothesis, and $\mu(N) = 0$. Fix $a \in \mathbb{R}$. By assumption $A = \{f < a\}$ is measurable, and if we can show that $B = \{g < a\}$ is measurable then we will be done. Now, $B \cap A^c \subset N$, so by completeness $A \cap B^c$ is measurable, hence

$$B = A \cup (B \cap A^c)$$

is measurable. □

B.3 Integration

We now wish to define a notion of integration for measurable functions on some measure space $(E, \mathcal{E}, \mu)$. We approach this by first considering the case of non-negative measurable functions. These can be approximated from below by simple functions, which are finite linear combinations of characteristic functions on which the integral can be easily defined.

We say f is *simple* if

$$f = \sum_{n=1}^{k} \alpha_n \mathbb{1}_{A_n}$$

where $\alpha_n \in \mathbb{R}$ and $A_n \in \mathcal{E}$. For a non-negative simple function it is natural to define the integral as:

$$\mu(f) := \sum_{n=1}^{k} \alpha_n \mu(A_n)$$

Here, by convention $0 \cdot \infty = 0$. Alternative notations which we will make use of are:

$$\mu(f) = \int_E f d\mu = \int_E f(x) d\mu(x)$$

We note that α_n, A_n are not uniquely determined by f, however $\mu(f)$ is independent of the particular representation we choose.

Exercise B.8. a) Show that if $0 \leqslant \alpha_n, \beta_n < \infty$, $A_n, B_n \in \mathcal{E}$ satisfy

$$\sum_{n=1}^{k} \alpha_n \mathbb{1}_{A_n} = \sum_{n=1}^{l} \beta_n \mathbb{1}_{B_n},$$

then

$$\sum_{n=1}^{k} \alpha_n \mu(A_n) = \sum_{n=1}^{l} \beta_n \mu(B_n).$$

b) If f, g are simple functions and $a, b \geqslant 0$, show that

i) $\mu(af + bg) = a\mu(f) + b\mu(g)$.
ii) If $f \leqslant g$ then $\mu(f) \leqslant \mu(g)$.
iii) $f = 0$ a.e. if and only if $\mu(f) = 0$.

For a non-negative measurable function, we define the integral to be:

$$\mu(f) = \int_E f d\mu = \sup\{\mu(g) : g \text{ simple with } 0 \leqslant g \leqslant f\}.$$

By the results of Exercise B.8 this is consistent with the previous definition when f is simple. Note that $\mu(f)$ is permitted to take the value ∞. We also note that if f, g are non-negative measurable functions with $f \leqslant g$, then

$$\int_E f d\mu \leqslant \int_E g d\mu.$$

It also follows immediately from the definition that for any $\epsilon > 0$ there exists a simple function f_ϵ such that

$$\int_E |f - f_\epsilon| \, d\mu = \int_E (f - f_\epsilon) d\mu < \epsilon$$

To define the integral for functions which may take both positive and negative values, we first recall that if f is measurable then $f^+, f^-, |f|$ are non-negative measurable functions. We say that f is *integrable* if $\mu(|f|) < \infty$, in which case we define:

$$\mu(f) = \mu(f^+) - \mu(f^-).$$

Note that $f \leqslant g$ if and only if $f^+ \leqslant g^+$ and $f^- \geqslant g^-$, so that $f \leqslant g$ implies $\mu(f) \leqslant \mu(g)$. In particular, we have that $|\mu(f)| \leqslant \mu(|f|)$. By our comment above, for any $\epsilon > 0$ we can find a simple function f_ϵ such that

$$\int_E |f - f_\epsilon| \, d\mu < \epsilon$$

since we can approximate both f^+ and f^- by appropriate simple functions.

If at most one of $\mu(f^+)$ or $\mu(f^-)$ is infinite, then we can still define $\mu(f)$ by the same formula, but if both $\mu(f^+)$ and $\mu(f^-)$ are infinite then we can't sensibly assign a value to $\mu(f)$. We can also consider the case where f takes values in $\mathbb{R}^n$. In this case we pick a basis $(e_i)_{i=1}^n$ for $\mathbb{R}^n$ and write $f = \sum_{i=1}^n f_i e_i$. We say f is integrable if each f_i is integrable and we define:

$$\int_E f d\mu = \sum_{i=1}^n \left(\int_E f_i d\mu \right) e_i$$

This naturally gives a definition for functions taking complex values by the isomorphism $\mathbb{C} \simeq \mathbb{R}^2$.

B.3.1 Convergence theorems

A fundamental result in the Lebesgue theory of integration is the monotone convergence theorem, sometimes called Beppo Levi's Lemma. Suppose $(E, \mathcal{E}, \mu)$ is a measure space and that $(f_n)_{n=1}^{\infty}$ is a sequence of non-negative measurable functions which is increasing, i.e. $f_n(x) \leqslant f_{n+1}(x)$ for all $x \in E$ and $n \geqslant 1$. Then for each $x \in E$ the limit $f(x) = \lim_{n\to\infty} f_n(x)$ exists in $[0, \infty]$. We know that f is measurable, and the monotone convergence theorem asserts that $\mu(f) = \lim_{n\to\infty} \mu(f_n)$.

Theorem B.19 (Monotone convergence theorem)**.** *Let $(f_n)_{n=1}^{\infty}$ be an increasing sequence of non-negative integrable functions on a measure space $(E, \mathcal{E}, \mu)$ converging to f. Then*

$$\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu.$$

Proof. Let $M = \sup_n \mu(f_n)$. We wish to show that $M = \mu(f)$. Since f_n is an increasing sequence, we have $f_n \leqslant f$ so that $\mu(f_n) \leqslant \mu(f)$. As this holds for all n, we deduce:

$$M \leqslant \mu(f) = \sup\{\mu(g) : g \text{ simple}, \ g \leqslant f\}.$$

If we can show that for any simple function g with $0 \leqslant g \leqslant f$ we have $\mu(g) \leqslant M$ then we're done. Suppose

$$g = \sum_{i=1}^{m} a_k \mathbb{1}_{A_k}$$

is such a function, where we may assume $A_k \in \mathcal{E}$ are disjoint without loss of generality. We define

$$g_n(x) = \min\{g(x), 2^{-n} \lfloor 2^n f_n \rfloor\}.$$

Then $(g_n)_{n=1}^{\infty}$ is an increasing sequence of simple functions, satisfying $g_n \leqslant f_n \leqslant f$ and $g_n \to g$. Fix $0 < \epsilon < 1$. Define the sets $A_{k,n}$ by

$$A_{k,n} = \{x \in A_k : g_n(x) \geqslant (1 - \epsilon) a_k\}$$

Then since g_n is an increasing sequence, we have $A_{k,n} \subset A_{k,n+1}$. So by countable additivity we have $\mu(A_{k,n}) \to \mu(A_k)$ as $n \to \infty$. By construction we have

$$\mathbb{1}_{A_k} g_n \geqslant (1 - \epsilon) a_k \mathbb{1}_{A_{k,n}}$$

so

$$\mu(\mathbb{1}_{A_k} g_n) \geqslant (1 - \epsilon) a_k \mu(A_{k,n})$$

Now, noting that $g_n = \sum_{k=1}^{m} \mathbb{1}_{A_k} g_n$, and using the linearity result of Exercise B.8 we see

$$\mu(g_n) \geqslant (1 - \epsilon) \sum_{k=1}^{m} a_k \mu(A_{k,n}) \to (1 - \epsilon) \sum_{k=1}^{m} a_k \mu(A_k) = (1 - \epsilon)\mu(g).$$

Now, $\mu(g_n) \leqslant \mu(f_n) \leqslant M$, so we have $(1-\epsilon)\mu(g) \leqslant M$ for any $\epsilon > 0$, hence $\mu(g) \leqslant M$. □

A straightforward corollary of this result is the following:

Corollary B.20. *Suppose $(f_n)_{n=1}^{\infty}$ is a sequence of non-negative measurable functions on a measure space $(E, \mathcal{E}, \mu)$. Then*

$$\int_E \left(\sum_{n=1}^{\infty} f_n \right) d\mu = \sum_{n=1}^{\infty} \left(\int_E f_n d\mu \right).$$

Another useful corollary:

Corollary B.21. *Suppose $(f_n)_{n=1}^{\infty}$ is a* decreasing *sequence of bounded measurable functions on a measure space $(E, \mathcal{E}, \mu)$. Then*

$$\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu.$$

Proof. We take $g_n = f_1 - f_n$, since the f_n are bounded this is well defined (i.e we don't have to assign a value to $\infty - \infty$). Then $(g_n)_{n=1}^{\infty}$ is an increasing sequence and we can apply the usual monotone convergence theorem. □

> **Exercise($*$).** Give an example to show that Corollary B.21 fails if the boundedness assumption is dropped.

With the monotone convergence theorem in hand, we can readily show that the integral satisfies the properties we would expect.

Theorem B.22. *Suppose f, g are non-negative measurable functions on a measure space $(E, \mathcal{E}, \mu)$ and $a, b \geqslant 0$ are constants. Then:*

i) $\displaystyle\int_E (af + bg) d\mu = a \int_E f d\mu + b \int_E g d\mu$

ii) If $f \leqslant g$ then $\displaystyle\int_E f d\mu \leqslant \int_E g d\mu$

iii) $\displaystyle\int_E f d\mu = 0$ *if and only if $f = 0$ almost everywhere.*

Proof. Let

$$f_n(x) = \min\{2^n \lfloor 2^{-n} f(x) \rfloor, n\}, \qquad g_n(x) = \min\{2^n \lfloor 2^{-n} g(x) \rfloor, n\}.$$

Then $(f_n)_{n=1}^{\infty}, (g_n)_{n=1}^{\infty}$ is an increasing sequence of non-negative simple functions tending to f, g respectively, and clearly $(af_n + bg_n)_{n=1}^{\infty}$ is an increasing sequence tending to $af + bg$. Since these are simple functions, we have:

$$\int_E (af_n + bg_n) d\mu = a \int_E f_n d\mu + b \int_E g_n d\mu$$

and by the monotone convergence theorem, we can take the limit $n \to \infty$ to establish *i*). Point *ii*) we already noted follows directly from the definition of the integral.

Finally, if $f = 0$ almost everywhere, then we have $\mu(g) = 0$ for any simple $g \leqslant f$, and thus $\mu(f) = 0$. Now suppose $f(x) \neq 0$ almost everywhere. Then there exists $\epsilon > 0$ such that if $A = \{f \geqslant \epsilon\}$ then $\mu(A) > 0$. Then $g = \mathbb{1}_A \epsilon$ is a non-negative simple function with $g \leqslant f$ and $\mu(g) = \epsilon\mu(A) > 0$, hence $\mu(f) > 0$. □

We can extend this result to functions taking values in $\mathbb{R}$ as follows:

Theorem B.23. *Suppose f, g are integrable functions on a measure space $(E, \mathcal{E}, \mu)$ and $a, b \in \mathbb{R}$ are constants. Then:*

i) $\displaystyle\int_E (af + bg)d\mu = a\int_E fd\mu + b\int_E gd\mu$

ii) If $f \leqslant g$ then $\displaystyle\int_E fd\mu \leqslant \int_E gd\mu$

iii) If $f = 0$ almost everywhere then $\displaystyle\int_E fd\mu = 0.$

Proof. Note that it follows immediately from the definition that $\mu(-f) = -\mu(f)$. Suppose then that $a > 0$. We have:

$$\mu(af) = \mu(af^+) - \mu(af^-) = a\mu(f^+) - a\mu(f^-) = a\mu(f).$$

We also note that $(f + g)^+ - (f + g)^- = f + g = f^+ + g^+ - f^- - g^-$. As a consequence $(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+$, where both sides are sums of non-negative measurable functions, hence:

$$\mu((f + g)^+) + \mu(f^-) + \mu(g^-) = \mu((f + g)^-) + \mu(f^+) + \mu(g^+)$$

and on rearranging:

$$\mu(f + g) = \mu((f + g)^+) - \mu((f + g)^-) = \mu(f^+) - \mu(f^-) + \mu(g^+) - \mu(g^-) = \mu(f) + \mu(g).$$

Combining our observations gives i). Noting that $f \leqslant g$ implies $0 \leqslant g - f$, we deduce $0 \leqslant \mu(g) - \mu(f)$ and thus ii) holds. Finally, if $f = 0$ almost everywhere, then $f^+, f^- = 0$ almost everywhere thus $\mu(f) = 0$. □

Suppose $(E, \mathcal{E}, \mu)$ is a measure space. If $A \in \mathcal{E}$ and f is integrable, then so is $f\mathbb{1}_A$. Recall also that A inherits a measure space structure in a natural way $(A, \mathcal{E}|_A, \mu|_A)$. It is relatively straightforward to see that $f|_A$ is integrable, and that we can define unambiguously

$$\int_A fd\mu := \int_E f\mathbb{1}_A d\mu = \int_A f|_A d\mu|_A,$$

By our linearity result i) above, if A, B are disjoint measurable sets, then

$$\int_A fd\mu + \int_B fd\mu = \int_{A\cup B} fd\mu.$$

We also note that by ii) we have that if $|f| \leqslant K$ almost everywhere and $\mu(E) < \infty$, then f is integrable and

$$\left|\int_E fd\mu\right| \leqslant K\mu(E).$$

A useful consequence of the monotone convergence theorem connects the Lebesgue integral to the Riemann integral.

Theorem B.24. *Let $A = (a_1, b_1] \times \cdots \times (a_n, b_n]$ be a rectangle in $\mathbb{R}^n$, and suppose $f : A \to \mathbb{R}$ is bounded. Then f is Riemann integrable*[4] *if and only if f is continuous almost everywhere. If so, f is integrable with respect to Lebesgue measure, μ, on A and moreover*

$$\mathscr{R}\int_A f(x)dx = \int_A f d\mu,$$

where $\mathscr{R}\int$ denotes the Riemann integral.

Proof. Since f is bounded, we may assume that $0 \leqslant f \leqslant K$ for some K, without loss of generality. We consider a sequence of partitions $\mathcal{P}_n$ of A such that $\mathcal{P}_{n+1}$ is a refinement of $\mathcal{P}_n$ and the mesh of $\mathcal{P}_n \to 0$. Correspondingly, we construct two sequences of functions, $\underline{f}_n, \overline{f}_n$ by

$$\underline{f}_n = \sum_{\pi \in \mathcal{P}_n} \inf_\pi f \mathbb{1}_\pi, \quad \overline{f}_n = \sum_{\pi \in \mathcal{P}_n} \sup_\pi f \mathbb{1}_\pi$$

which satisfy

$$0 \leqslant \underline{f}_n \leqslant \underline{f}_{n+1} \leqslant f \leqslant \overline{f}_{n+1} \leqslant \overline{f}_n \leqslant K.$$

Since each $\pi \in \mathcal{P}_n$ is a rectangle, it is certainly Lebesgue measurable and so $\underline{f}_n, \overline{f}_n$ are in fact simple functions. Moreover,

$$\int_A \underline{f}_n d\mu = L(f, \mathcal{P}_n), \qquad \int_A \overline{f}_n d\mu = U(f, \mathcal{P}_n),$$

where U, L are the usual upper and lower sums associated to a partition. The function f is Riemann integrable if and only if $L(f, \mathcal{P}_n), U(f, \mathcal{P}_n)$ have a common limit as $n \to \infty$, i.e:

$$U(f, \mathcal{P}_n) \to \mathscr{R}\int_A f(x)dx, \qquad L(f, \mathcal{P}_n) \to \mathscr{R}\int_A f(x)dx, \qquad \text{as } n \to \infty.$$

$(\underline{f}_n)_{n=1}^\infty$ is a monotone increasing sequence, bounded above by f, so there exists a bounded measurable function $\underline{f} \leqslant f$ such that $\underline{f} = \lim_{n\to\infty} \underline{f}_n = \sup \underline{f}_n$. Similarly, there exists a bounded measurable function $\overline{f} \geqslant f \geqslant \underline{f}$ such that $\overline{f} = \lim_{n\to\infty} \overline{f}_n = \inf \underline{f}_n$. By applying monotone convergence to $(\underline{f}_n)_{n=1}^\infty$ and $(\overline{f}_n)_{n=1}^\infty$ we have:

$$\lim_{n\to\infty} \int_A \underline{f}_n d\mu = \int_A \underline{f} d\mu \leqslant \int_A \overline{f} d\mu \leqslant \lim_{n\to\infty} \int_A \overline{f}_n d\mu$$

We deduce that f is Riemann integrable if and only if

$$\int_A \underline{f} d\mu = \int_A \overline{f} d\mu = \mathscr{R}\int_A f(x)dx.$$

This occurs if and only if $\underline{f} = \overline{f}$ almost everywhere.

We define the set of boundary points of $\mathcal{P}_n$ to be:

$$B_n = \bigcup_{\pi \in \mathcal{P}_n} \partial\pi \cap A$$

[4] For a discussion of the Riemann integral in $\mathbb{R}^n$ see Spivak: "Calculus on maniolds".

Clearly $\mu(B_n) = 0$, so the set $B = \cup_n B_n$ also has measure zero. Suppose $x \notin B$, then $\underline{f}(x) = \overline{f}(x)$ if and only if f is continuous at x. We conclude that f is Riemann integrable if and only if f is continuous almost everywhere. Since $\underline{f} \leqslant f \leqslant \overline{f}$ in this case, we deduce that f is almost everywhere equal to $\underline{f}$ and the result follows by Lemma B.18. □

This result, that Riemann integrability is equivalent to almost-everywhere continuity, is known as *Lebesgue's criterion for integrability.* In practice, many of the explicit integrals we encounter are Riemann integrals, and this gives us access to the standard toolkit to compute them. Where there's no possibility for ambiguity, we will often use the standard notation $\int dx$ or $\int d^n x$, etc. to denote Lebesgue integration.

The next convergence result for integrals we shall require allows us to drop the assumption that our sequence is monotone, but at the cost of a weakened result.

Lemma B.25 (Fatou's Lemma)**.** *Suppose $(f_n)_{n=1}^{\infty}$ is a sequence of non-negative measurable functions on a measure space $(E, \mathcal{E}, \mu)$. Then*

$$\int_E \liminf_{n\to\infty} f_n d\mu \leqslant \liminf_{n\to\infty} \int_E f_n d\mu$$

Proof. Let $g_n = \inf_{m \geqslant n} f_m$. Then $(g_n)_{n=1}^{\infty}$ is an increasing sequence of non-negative measurable functions, which tends to $\liminf f_n$. Thus by monotone convergence

$$\int_E g_n d\mu \to \int_E \liminf_{n\to\infty} f_n d\mu.$$

On the other hand, for $k \geqslant n$ we have:

$$g_n \leqslant f_k,$$

hence

$$\int_E g_n d\mu \leqslant \int_E f_k d\mu \text{ for all } k \geqslant n \quad \Longrightarrow \quad \int_E g_n d\mu \leqslant \inf_{k \geqslant n} \int_E f_k d\mu.$$

Now, as $n \to \infty$

$$\inf_{k \geqslant n} \int_E f_k d\mu \to \liminf_{n\to\infty} \int_E f_n d\mu,$$

and we're done. □

> **Exercise(∗).** Construct a sequence $(f_n)_{n=1}^{\infty}$ of functions $f_n : [0,1] \to [0,\infty)$ satisfying the hypotheses of Fatou's Lemma such that the inequality is strict.

The next convergence result we shall establish is an especially useful one, and in particular will be invoked on many occasions during the course.

Theorem B.26 (The Dominated Convergence Theorem)**.** *Suppose that $(E, \mathcal{E}, \mu)$ is a measure space and that $(f_n)_{n=1}^{\infty}$ is a sequence of measurable functions such that:*

i) There exists an integrable function g such that $|f_n| \leqslant g$.

ii) $f_n(x) \to f(x)$ for all x.

Then f is integrable and:

$$\int_E f_n d\mu \to \int_E f d\mu.$$

Proof. By Theorem B.15, f is measurable, and since $|f| \leqslant g$, we further have that f is integrable. We have

$$0 \leqslant g \pm f_n \to g \pm f,$$

so that $\liminf g \pm f_n = g \pm f$. By Fatou and properties of $\liminf, \limsup$ we have:

$$\int_E g d\mu + \int_E f d\mu = \int_E \liminf (g + f_n) d\mu \leqslant \liminf \int_E (g + f_n) d\mu = \int_E g d\mu + \liminf \int_E f_n d\mu$$
$$\int_E g d\mu - \int_E f d\mu = \int_E \liminf (g - f_n) d\mu \leqslant \liminf \int_E (g - f_n) d\mu = \int_E g d\mu - \limsup \int_E f_n d\mu$$

Rearranging, we have:

$$\int_E f d\mu \leqslant \liminf \int_E f_n d\mu \leqslant \limsup \int_E f_n d\mu \leqslant \int_E f d\mu,$$

hence

$$\liminf \int_E f_n d\mu = \limsup \int_E f_n d\mu = \int_E f d\mu,$$

and we're done. □

We note that the hypotheses can be weakened slightly: suppose the hypotheses hold almost everywhere, so that $X = \{x \in E : \text{ hypotheses fail}\}$ has measure zero, then by applying the Dominated Convergence Theorem to $f_n \mathbb{1}_{X^c}$, we can recover the same result.

Exercise B.9. Here μ is the Lebesgue measure on $\mathbb{R}$.

a) Show that $f : [0,1] \to \mathbb{R}$ given by $f(x) = \frac{1}{\sqrt{x}}$ is Lebesgue integrable, and that

$$\int_{[0,1]} f d\mu = \lim_{\epsilon \to 0} \mathscr{R} \int_\epsilon^1 \frac{1}{\sqrt{x}} dx.$$

b) Suppose $f : [0,1] \to \mathbb{R}$ is Riemann integrable on every interval $[\epsilon, 1]$, $\epsilon > 0$ and moreover

$$\mathscr{R} \int_\epsilon^1 |f(x)|\, dx \leqslant C$$

for some C independent of ϵ. Show that f is Lebesgue integrable with

$$\int_{[0,1]} f d\mu = \lim_{\epsilon \to 0} \mathscr{R} \int_\epsilon^1 f(x) dx.$$

c) Suppose $f : \mathbb{R} \to \mathbb{R}$ is Riemann integrable on every interval $[-R, R]$ and moreover

$$\mathscr{R} \int_{-R}^R |f(x)|\, dx \leqslant C$$

for some C independent of R. Show that f is Lebesgue integrable with

$$\int_{\mathbb{R}} f d\mu = \lim_{R\to\infty} \mathscr{R} \int_{-R}^{R} f(x)dx.$$

Give an example of a function such that

$$\lim_{R\to\infty} \mathscr{R} \int_{-R}^{R} f(x)dx$$

exists, but $f : \mathbb{R} \to \mathbb{R}$ is *not* Lebesgue integrable.

B.3.2 Product measures and Tonelli–Fubini

Given two measure spaces $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$, we wish to construct a measure space on $E \times F$. We say a subset $E \times F$ is a rectangle if it is of the form $A \times B$, with $A \in \mathcal{E}$, $B \in \mathcal{F}$. We denote by $\mathcal{E} \boxtimes \mathcal{F}$ the collection of finite disjoint unions of rectangles. Note that if $A_i \in \mathcal{E}$, $B_i \in \mathcal{F}$ then

$$\begin{aligned}(A_1 \times B_1) \cap (A_2 \times B_2) &= (A_1 \cap A_2) \times (B_1 \cap B_2),\\ (A_1 \times B_1) \cup (A_2 \times B_2) &= (A_1 \times B_1 \setminus B_2) \cup ((A_1 \cup A_2) \times (B_1 \cap B_2)) \cup (A_2 \times B_2 \setminus B_1)\end{aligned}$$

where the right-hand side of the second line is a disjoint union of rectangles. Finally, since

$$(A_1 \times B_1)^c = (E \times B_1^c) \cup (A_1^c \times F),$$

we see that $\mathcal{E} \boxtimes \mathcal{F}$ is an algebra (hence a ring). We denote by $\mathcal{E} \otimes \mathcal{F}$ the σ-algebra generated by $\mathcal{E} \boxtimes \mathcal{F}$. We define a set function $\pi : \mathcal{E} \boxtimes \mathcal{F} \to [0, \infty]$ by

$$\pi\left(\bigcup_{i=1}^{N} (A_i \times B_i)\right) = \sum_{i=1}^{N} \mu(A_i)\nu(B_i),$$

where the rectangles $A_i \times B_i \in \mathcal{E} \times \mathcal{F}$, $i = 1, \ldots, N$, are assumed to be disjoint. Now suppose that $(A_j \times B_j)_{j=1}^{\infty}$ is a sequence of disjoint rectangles such that

$$\bigcup_{j=1}^{\infty} A_j \times B_j = A \times B \in \mathcal{E} \times \mathcal{F}.$$

We claim that

$$\mu(A)\mu(B) = \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j).$$

To see this, we note:

$$\mathbb{1}_A(x)\mathbb{1}_B(y) = \mathbb{1}_{A\times B}(x, y) = \sum_{j=1}^{\infty} \mathbb{1}_{A_j\times B_j}(x, y) = \sum_{j=1}^{\infty} \mathbb{1}_{A_j}(x)\mathbb{1}_{B_j}(y)$$

Integrating with respect to x, using Corollary B.20 we see

$$\mu(A)\mathbb{1}_B(y) = \sum_{j=1}^{\infty} \mathbb{1}_{B_j}(y) \int \mathbb{1}_{A_j} d\mu = \sum_{j=1}^{\infty} \mathbb{1}_{B_j}(y)\mu(A_j)$$

Integrating again with respect to y, by the same argument we find:

$$\mu(A)\mu(B) = \sum_{j=1}^{\infty} \mu(A_j)\nu(B_j).$$

Note that this immediately implies that $\pi(C)$ is well defined for $C \in \mathcal{E} \boxtimes \mathcal{F}$, independent of how C is represented as a finite union of rectangles. We also note that $\mathcal{E} \boxtimes \mathcal{F}, \pi$ satisfy the conditions of Carathéodory's theorem, Theorem B.3, thus we can define an outer measure π^* on $E \times F$, whose restriction to $\mathcal{E} \otimes \mathcal{F}$ gives a measure, which agrees with π on $\mathcal{E} \boxtimes \mathcal{F}$. We call this measure on $\mathcal{E} \otimes \mathcal{F}$ the product measure, $\mu \times \nu$.

Note that the product measure $\mu \times \nu$ will not in general be unique. However, it will be if $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ are σ-finite. We say $(E, \mathcal{E}, \mu)$ is σ-finite if there exists a countable collection $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$ of disjoint measurable sets, with $\mu(A_i) < \infty$, such that $E = \cup A_i$. If both $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ are σ-finite, then $\mathcal{E} \boxtimes \mathcal{F}$ satisfies the conditions to enable us to apply Corollary B.9 to deduce that $\mu \times \nu$ is the unique measure on $\mathcal{E} \otimes \mathcal{F}$ such that

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B).$$

A brief note of caution before we consider integration on product spaces. If E, F are topological spaces and $\mathcal{E}, \mathcal{F}$ are the Borel σ-algebras on their respective spaces, then $\mathcal{E} \otimes \mathcal{F}$ contains the Borel σ-algebra of $E \times F$ with the product topology. However, the two need not be equal in general. One important case where we do have equality is when E, F are σ-compact metric spaces[5]. In particular this is the case when $E = \mathbb{R}^n$, $F = \mathbb{R}^m$. By the uniqueness of Lebesgue measure, we have that the product measure restricted to $\mathcal{B}(\mathbb{R}^n) \times \mathcal{B}(\mathbb{R}^m) = \mathcal{B}(\mathbb{R}^n \times \mathbb{R}^m)$ is the Lebesgue measure on $\mathbb{R}^{n+m}$.

We now wish to consider integration of a measurable function defined on $E \times F$. If $f : E \times F \to \mathbb{R}$, and $x \in E, y \in F$, we define the $x-$section, f_x and y-section, f^y as:

$$f_x(y) = f^y(x) = f(x, y).$$

We also introduce the x-section and y-section of a set $A \subset E \times F$ as:

$$A_x = \{y \in F : (x, y) \in A\}, \qquad A^y = \{x \in F : (x, y) \in A\}.$$

Note that $A_x \subset F$, $A^y \subset E$ and we have

$$(\mathbb{1}_A)_x = \mathbb{1}_{A_x}, \qquad (\mathbb{1}_A)^y = \mathbb{1}_{A^y},$$

Lemma B.27. *If $A \in \mathcal{E} \otimes \mathcal{F}$, then $A_x \in \mathcal{F}$ for all $x \in E$, $A^y \in \mathcal{E}$ for all $y \in F$. More generally, if f is $\mathcal{E} \otimes \mathcal{F}$-measurable, then f_x is $\mathcal{F}$-measurable and f^y is $\mathcal{E}$-measurable.*

[5] A topological space is σ-compact if it is the union of countably many compact sets.

Proof. Let

$$\mathcal{C} = \{A \subset E \times F : A_x \in \mathcal{F} \text{ for all } x \in E, A^y \in \mathcal{E} \text{ for all } y \in F\}$$

Certainly every measurable rectangle is in $\mathcal{C}$. We also have:

$$\left(\bigcup_{j=1}^{\infty} A_j\right)_x = \bigcup_{j=1}^{\infty} (A_j)_x, \qquad (A_x)^c = (A^c)_x,$$

and similarly for A^y, so that $\mathcal{C}$ is a σ-algebra and thus $\mathcal{E} \otimes \mathcal{F} \subset \mathcal{C}$. For the final part we note that for

$$(f_x)^{-1}(S) = \left(f^{-1}(S)\right)_x, \qquad (f^y)^{-1}(S) = \left(f^{-1}(S)\right)^y,$$

whence the result follows. □

Next we prove a special case of the Tonelli and Fubini theorems, where we restrict attention to the characteristic functions of a measurable set/

Lemma B.28. *Suppose μ, ν are finite, and let $A \in \mathcal{E} \otimes \mathcal{F}$. Then*

$$x \mapsto \nu(A_x), \qquad y \mapsto \mu(A^y)$$

are measurable functions, and

$$(\mu \times \nu)(A) = \int_E \nu(A_x) d\mu(x) = \int_E \mu(A^y) d\nu(y).$$

Proof. Let $\mathcal{C}$ consist of all sets $A \in \mathcal{E} \otimes \mathcal{F}$ for which the conclusion of the Lemma holds. Clearly $\mathcal{C}$ contains all rectangles, and these form a π-system. If we can show that $\mathcal{C}$ is a d-system, then we will be done by Lemma B.2.

Clearly $E \times F \in \mathcal{C}$. Suppose $A, B \in \mathcal{C}$ with $B \subset A$. Then$(A \setminus B)_x = A_x \setminus B_x$, so[6] $\nu((A \setminus B)_x) = \nu(A_x) - \nu(B_x)$, hence $x \mapsto \nu((A \setminus B)_x)$ is measurable, and

$$\begin{aligned}(\mu \times \nu)(A \setminus B) = (\mu \times \nu)(A) - (\mu \times \nu)(B) &= \int_E \nu(A_x) d\mu(x) - \int_E \nu(B_x) d\mu(x) \\ &= \int_E \nu((A \setminus B)_x) d\mu(x)\end{aligned}$$

A similar argument for $(A \setminus B)^y$ shows $A \setminus B \in \mathcal{C}$.

Now suppose $A_n \in \mathcal{C}$ with $A_n \subset A_{n+1}$ and let $A = \cup_n A_n$. Then by countable additivity we have $(x \mapsto \nu((A_n)_x))_{n=1}^{\infty}$ is a monotone increasing sequence of functions with limit $\nu(A_x)$. By monotone convergence we have $\nu(A_x)$ is measurable, with

$$\int_E \nu(A_x) d\mu(x) = \lim_{n\to\infty} \int_E \nu((A_n)_x) d\mu(x) = \lim_{n\to\infty} (\mu \times \nu)(A_n) = \mu \times \nu(A),$$

where in the final inequality we use countable additivity for $\mu \times \nu$. A similar argument for $\mu(A^y)$ establishes that $A \in \mathcal{C}$ and we're done. The extension to the case where μ, ν are assumed σ-finite is straightforward, and left as an exercise. □

[6] This is where the assumption that ν is finite is required

Exercise($*$). Show that Lemma B.28 holds if μ, ν are only assumed to be σ-finite.

We now prove two very closely related results, Tonelli's Theorem and Fubini's Theorem. They are often referred to together as the Tonelli-Fubini theorem.

Theorem B.29 (Tonelli)**.** *Assume $(E, \mathcal{E}, \mu)$, $(F, \mathcal{F}, \nu)$ are σ-finite measure spaces. Suppose $f : E \times F \to [0, \infty]$ is a non-negative measurable function. Then so are f_x, f^y, and setting:*

$$h(x) = \int_F f_x(y) d\nu(y), \qquad g(y) = \int_E f^y(x) d\mu(x),$$

we have that $h : E \to [0, \infty]$, $g : F \to [0, \infty]$ are non-negative measurable functions on their respective domains with:

$$\int_{E\times F} f d(\mu \times \nu) = \int_E h d\mu = \int_F g d\nu. \tag{B.2}$$

Proof. Take $(f_n)_{n=1}^\infty$ a monotone increasing sequence of non-negative simple functions with $f_n \to f$. Letting

$$h_n(x) = \int_F (f_n)_x(y) d\nu(y), \qquad g_n(y) = \int_E (f_n)^y(x) d\mu(x),$$

we have

$$\int_{E\times F} f_n d(\mu \times \nu) = \int_E h_n d\mu = \int_F g_n d\nu, \tag{B.3}$$

by the previous Lemma and the linearity of the integral. For each $x \in E$, we have that $((f_n)_x)_{n=1}^\infty$ is a monotone increasing sequence with $(f_n)_x \to f_x$ and similarly for $((f_n)^y)_{n=1}^\infty$. Thus, we have $(h_n)_{n=1}^\infty$ is an increasing sequence of functions with $h_n \to h$, and similarly for g_n by the monotone convergence theorem. Thus we can pass to the limit in (B.3) by the monotone convergence theorem to obtain (B.2). □

Theorem B.30 (Fubini)**.** *Assume $(E, \mathcal{E}, \mu)$, $(F, \mathcal{F}, \nu)$ are σ-finite measure spaces. Suppose $f : E \times F \to \mathbb{R}$ is an integrable function. Then $f_x : F \to \mathbb{R}$ is integrable for μ-almost every $x \in E$, as is $f^y : E \to \mathbb{R}$ for ν-almost every y. Thus*

$$h(x) = \int_F f_x(y) d\nu(y), \qquad g(y) = \int_E f^y(x) d\mu(x), \tag{B.4}$$

are defined almost everywhere. We have that $h : E \to \mathbb{R}$, $g : F \to \mathbb{R}$ are integrable functions on their respective domains, and:

$$\int_{E\times F} f d(\mu \times \nu) = \int_E h d\mu = \int_F g d\nu. \tag{B.5}$$

Proof. Write $f = f^+ - f^-$, with $f^\pm$ non-negative and integrable. By Tonelli applied to $f^\pm$ we find $h^\pm$, $g^\pm$ such that

$$\int_{E\times F} f^\pm d(\mu \times \nu) = \int_E h^\pm d\mu = \int_F g^\pm d\nu.$$

The first integral is finite by assumption, so we must have that $h^\pm, g^\pm$ are finite almost everywhere, and moreover are integrable. Thus $h = h^+ - h^-$ and $g = g^+ - g^-$ satisfy (B.4) and we also deduce (B.5). $\square$

In the particular case where $E = \mathbb{R}^n$, $F = \mathbb{R}^m$ equipped with their Borel sets and Lebesgue measure, then we conclude that if $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ with

$$\int_{\mathbb{R}^{n+m}} |f(x,y)|\, dxdy < \infty$$

then

$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} f(x,y)dy \right) dx = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} f(x,y)dx \right) dy = \int_{\mathbb{R}^{n+m}} f(z)dz$$

with obvious notation.

In combination, Tonelli–Fubini together with the Dominated Convergence Theorem are very powerful, and typically suffice for the majority of convergence related results that we require in standard analysis.

Exercise B.10. Let $(f_n)_{n=1}^\infty$ be a sequence of measurable functions $f_n : \mathbb{R}^m \to \mathbb{R}$ such that

$$\sum_{n=1}^\infty \int_{\mathbb{R}^m} |f_n|\, dx < \infty.$$

Show that:

$$f(x) = \sum_{n=1}^\infty f_n(x)$$

converges for a.e. $x \in \mathbb{R}^m$, and

$$\int_{\mathbb{R}^m} f dx = \sum_{n=1}^\infty \int_{\mathbb{R}^m} f_n dx.$$

B.4 The L^p-spaces

Given a measure space $(E, \mathcal{E}, \mu)$, we say that a measurable complex-valued[7] function f belongs to $\mathcal{L}^p(E,\mu)$ for some $p \leqslant 1 < \infty$ if

$$\|f\|_{L^p} := \left(\int_E |f|^p \, d\mu \right)^{\frac{1}{p}} = (\mu(|f|^p))^{\frac{1}{p}} < \infty.$$

We say that $f \in \mathcal{L}^\infty(E,\mu)$ if f is bounded almost everywhere, that is there exists $0 \leqslant K < \infty$ such that

$$\mu(\{|f(x)| > K\}) = 0.$$

If so, then we define

$$\|f\|_{L^\infty} := \inf\{K : \mu(\{|f(x)| > K\}) = 0\}.$$

We can show:

[7] We can also assume f is real-valued

Lemma B.31. *For* $1 \leqslant p \leqslant \infty$*, the function* $f \mapsto \|f\|_{L^p}$ *defines a* seminorm *on* $\mathcal{L}^\infty(E, \mu)$*. That is:*

i) $\|\cdot\|_{L^p}$ *is non-negative:*

$$\|f\|_{L^p} \geqslant 0, \qquad \text{for all } f \in \mathcal{L}^p(E, \mu).$$

ii) $\|\cdot\|_{L^p}$ *is homogeneous:*

$$\|\lambda f\|_{L^p} = |\lambda| \, \|f\|_{L^p}, \qquad \text{for all } f \in \mathcal{L}^p(E, \mu), \lambda \in \mathbb{C}.$$

iii) $\|\cdot\|_{L^p}$ *satisfies the triangle inequality:*

$$\|f + g\|_{L^p} \leqslant \|f\|_{L^p} + \|g\|_{L^p}, \qquad \text{for all } f, g \in \mathcal{L}^p(E, \mu).$$

Proof. See Exercise B.11. □

We note that the crucial property that is missing and prevents $\mathcal{L}^p(E, \mu)$ from being a normed space is *positivity*, i.e. that $\|f\|_{L^p} = 0$ if and only if $f = 0$. By Theorem B.22, we know that $\|f\|_{L^p} = 0$ if and only if $f = 0$ holds *almost everywhere*. In order to construct a normed space, we must quotient out the elements of $\mathcal{L}^p(E, \mu)$ which satisfy $\|f\|_{L^p} = 0$. To do this, we introduce an equivalence relation according to:

$$f \sim g \qquad \Longleftrightarrow \qquad f - g = 0 \text{ a.e.}$$

It is straightforward to see that $\sim$ defines an equivalence relation on $\mathcal{L}^p(E, \mu)$ and moreover, by the reverse triangle inequality

$$f \sim g \qquad \Longrightarrow \qquad \|f\|_{L^p} = \|g\|_{L^p}.$$

Thus we can define a new space

$$L^p(E, \mu) = \mathcal{L}^p(E, \mu)/\sim,$$

and $\|\cdot\|_{L^p}$ descends to a norm on the quotient space by:

$$\|[f]_\sim\|_{L^p} := \|f\|_{L^p}.$$

In practice, we usually elide the distinction between the function $f \in \mathcal{L}^p(E, \mu)$ and the equivalence class of functions $[f]_\sim \in L^p(E, \mu)$, so it is standard to speak of a function f belonging to $L^p(E, \mu)$. One should always remember, however, that in general statements about elements of $L^p(E, \mu)$ hold at most almost everywhere. It is immediate that we have

Lemma B.32. *The space* $L^p(E, \mu)$*, equipped with the norm* $\|\cdot\|_{L^p}$*, is a normed vector space.*

In the case where $E = \mathbb{R}^n$ equipped with the σ-algebra of Lebesgue measurable sets and the Lebesgue measure, we typically write $L^p(\mathbb{R}^n)$ instead of $L^p(\mathbb{R}^n, dx)$ to denote the associated spaces.

Exercise B.11. Let $(E, \mathcal{E}, \mu)$ be a measure space. Show that $\|\cdot\|_{L^p}$ defines a seminorm on $\mathcal{L}^p(E, \mu)$ for $1 \leqslant p \leqslant \infty$:

a) First check that the homogeneity and non-negativity properties are satisfied.

b) Establish the triangle inequality for the special cases $p = 1, \infty$.

c) Next prove Young's inequality: if $a, b \in \mathbb{R}_+$ and $p, q > 1$ with $p^{-1} + q^{-1} = 1$ then:

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}$$

Hint: set $t = p^{-1}$, *consider the function* $\log\left[ta^p + (1-t)b^q\right]$ *and use the concavity of the logarithm*

d) With[8] $p, q \geqslant 1$ such that $p^{-1} + q^{-1} = 1$, show that if $\|f\|_{L^p} = 1$ and $\|g\|_{L^q} = 1$ then

$$\int_E |fg| \, d\mu \leqslant 1$$

Deduce Hölder's inequality:

$$\int_E |fg| \, d\mu \ \leqslant \|f\|_{L^p} \|g\|_{L^q}, \qquad \text{for all } f \in \mathcal{L}^p(E, \mu), \ \ y \in \mathcal{L}^q(E, \mu).$$

e) Show that if $f, g \in \mathcal{L}^p(E, \mu)$

$$\|f + g\|_{L^p}^p \leqslant \int_E |f| \, |f + g|^{p-1} \, d\mu + \int_E |g| \, |f + g|^{p-1} \, d\mu$$

Apply Hölder's inequality to deduce:

$$\|f + g\|_{L^p}^p \leqslant \left(\|f\|_{L^p} + \|g\|_{L^p}\right) \|f + g\|_{L^p}^{p-1}$$

and conclude

$$\|f + g\|_{L^p} \leqslant \|f\|_{L^p} + \|g\|_{L^p}.$$

This is Minkowski's inequality.

Exercise B.12. a) Suppose that $\mu(E) < \infty$. Show that if $f \in \mathcal{L}^p(E, \mu)$, then $f \in \mathcal{L}^q(E, \mu)$ for any $1 \leqslant q < p$, with

$$\|f\|_{L^q} \leqslant \mu(E)^{\frac{p-q}{qp}} \|f\|_{L^p}.$$

b) Suppose that $f \in \mathcal{L}^{p_0}(E, \mu) \cap \mathcal{L}^{p_1}(E, \mu)$ with $p_0 < p_1 \leqslant \infty$. For $0 \leqslant \theta \leqslant 1$, define p_θ by

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}.$$

Show that $f \in \mathcal{L}^{p_\theta}(E, \mu)$ with

$$\|f\|_{L^{p_\theta}} \leqslant \|f\|_{L^{p_0}}^{1-\theta} \|f\|_{L^{p_1}}^{\theta}.$$

[8] We permit p, q to take the value ∞ with the convention $\infty^{-1} = 0$

B.4.1 Completeness

The most important property of the L^p-spaces is that in fact, they are complete, i.e. they are Banach spaces. To establish this, we first prove the following result.

Lemma B.33. *Suppose $(E, \mathcal{E}, \mu)$ is a measure space and $1 \leqslant p < \infty$. Let $(g_n)_{n=1}^{\infty}$ be a sequence with $g_n \in L^p(E, \mu)$ such that*

$$\sum_{n=1}^{\infty} \|g_n\|_{L^p} < \infty$$

then there exists $f \in L^p(E, \mu)$ such that

$$\sum_{n=1}^{\infty} g_n = f,$$

where the sum converges pointwise almost everywhere, and in $L^p(E, \mu)$.

Proof. Fix representative[9] functions $\tilde{g}_n \in \mathcal{L}^p(E, \mu)$ corresponding to each $g_n \in L^p(E, \mu)$. Define $h_n, h : E \to [0, \infty]$ by

$$h_n = \sum_{k=1}^{n} |\tilde{g}_k|, \qquad h = \sum_{k=1}^{\infty} |\tilde{g}_k|.$$

Note that $(h_n)_{n=1}^{\infty}$ is a monotone increasing sequence of non-negative measurable functions, converging pointwise to h, so by the monotone convergence theorem we have

$$\int_E h^p d\mu = \lim_{n \to \infty} \int_E h_n^p d\mu.$$

By Minkowski's inequality we see

$$\|h_n\|_{L^p} \leqslant \sum_{k=1}^{n} \|g_k\|_{L^p} \leqslant K =: \sum_{k=1}^{\infty} \|g_k\|_{L^p}.$$

It follows that $h \in \mathcal{L}^p(E, \mu)$ with $\|h\|_{L^p} \leqslant K$, which in particular implies that h is finite almost everywhere. At each point x such that $h(x) < \infty$, we have that $\sum_{k=1}^{\infty} \tilde{g}_k(x)$ converges absolutely, hence converges by the completeness of $\mathbb{C}$. We deduce that $\sum_{k=1}^{\infty} \tilde{g}_k$ converges pointwise almost everywhere and we define:

$$f(x) = \begin{cases} \sum_{k=1}^{\infty} \tilde{g}_k(x) & \text{if the sum converges} \\ 0 & \text{otherwise} \end{cases}$$

Now, we have that $|f| \leqslant h$, which implies $\|f\|_{L^p} \leqslant \|h\|_{L^p} \leqslant K$, and moreover:

$$\left| f - \sum_{k=1}^{n} \tilde{g}_k \right|^p \leqslant \left(|f| + \sum_{k=1}^{n} |\tilde{g}_k| \right)^p \leqslant (2h)^p.$$

[9] We typically don't state this point explicitly, but on this occasion we will make the distinction between $\mathcal{L}^p$ and L^p

Since h^p is integrable, by the Dominated Convergence Theorem (Thm B.26) we deduce that

$$\int_E \left| f - \sum_{k=1}^{n} \tilde{g}_k \right|^p d\mu \to 0, \qquad \text{as } n \to \infty,$$

which implies $\sum_{k=1}^{\infty} \tilde{g}_k$ converges to f in L^p. Noting that a different choice of representatives $\tilde{g}_n$ will result in h_n, h, f which differ from those defined above only on a set of measure zero, since the union of countably many sets of measure zero also has measure zero we are done. □

With this result in hand, we are able to establish

Theorem B.34 (Riesz-Fischer Theorem)**.** *Suppose* $(E, \mathcal{E}, \mu)$ *is a measure space and* $1 \leqslant p \leqslant \infty$. *Then* $L^p(E, \mu)$ *is complete.*

Proof. To prove completeness, suppose $(f_n)_{n=1}^{\infty}$ is a Cauchy sequence with respect to the L^p-norm. It suffices to show there exists $f \in L^p(E, \mu)$ with $f_n \to f$ in L^p. We split the cases $p < \infty$ and $p = \infty$.

1. First suppose $1 \leqslant p < \infty$. Then by the Cauchy property we can find a subsequence $(f_{n_k})_{k=1}^{\infty}$ such that
$$\left\| f_{n_{k+1}} - f_{n_k} \right\|_{L^p} < 2^{-k}.$$
Set $g_k = f_{n_{k+1}} - f_{n_k}$. By construction we have:
$$\sum_{k=1}^{\infty} \|g_k\|_{L^p} < \sum_{k=1}^{\infty} 2^{-k} = 1,$$
so by Lemma B.33 there exists $g \in L^p(E, \mu)$ such that
$$\sum_{k=1}^{\infty} g_k = g$$
with the sum converging pointwise a.e. and in L^p. Noting that $f_{n_{j+1}} = f_{n_1} + \sum_{k=1}^{j} g_k$, we deduce that $(f_{n_k})_{k=1}^{\infty}$ converges in L^p to some $f \in L^p(E, \mu)$. It follows by a standard argument using the fact it is a Cauchy sequence that $(f_n)_{n=1}^{\infty}$ converges to f in L^p.

2. Now consider the case $p = \infty$. Since (f_n) is Cauchy in $L^\infty(E, \mu)$, for each $m \in \mathbb{N}$ there exists n such that for any $j, m \geqslant n$ we have
$$|f_j(x) - f_k(x)| < \frac{1}{m} \qquad \text{for all } x \in N_{j,k,m}^c$$
where $\mu(N_{j,k,m}) = 0$. Let
$$N = \bigcup_{j,k,m} N_{j,k,m},$$

then $\mu(N) = 0$ and further we have that for any $m \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that for $j, k \geqslant n$:

$$\sup_{x \in N^c} |f_j(x) - f_k(x)| < \frac{1}{m}, \tag{B.6}$$

so by the completeness of $\mathbb{C}$ we have that for each $x \in N^c$ we have $f_j(x) \to f(x)$ for some $f(x) \in \mathbb{C}$ We let $f(x) = 0$ for $x \in N$, so that $f : E \to \mathbb{C}$. Sending $k \to \infty$ in (B.6) we see that for $j \geqslant n$

$$\sup_{x \in N^c} |f_j(x) - f(x)| < \frac{1}{m},$$

whence we conclude that $\|f\|_{L^\infty} < \infty$ and $f_j \to f$ in L^∞. □

Note that we have in fact proved the stronger result:

Corollary B.35. *Suppose $(f_n)_{n=1}^\infty$ is a Cauchy sequence in $L^p(E, \mu)$ for $1 \leqslant p \leqslant \infty$. Then there exists a subsequence $(f_{n_k})_{k=1}^\infty$ which converges pointwise almost everywhere.*

B.4.2 Density

It is often useful, when discussing topological spaces to identify dense subsets consisting of 'nice' or 'concrete' objects, for example elements of $\mathbb{Q}$ can be easily discussed, while a general element of $\mathbb{R}$ is typically expressible only as some limit of elements of $\mathbb{Q}$. In the main body of the course we shall establish that $C_c^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$. For a general measure space, we don't necessarily have a notion of continuity or smoothness, but we can show

Theorem B.36. *Let S be the set of all complex, measurable, simple functions on E such that:*

$$\mu(\{x : s(x) \neq 0\}) < \infty.$$

Then S is dense in $L^p(E, \mu)$ for $1 \leqslant p < \infty$.

Proof. Clearly $S \subset L^p(E, \mu)$. Now, suppose $f \geqslant 0$ with $f \in L^p(E, \mu)$ and let

$$f_n(x) = \min\{2^n \lfloor 2^{-n} f(x) \rfloor, n\}.$$

We have $f_n \in S$ and $0 \leqslant f_n \leqslant f$, so that $f_n \in L^p(E, \mu)$. Further, we know that $f_n(x) \to f(x)$ and moreover

$$|f - f_n|^p \leqslant |f|^p,$$

so by the Dominated convergence Theorem (Thm B.26) we deduce

$$\int_E |f - f_n|^p \, d\mu \to 0$$

hence $f_n \to f$ in L^p. A general (i.e. complex valued) element of $L^p(E, \mu)$ may be written as:

$$f = f_r^+ - f_r^- + i(f_i^+ - f_i^-),$$

where $f_r^\pm, f_i^\pm$ are non-negative elements of $L^p(E, \mu)$, hence the result follows. □