



Cairo University
Faculty of Sciences
Department of Mathematics

Functional Analysis II

Mostafa SABRI

Contents

1	LCS and Distributions	1
1.1	Introduction	1
1.2	Locally convex spaces	2
1.3	Fréchet Spaces	6
1.4	The inductive limit topology	8
1.5	Distributions	11
1.6	PoU and sheaf structure	13
1.7	Distributions with compact support	15
1.8	Further results and applications	16
1.9	Exercises	16
2	Fourier analysis	21
2.1	Convolutions	21
2.2	The Fourier transform	24
2.3	Sobolev spaces and embeddings	27
2.4	Paley-Wiener Theorem	32
2.5	Further results	33
2.6	Exercises	34
3	Some Important Theorems	39
3.1	Measure theory	39
3.2	Separation of convex sets	41
3.3	Banach-Alaoglu	42
3.4	Markov-Kakutani and Haar	44
3.5	Krein-Milman	46
3.6	Stone-Weierstrass vs Monotone Class	47
3.7	Appendix	50
4	Spectral theory	55
4.1	Compact and Hilbert-Schmidt operators	55
4.2	Spectral theorem for compact operators	58
4.3	Application : Peter-Weyl Theorem	60
4.4	The spectrum of a bounded operator	62
4.5	Continuous functional calculus	66
4.6	Borel functional calculus	67
4.7	Further results	69
4.8	Exercises	70
	Bibliography	75

Chapter 1

LCS and Distributions

1.1 Introduction

In this chapter we present the basics of the theory of *distributions*, also known as *generalized functions*.¹ Some motivation to introduce them:

- Differentiation of functions is not well behaved. For instance, there exist continuous functions which are nowhere differentiable. Another example : if a sequence (f_k) converges to some function f pointwise or uniformly, then (f'_k) does not necessarily converge to f' in the same sense.
- One would like to generalize the concept of a function to encompass objects like the “Dirac delta function”, which was defined by Dirac as a function δ satisfying $\delta(x) = 0$ if $x \neq 0$, $\delta(0) = +\infty$ and $\int_{\mathbb{R}} f(x)\delta(x) dx = f(0)$ for sufficiently smooth $f : \mathbb{R} \rightarrow \mathbb{R}$. These properties are of course contradictory, because if $\delta(x) = 0$ for all $x \neq 0$, then $f(x)\delta(x) = 0$ a.e., so $\int_{\mathbb{R}} f(x)\delta(x) dx = 0$. This is why Von Neumann objected its use during his work on the foundations of quantum mechanics. But despite this, physicists continued to use it and obtain quite successful results.

The theory of distributions was founded by Laurent Schwartz in the 1940’s, with some earlier results by Sergei Sobolev. The idea is that, instead of *restricting* the use of differentiation to smooth functions, one can instead *enlarge* the space of functions, so that it contains derivatives of functions which are not differentiable in the classical sense. For example, $f(x) = |x|$ will be differentiable on \mathbb{R} , with derivative $f'(x) = -1$ if $x < 0$ and $f'(x) = 1$ if $x > 0$. The way we’ll do this is to first define a convenient topology on the space of smooth functions with compact support, then let the space of “generalized functions” be its topological dual, and finally define on it calculus concepts like differentiation through the duality bracket. As a result,

- On this larger space of distributions, *any* generalized function is differentiable. Moreover, if (f_k) converges to f (in this dual topology), then $\partial^\alpha f_k$ converges to $\partial^\alpha f$ for all α . Finally, if the function is smooth, then “generalized differentiation” is the same as classic differentiation.
- The Dirac delta function has a simple interpretation as a distribution, and can indeed be seen as a limit of well behaved functions getting more and more singular at 0 (which is the physicist’s intuition). We call these “delta sequences”.
- As a bonus we get interesting applications in the theory of partial differential equations. It turns out that if $P(D)$ is a partial differential operator and if E is a distribution, then there always exist a solution to the equation $P(D)E = \delta$. This

1. Sections 1-4 simply unzip [15], the rest is taken from [32].

result is due to Malgrange and Ehrenpreis, and such solutions E are called “fundamental solutions”. If now we’d like to solve the PDE $P(D)u = v$, and if v has compact support, it suffices to choose $u = E * v$. Then u will be a solution.

- Let us finally mention that the famous “Schwartz kernel (or nuclear) theorem” concerning bilinear forms on functions of rapid decay was the starting point of a deep study by Grothendieck of spaces in which such theorems hold. His PhD thesis was published in [18], and the main results were summarized later in [17]; a paper which had a major impact in Banach space theory.

To conclude this introduction, let us mention that the idea of generalizing classical concepts by duality is quite fertile. For example, one uses the same idea to define “generalized eigenfunctions” in Spectral Theory. Here, starting from a Hilbert space \mathcal{H} , we first construct a smaller space $\mathcal{H}_+ \hookrightarrow \mathcal{H}$, define its dual \mathcal{H}_- , so that $\mathcal{H}_+ \hookrightarrow \mathcal{H} \hookrightarrow \mathcal{H}_-$, and let generalized eigenfunction live in \mathcal{H}_- .

1.2 Locally convex spaces

Definition 1.1. Let X be a vector space over $\mathbb{K} = \mathbb{R}$ or \mathbb{C} . A function $p : X \rightarrow \mathbb{R}$ is called a *seminorm* if

- (i) $p(x + y) \leq p(x) + p(y)$ for $x, y \in X$ (*sub-additivity*),
- (ii) $p(\lambda x) = |\lambda| \cdot p(x)$ for $\lambda \in \mathbb{K}$ and $x \in X$ (*positive homogeneity*).

Lemma 1.2. Let p be a seminorm on a vector space X . Then

- (a) $p(0) = 0$,
- (b) $p(x) \geq 0$,
- (c) $|p(x) - p(y)| \leq p(x - y)$,

Proof. By homogeneity, $p(0) = p(0 \cdot x) = 0 \cdot p(x) = 0$, which is (a).

By sub-additivity, $p(x) \leq p(x - y) + p(y)$, so $p(x - y) \geq p(x) - p(y)$. Substituting x for y we get $p(x - y) = |-1| \cdot p(y - x) \geq p(y) - p(x)$, so (c) follows. In particular, $p(x) \geq |p(x) - p(0)| \geq 0$, which is (b). \square

Definition 1.3. Given a seminorm on a vector space X , we call

$$B_{y,\varepsilon}(p) = \{x \in X : p(x - y) < \varepsilon\}$$

the p -*semiball* of radius $\varepsilon > 0$ centered at $y \in X$.

Given a finite collection of seminorms p_1, \dots, p_k , we call

$$B_{y,\varepsilon}(p_1, \dots, p_k) = B_{y,\varepsilon}(p_1) \cap \dots \cap B_{y,\varepsilon}(p_k)$$

the (p_1, \dots, p_k) -*semiball* of radius $\varepsilon > 0$ centered at $y \in X$.

Definition 1.4. Let X be a vector space and let \mathcal{P} be a family of seminorms on X . We define a topology $\tau_{\mathcal{P}}$ on X by calling $A \subseteq X$ *open* if $\forall x \in A, \exists \varepsilon > 0$ and $p_1, \dots, p_k \in \mathcal{P}$ such that $B_{x,\varepsilon}(p_1, \dots, p_k) \subseteq A$.

With this topology, we call (X, \mathcal{P}) a *locally convex space*.²

2. Remark on terminology: a topological vector space in general is called locally convex if there is a base for the topology that consists of convex sets. It can be shown that a space is locally convex in this sense iff its topology is generated by a family of seminorms; see [32]. The reader will prove some results in this direction in the exercises.

Lemma 1.5. *Let (X, \mathcal{P}) be a locally convex space. Then*

1. $\tau_{\mathcal{P}}$ is a topology on X .
2. The collection $\{B_{y,\varepsilon}(p_1, \dots, p_k) : y \in X, \varepsilon > 0 \text{ and } p_1, \dots, p_k \in \mathcal{P}\}$ is a base for $\tau_{\mathcal{P}}$.
3. The collection $\{B_{y,\varepsilon}(p) : y \in X, \varepsilon > 0 \text{ and } p \in \mathcal{P}\}$ is a subbase for $\tau_{\mathcal{P}}$.
4. If p_1 and p_2 are seminorms, then $\max(p_1, p_2)$ is a seminorm.

5. If

$$\overline{\mathcal{P}} = \mathcal{P} \cup \{\max\{p_1, \dots, p_k\} : p_1, \dots, p_k \in \mathcal{P}, k \in \mathbb{N}\},$$

then $\tau_{\mathcal{P}} = \tau_{\overline{\mathcal{P}}}$.

6. The collection $\{B_{y,\varepsilon}(p) : y \in X, \varepsilon > 0 \text{ and } p \in \overline{\mathcal{P}}\}$ is a base for $\tau_{\mathcal{P}}$.
7. If $p \in \mathcal{P}$ and p' is a seminorm satisfying $p'(x) \leq cp(x)$ for all $x \in X$ and some constant $c > 0$, then $\tau_{\mathcal{P}'} = \tau_{\mathcal{P}}$, where $\mathcal{P}' = \mathcal{P} \cup \{p'\}$.

The advantage of working with $\overline{\mathcal{P}}$ is that, similar to metric spaces, the semiballs $B_{y,\varepsilon}(p)$, $p \in \overline{\mathcal{P}}$ become a base, and we get rid of finite intersections $B_{y,\varepsilon}(p_1, \dots, p_k)$.

The point of the last item is that, if \mathcal{P}' is given initially, then using this result we can remove redundant seminorms from \mathcal{P}' to “clean it up” without affecting the topology.

Proof. 1. Clearly, $X \in \tau_{\mathcal{P}}$. Also, $\emptyset \in \tau_{\mathcal{P}}$ because any element in \emptyset (there are none) satisfies any property. Next, if $U, V \in \tau_{\mathcal{P}}$, then for any $x \in U \cap V$, we may find $\varepsilon_i > 0$ and $p_i, q_i \in \mathcal{P}$ such that $B_{x,\varepsilon_1}(p_1, \dots, p_k) \subseteq U$ and $B_{x,\varepsilon_2}(q_1, \dots, q_{k'}) \subseteq V$. Let $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$. Then $B_{x,\varepsilon}(p_1, \dots, p_k, q_1, \dots, q_{k'}) = B_{x,\varepsilon}(p_1, \dots, p_k) \cap B_{x,\varepsilon}(q_1, \dots, q_{k'}) \subseteq U \cap V$, so $U \cap V \in \tau_{\mathcal{P}}$. Finally, if $\{U_\alpha\}_{\alpha \in I}$ is any collection in $\tau_{\mathcal{P}}$ and $x \in \cup_{\alpha \in I} U_\alpha$, then $x \in U_\alpha$ for some α , so we may find $\varepsilon > 0$ and p_i such that $B_{x,\varepsilon}(p_1, \dots, p_k) \subseteq U_\alpha \subseteq \cup_{\alpha \in I} U_\alpha$, so the union is in $\tau_{\mathcal{P}}$. Hence, $\tau_{\mathcal{P}}$ is a topology.

2. Let $U \in \tau_{\mathcal{P}}$. Then $\forall x \in U \exists \varepsilon_x > 0$ and $p_{i,x}$ such that $B_{x,\varepsilon_x}(p_{1,x}, \dots, p_{k,x}) \subseteq U$. Hence, $U \subseteq \cup_{x \in U} B_{x,\varepsilon_x}(p_{1,x}, \dots, p_{k,x}) \subseteq U$, so $U = \cup_{x \in U} B_{x,\varepsilon_x}(p_{1,x}, \dots, p_{k,x})$. Thus, $\{B_{x,\varepsilon}(p_1, \dots, p_k)\}_{x \in U, \varepsilon > 0, p_i \in \mathcal{P}}$ is a base.

3. By item 2, $U = \cup_{x \in U} (B_{x,\varepsilon_x}(p_{1,x}) \cap \dots \cap B_{x,\varepsilon_x}(p_{k,x}))$. Thus, $\{B_{x,\varepsilon}(p)\}_{x \in U, \varepsilon > 0, p \in \mathcal{P}}$ is a subbase.

4. Easy, left for the reader.

5. If $U \in \tau_{\mathcal{P}}$, then $U \in \tau_{\overline{\mathcal{P}}}$ because $\mathcal{P} \subseteq \overline{\mathcal{P}}$. In detail: if $U \in \tau_{\mathcal{P}}$ and $x \in U$, there exist $\varepsilon > 0$ and $p_i \in \mathcal{P}$ such that $B_{x,\varepsilon}(p_1, \dots, p_k) \subseteq U$. In particular, there exist $p_i \in \overline{\mathcal{P}}$ such that $B_{x,\varepsilon}(p_1, \dots, p_k) \subseteq U$, so $U \in \tau_{\overline{\mathcal{P}}}$.

If $U \in \tau_{\overline{\mathcal{P}}}$, $\exists \varepsilon > 0$ and $q_i \in \overline{\mathcal{P}}$ such that $B_{x,\varepsilon}(q_1, \dots, q_k) \subseteq U$. Suppose $q_i = \max(p_i^{(1)}, \dots, p_i^{(m_i)})$. Then $B_{x,\varepsilon}(q_i) = B_{x,\varepsilon}(p_i^{(1)}) \cap \dots \cap B_{x,\varepsilon}(p_i^{(m_i)}) = B_{x,\varepsilon}(p_i^{(1)}, \dots, p_i^{(m_i)})$. Hence, $B_{x,\varepsilon}(p_1^{(1)}, \dots, p_1^{(m_1)}, p_2^{(1)}, \dots, p_k^{(m_k)}) = B_{x,\varepsilon}(q_1, \dots, q_k) \subseteq U$, so $U \in \tau_{\mathcal{P}}$.

6. By item 2, if $U \in \tau_{\mathcal{P}}$, then $U = \cup_{x \in U} B_{x,\varepsilon}(p_{1,x}, \dots, p_{k,x})$, where $p_{i,x} \in \mathcal{P}$. Hence, $U = \cup_{x \in U} B_{x,\varepsilon}(q_x)$, where $q_x = \max(p_{i,x}) \in \overline{\mathcal{P}}$. Hence, $\{B_{x,\varepsilon}(q)\}_{q \in \overline{\mathcal{P}}}$ is a base.

7. If $U \in \tau_{\mathcal{P}}$, then $U \in \tau_{\mathcal{P}'}$ because $\mathcal{P} \subseteq \mathcal{P}'$. Conversely, if $U \in \tau_{\mathcal{P}'}$, then given $x \in U$, there exist $\varepsilon > 0$, $p_i \in \mathcal{P}'$ such that $B_{x,\varepsilon}(p_1, \dots, p_k) \subseteq U$. If all $p_i \in \mathcal{P}$, for all $x \in U$, then $U \in \tau_{\mathcal{P}}$. Otherwise, say $p_i = p'$. Then $p'(x) \leq cp(x)$ implies $B_{y,\varepsilon}(p) \subseteq B_{y,c\varepsilon}(p')$. Hence, $B_{x,\varepsilon}(p_1) \cap \dots \cap B_{x,\varepsilon}(p_{i-1}) \cap B_{x,\varepsilon/c}(p) \cap \dots \cap B_{x,\varepsilon}(p_k) \subseteq B_{x,\varepsilon}(p_1, \dots, p_k) \subseteq U$. Hence, $U \in \tau_{\mathcal{P}}$ (take $\varepsilon' = \min(\varepsilon, \varepsilon/c)$). \square

Reminder 1.6. Recall that a sequence $\{x_k\} \subseteq X$ in a topological space converges if for any open set $\omega \subseteq X$ containing x , we have $x_k \in \omega$ for all large k .

Lemma 1.7. *Let (X, \mathcal{P}) be a locally convex space. Then $x_k \rightarrow x$ iff $p(x_k - x) \rightarrow 0$ for all $p \in \mathcal{P}$.*

Proof. Suppose $x_k \rightarrow x$ and let $p \in \mathcal{P}$. Then given $\varepsilon > 0$, there exists $k_0 > 0$ such that $x_k \in B_{x, \varepsilon}(p)$ for all $k \geq k_0$. Thus, $p(x - x_k) < \varepsilon$ for all $k \geq k_0$, so $p(x - x_k) \rightarrow 0$.

Conversely, suppose $p(x - x_k) \rightarrow 0$ for any $p \in \mathcal{P}$ and let ω be an open set containing x . Then there exists $\delta > 0$ and $p_i \in \mathcal{P}$ such that $B_{x, \delta}(p_1, \dots, p_k) \subseteq \omega$. Now we may find k_i such that $p_i(x_k - x) < \delta$ for all $k \geq k_i$. Let $k_0 = \max k_i$. Then $p_i(x - x_k) < \delta$ for all $k \geq k_0$ and all i . Hence, $x_k \in B_{x, \delta}(p_1, \dots, p_k) \subseteq \omega$ for all $k \geq k_0$. \square

The reader knows from the theory of Banach spaces that a linear operator is continuous iff it is bounded. For locally convex spaces, we have the following.

Theorem 1.8. *Let (X, \mathcal{P}) and (Y, \mathcal{Q}) be locally convex spaces. Then*

(a) *$f : X \rightarrow Y$ is continuous iff $\forall x \in X, \varepsilon > 0, q \in \overline{\mathcal{Q}}, \exists p \in \overline{\mathcal{P}}, \delta > 0$ such that*

$$z \in B_{x, \delta}(p) \implies q(f(x) - f(z)) < \varepsilon. \quad (\star)$$

(b) *If moreover f is linear, then f is continuous iff $\forall q \in \overline{\mathcal{Q}}, \exists p \in \overline{\mathcal{P}}, C > 0$ such that*

$$q(f(x)) \leq Cp(x) \quad \forall x \in X. \quad (\star\star)$$

(c) *A seminorm $q : X \rightarrow \mathbb{R}$ (not necessarily in \mathcal{P}) is continuous on (X, \mathcal{P}) iff $\exists p \in \overline{\mathcal{P}}, C > 0$ such that*

$$q(x) \leq Cp(x) \quad \forall x \in X.$$

Proof. (a) Suppose f is continuous and let $x \in X, \varepsilon > 0$ and $q \in \overline{\mathcal{Q}}$. Then $B_{f(x), \varepsilon}(q)$ is open in Y , so $f^{-1}(B_{f(x), \varepsilon}(q))$ is open in X . But $x \in f^{-1}(B_{f(x), \varepsilon}(q))$, so $\exists \delta > 0, p \in \overline{\mathcal{P}}$ such that $B_{x, \delta}(p) \subseteq f^{-1}(B_{f(x), \varepsilon}(q))$. Hence, $z \in B_{x, \delta}(p)$ implies $f(z) \in B_{f(x), \varepsilon}(q)$ implies $q(f(z) - f(x)) < \varepsilon$.

Conversely, suppose (\star) is true, let $U \subseteq Y$ be open and let $x \in f^{-1}(U)$. Then $f(x) \in U$, so $\exists \varepsilon > 0, q \in \overline{\mathcal{Q}}$ such that $B_{f(x), \varepsilon}(q) \subseteq U$. By (\star) we may find $\delta > 0, p \in \overline{\mathcal{P}}$ such that $B_{x, \delta}(p) \subseteq f^{-1}(B_{f(x), \varepsilon}(q)) \subseteq f^{-1}(U)$. Hence, $f^{-1}(U)$ is open.

(b) If $(\star\star)$ is true, by linearity $q(f(x) - f(z)) \leq Cp(x - z)$. So given $\varepsilon > 0$, taking $\delta = \varepsilon/C$, we have $z \in B_{x, \delta}(p)$ implies $p(x - z) < \delta$ implies $q(f(x) - f(z)) \leq C\delta = \varepsilon$.

Conversely, suppose (\star) is true. Then there is $\delta > 0$ and $p \in \overline{\mathcal{P}}$ such that $z \in B_{0, \delta}(p)$ implies $q(f(z)) \leq 1$. Let $x \in X$ with $p(x) \neq 0$. Then $z = \frac{\delta}{2p(x)}x$ satisfies $p(z) = \frac{\delta}{2} < \delta$, hence $1 \geq q(f(z)) = \frac{\delta}{2p(x)}q(f(x))$. We thus get $(\star\star)$ with $C = \frac{2}{\delta}$.

For the case $p(z) = 0$, we must show $q(f(z)) = 0$. By (\star) , for any $\varepsilon > 0$ we have $z \in B_{0, \delta}(p)$ implies $q(f(z)) < \varepsilon$. In particular, $p(z) = 0$ implies $q(f(z)) < \varepsilon$. As $\varepsilon > 0$ is arbitrary, we have $q(f(z)) = 0$. Thus, $(\star\star)$ remains true in this case.

(c) Part (c) is analogous to (b) and is left to the reader. \square

Corollary 1.9. *Let (X, \mathcal{P}) and (Y, \mathcal{Q}) be locally convex with $X \subset Y$. Then the following statements are equivalent:*

- (i) *The embedding $X \hookrightarrow Y$ is continuous.*
- (ii) *$\forall q \in \overline{\mathcal{Q}} \exists p \in \overline{\mathcal{P}}, C > 0$ such that $q(x) \leq Cp(x)$ for all $x \in X$.*
- (iii) *The restriction of every seminorm of (Y, \mathcal{Q}) to X is continuous on (X, \mathcal{P}) .*

Proof. (i) is equivalent to (ii) by Theorem 1.8 (b), taking $f : X \rightarrow Y$ as the inclusion map $f(x) = x$.

(ii) implies (iii) by Theorem 1.8 (c). Next, if (iii) holds and $q \in \overline{\mathcal{Q}}$, then $q = \max_{1 \leq i \leq k} (q_i)$ for some $q_i \in \mathcal{Q}$. By Theorem 1.8 (c), $\exists p_i \in \overline{\mathcal{P}}$ and $C_i > 0$ such that $q_i(x) \leq C_i p_i(x)$. Let $C = \max C_i$ and $p = \max p_i$. Then $q(x) \leq Cp(x)$ and (ii) holds. \square

In view of Lemma 1.7, we make the following definition:

Definition 1.10. Let (X, \mathcal{P}) be a locally convex space. We say that

- $\{x_k\}$ is *Cauchy* if for any $p \in \mathcal{P}$, $p(x_j - x_k) \rightarrow 0$ as $j, k \rightarrow \infty$.
- $A \subseteq X$ is *bounded* if for any $p \in \mathcal{P}$, $\sup_{x \in A} p(x) < \infty$.

Lemma 1.11. A Cauchy sequence is bounded.

Proof. If $\{x_k\}$ is Cauchy, then for any $p \in \mathcal{P}$ we may find m_0 such that $p(x_j - x_k) < 1$ for $j, k \geq m_0$. Hence, $p(x_k) \leq p(x_{m_0}) + 1$ for all $k \geq m_0$, so $\sup_k p(x_k) \leq C$, where $C = \max(p(x_1), \dots, p(x_{m_0-1}), p(x_{m_0}) + 1)$. \square

Definition 1.12. A family of seminorms \mathcal{P} on X is said to be *separating* or *sufficient* if for any $x \in X \setminus \{0\}$, there exists $p \in \mathcal{P}$ such that $p(x) \neq 0$.

Lemma 1.13. Let (X, \mathcal{P}) be locally convex with \mathcal{P} separating. Then $\tau_{\mathcal{P}}$ is Hausdorff.

Proof. Let $x, y \in X$, $x \neq y$. Then $\exists p \in \mathcal{P}$ with $\delta := p(x - y) > 0$. Let $U = B_{x, \delta/2}(p)$ and $V = B_{y, \delta/2}(p)$. Then U, V are open, satisfy $x \in U$, $y \in V$ and $U \cap V = \emptyset$. \square

Reminder 1.14. Recall that a topological space (X, τ) is said to be *metrizable* if there is a metric d on X which induces τ . If X is a vector space, we say that a metric is *translation invariant* if $d(x + z, y + z) = d(x, y)$ for all $z \in X$.

Theorem 1.15. Let (X, \mathcal{P}) be a locally convex space with \mathcal{P} countable and separating. Then (X, \mathcal{P}) is metrizable with a translation invariant metric.³

Proof. Let $\mathcal{P} = \{p_1, p_2, \dots\}$ and let $\{\alpha_k\} \subset \mathbb{R}_+^*$ with $\alpha_k \rightarrow 0$. Define

$$d(x, y) = \max_k \frac{\alpha_k p_k(x - y)}{1 + p_k(x - y)} \quad \text{for } x, y \in X.$$

The maximum is well-defined because $\frac{p_k}{1+p_k} \leq 1$ and $\alpha_k \rightarrow 0$. Since $\alpha_k > 0$ for all k , we have $d(x, y) = 0$ iff $p_k(x - y) = 0$ for all k iff $x = y$ because \mathcal{P} is separating. Next, since $f(t) = \frac{t}{1+t}$ is increasing, we have $a \leq b + c$ implies $\frac{a}{1+a} \leq \frac{b+c}{1+b+c} \leq \frac{b}{1+b} + \frac{c}{1+c}$ and the triangle inequality follows. Since d is clearly symmetric, it follows that d is a metric.

Now let $\varepsilon > 0$ and let $B_{x, \varepsilon}(d) = \{y \in X : d(x, y) < \varepsilon\}$ be a metric ball. If $\alpha_k < \varepsilon \forall k$, then $d(x, y) < \varepsilon$ for any y , so $B_{x, \varepsilon}(d) = X \in \tau_{\mathcal{P}}$. Otherwise, there are finitely many i such that $\alpha_i \geq \varepsilon$ (since $\alpha_k \rightarrow 0$), say $\alpha_{i_1}, \dots, \alpha_{i_m}$. Then a simple calculation shows that $B_{x, \varepsilon}(d) = B_{x, \delta_1}(p_{i_1}) \cap \dots \cap B_{x, \delta_m}(p_{i_m})$, where $\delta_j = \frac{\varepsilon}{\alpha_j - \varepsilon}$. Hence, $B_{x, \varepsilon}(d) \in \tau_{\mathcal{P}}$ for all $\varepsilon > 0$. But if U is open in the metric topology, then $U = \cup_{x \in U} B_{x, \varepsilon_x}(d)$, hence, $U \in \tau_{\mathcal{P}}$.

3. The proof of many results in this course, including this one, can be simplified if you know about nets; see e.g. [28] or [41]. Nets generalize the concept of a sequence. A nice feature about nets is that they characterize the topology. For example, suppose we have two topologies τ_1 and τ_2 on a set X and we want to show that $\tau_1 = \tau_2$. Then it suffices to show that a net converges in (X, τ_1) iff it converges in (X, τ_2) . This is an easy exercise if you know that $x \in \overline{A}$ iff there exists a net $\{x_\nu\} \subset A$ converging to x (just show that consequently, A is closed in (X, τ_1) iff it is closed in (X, τ_2)).

A related feature is that given a set X , one can define a topology on X by stating which nets converge. See e.g. [42, Problem 11D].

Conversely, suppose $U \in \tau_{\mathcal{P}}$ and let $x \in U$. Then $\exists \varepsilon > 0$, $p_i \in \mathcal{P}$ such that $B_{x,\varepsilon}(p_1, \dots, p_k) \subseteq U$. Choose $r < \frac{\varepsilon}{1+\varepsilon} \min(\alpha_1, \dots, \alpha_k)$ and let $y \in B_{x,r}(d)$. Then

$$\frac{\alpha_i p_i(x-y)}{1+p_i(x-y)} \leq d(x,y) < r < \frac{\alpha_i \varepsilon}{1+\varepsilon} \quad \text{for } i = 1, \dots, k,$$

so $\frac{p_i(x-y)}{1+p_i(x-y)} < \frac{\varepsilon}{1+\varepsilon}$ and thus $p_i(x-y) < \varepsilon$. Hence, $B_{x,r}(d) \subseteq B_{x,\varepsilon}(p_1, \dots, p_k) \subseteq U$. Thus, U is open in the metric topology. \square

1.3 Fréchet Spaces

Definition 1.16. A *Fréchet space* is a locally convex space (X, \mathcal{P}) such that \mathcal{P} is countable, separating, and (X, \mathcal{P}) is complete.

By Theorem 1.15, any Fréchet space is metrizable with a translation invariant metric. Actually, the converse is also true: if X is a metrizable locally convex space, then its topology is generated by a countable separating family of seminorms. These seminorms are constructed using the Minkowski functional (see Exercise 2) and the fact that metric spaces are first countable. We will not pursue this here; see e.g. [28] and [32].

Notation. We denote $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Given $\alpha \in \mathbb{N}_0^n$ and $x \in \mathbb{R}^n$, we denote

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n} \quad \text{and} \quad D^\alpha = \partial^\alpha = \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}.$$

Finally, we let $|\alpha| = \alpha_1 + \dots + \alpha_n$.

Lemma 1.17. Let $K \subset \mathbb{R}^n$ be compact. Then the space $C(K)$ of continuous functions on K , endowed with the norm $\|f\|_{C(K)} = \sup_{x \in K} |f(x)|$, is a Banach space.

Proof. Exercise 8; just adapt the proof you learned for $C[a, b]$ in undergraduate courses. More generally, if $A \subseteq \mathbb{R}^n$ and $C_b(A)$ denotes the set of bounded continuous functions on A , then $C_b(A)$ is a Banach space in the supremum norm. \square

Lemma 1.18. Let $K \subset \mathbb{R}^n$ be compact and let $C^m(K) = \{f|_K : f \in C^m(\mathbb{R}^d)\}$, $m \in \mathbb{N}$. Then $C^m(K)$ is a Banach space in the norm $\|f\|_{C^m(K)} = \max_{|\alpha| \leq m} \|D^\alpha f\|_{C(K)}$.

Proof. Exercise 8. If (f_k) is Cauchy in $C^m(K)$, then $(D^\alpha f_k)$ is Cauchy in $C(K)$ for each $|\alpha| \leq m$. Now suppose $n = 1$, use Lemma 1.17, the fundamental theorem of calculus and induction. Generalize for arbitrary n using partial derivatives. \square

Lemma 1.19. Let $K \subset \mathbb{R}^n$ be compact and let $C^\infty(K) = \bigcap_{m \in \mathbb{N}} C^m(K)$. Then $C^\infty(K)$, endowed with the seminorms $p_m(f) = \|f\|_{C^m(K)}$, $m = 0, 1, 2, \dots$, is a Fréchet space.

Proof. Clearly, $\mathcal{P} = \{p_m\}$ is countable and separating. Let (f_k) be Cauchy in $C^\infty(K)$. Then (f_k) is Cauchy in $C^m(K)$ for each m , so by Lemma 1.18, we may find $g_m \in C^m(K)$ such that $p_m(f_k - g_m) \rightarrow 0$ as $k \rightarrow \infty$. But

$$p_0(g_0 - g_m) \leq p_0(f_k - g_0) + p_0(f_k - g_m) \leq p_0(f_k - g_0) + p_m(f_k - g_m) \rightarrow 0$$

as $k \rightarrow \infty$. Hence, $g_0 = g_m \in C^m(K)$ for every m . Hence, $g_0 \in C^\infty(K)$ and $f_k \rightarrow g_0$ in $C^\infty(K)$. \square

We now turn to $C^m(\Omega)$ where $\Omega \subseteq \mathbb{R}^n$ is an open set. We first give a technical lemma.

Lemma 1.20. *Let $\Omega \subseteq \mathbb{R}^n$ be open. Then there is a sequence of compact sets K_1, K_2, \dots such that K_j is contained in the interior of K_{j+1} and $\cup_j K_j = \Omega$.*

Proof. If $\Omega = \mathbb{R}^n$, take $K_j = \overline{B}_j = \{x \in \mathbb{R}^n : |x| \leq j\}$, where $|x|$ is the euclidean norm on \mathbb{R}^n . If $\Omega \neq \mathbb{R}^n$, let $A_j = \{x \in \Omega : \text{dist}(x, \partial\Omega) \geq \frac{1}{j}\}$. Then A_j is closed and lies in the interior of A_{j+1} . Take $K_j = A_j \cap \overline{B}_j$. Then K_j satisfies the requirements. \square

Lemma 1.21. *Let $\Omega \subseteq \mathbb{R}^n$ be open and let $K_j \subset \Omega$ be compact, $K_j \subset \overset{\circ}{K}_{j+1}$ and $\cup_j K_j = \Omega$. Endow $C^m(\Omega)$, $m \in \mathbb{N}$, with the seminorms $p_{j,m}(f) = \|f\|_{C^m(K_j)}$, $j = 1, 2, \dots$. Then $C^m(\Omega)$ is a Fréchet space.*

Similarly, $C^\infty(\Omega)$, endowed with the seminorms $p_{j,k}(f) = \|f\|_{C^k(K_j)}$, $j = 1, 2, \dots$ and $k = 0, 1, \dots$ is a Fréchet space.

Proof. Let (f_i) be Cauchy in $C^m(\Omega)$. By completeness of $C^m(K_j)$, we may find $g_j \in C^m(K_j)$ such that $p_{j,m}(f_i - g_j) \rightarrow 0$ as $i \rightarrow \infty$. But $p_{j,m}(g) \leq p_{j+1,m}(g)$ for $g \in C^m(K_{j+1})$, so as in Lemma 1.19, it follows that $g_j = g_{j+1}|_{K_j}$. Hence, we may define g on Ω by $g(x) = g_j(x)$ if $x \in K_j$. Then $g \in C^m(\Omega)$ and $p_{j,m}(f_i - g) \rightarrow 0$ as $i \rightarrow \infty$, for any j .

The case of $C^\infty(\Omega)$ is the same; just consider $p_{j,k}$ for arbitrary k instead of $p_{j,m}$. \square

Remark 1.22. We could also consider $C^m(\Omega)$ with the larger family $p_{K,m}(f) = \|f\|_{C^m(K)}$, where K runs over all compact subsets of Ω . But the topology on $C^m(\Omega)$ would actually be the same. Indeed, if $K \subset \Omega$ is compact, then $K \subset K_j$ for some j , so $p_K(f) \leq p_{K_j}(f)$, so we may remove p_K from the family without affecting the topology by Lemma 1.5.7.

Lemma 1.23. *Let $\Omega \subset \mathbb{R}^n$ be open and for $1 \leq p \leq \infty$, define*

$$L^p_{\text{loc}}(\Omega) = \{u : \Omega \rightarrow \mathbb{C} \text{ measurable, } u|_K \in L^p(K) \text{ for any compact } K \subset \Omega\}.$$

Then $L^p_{\text{loc}}(\Omega)$ is a Fréchet space when endowed with the seminorms $q_j(u) = \|u\|_{L^p(K_j)}$

Proof. Same proof as Lemma 1.21; you can write it as an exercise. \square

We now turn to the space of smooth functions with compact support. Recall that if $f : \Omega \rightarrow \mathbb{C}$ is continuous, we define the support of f by $\text{supp } f = \overline{\{x \in \Omega : f(x) \neq 0\}}$.

We should first check that such functions exist, since a nonzero analytic function cannot have compact support, and being analytic doesn't seem much stronger than being smooth.

Lemma 1.24. *Let $R > r > 0$. There exists $\phi \in C^\infty(\mathbb{R}^n)$ such that $\phi(x) = 1$ for $|x| \leq r$, $\phi(x) = 0$ for $|x| \geq R$ and $0 \leq \phi \leq 1$ on \mathbb{R}^n .⁴*

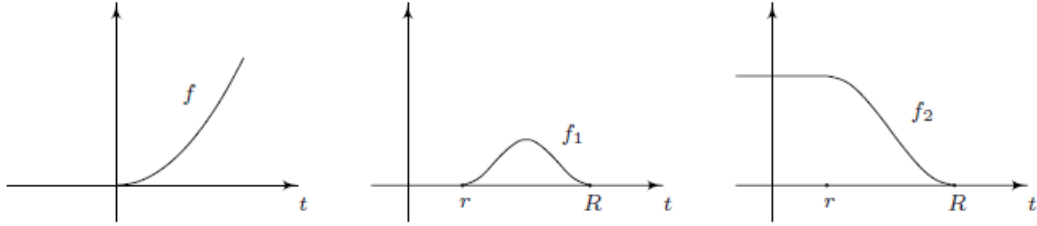
Proof. Let $f(t) = \begin{cases} e^{-1/t} & \text{if } t > 0, \\ 0 & \text{if } t \leq 0. \end{cases}$ Clearly, $f(t) \rightarrow 0$ as $t \searrow 0$. Moreover, $f^{(k)}(t) = \begin{cases} p_k(1/t)e^{-1/t} & \text{if } t > 0, \\ 0 & \text{if } t < 0, \end{cases}$ for some polynomials p_k . Since $p(1/t)e^{-1/t} \rightarrow 0$ as $t \searrow 0$ for any polynomial p , it follows that $f \in C^\infty(\mathbb{R})$. Now let $R > r > 0$, let $f_1(t) = f(t-r)f(R-t)$ and let $f_2(t) = \int_t^\infty f_1(s) ds$. Then $f_2(t) \geq 0$ for all t , equals 0 for $t \geq R$ and equals $C := \int_r^R f_1(s) ds > 0$ for $t \leq r$. Thus, $\phi(x) := \frac{1}{C} f_2(|x|)$ satisfies the desired properties. \square

Definition 1.25. Let $K \subset \mathbb{R}^n$ be compact. We define

$$\mathcal{D}_K = \{\varphi \in C^\infty(\mathbb{R}^n) : \text{supp } \varphi \subset K\}.$$

and endow it with the family $\mathcal{P}^\infty = \{p_m\}$, where $p_m(\varphi) = \|\varphi\|_{C^m}$, $m = 0, 1, 2, \dots$

4. Lemma and picture taken from [19].



- Remarks 1.26.** (i) This topology is the one induced by the embedding $\mathcal{D}_K \hookrightarrow C^\infty(K)$. Note that \mathcal{D}_K is a bit smaller than $C^\infty(K)$. For example, $\varphi_1 \equiv 1$ on K can be extended to a smooth function in $C^\infty(\mathbb{R}^n)$ and is thus in $C^\infty(K)$, but $\varphi_1 \notin \mathcal{D}_K$ because any smooth extension of φ_1 must be nonzero in a small neighborhood of K .
- (ii) If K contains an open ball, then by scaling and translating ϕ , we can find infinitely many smooth functions so that their supports are in K but do not intersect each other. This implies that \mathcal{D}_K is infinite-dimensional.
- (iii) \mathcal{D}_K is a Fréchet space, as it is a closed subspace of $C^\infty(K)$.
- (iv) Note that $\overline{\mathcal{P}^\infty} = \mathcal{P}^\infty$. Indeed, $\mathcal{P}^\infty \subseteq \overline{\mathcal{P}^\infty}$ is clear, and if $q \in \overline{\mathcal{P}^\infty}$, say $q = \max(q_{i_1}, \dots, q_{i_k})$ for some $q_{i_j} \in \mathcal{P}^\infty$, then $q_{i_j}(\varphi) = \|\varphi\|_{C^{i_j}}$. Take $m = \max(i_1, \dots, i_k)$. Then $q(\varphi) = p_m(\varphi)$, so $q \in \mathcal{P}^\infty$.

1.4 The inductive limit topology

Definition 1.27. Let $\Omega \subseteq \mathbb{R}^n$ be open. We define the *space of test functions* by

$$\mathcal{D}(\Omega) = \bigcup_{K \in \Omega} \mathcal{D}_K,$$

where $K \in \Omega$ means that K is a compact subset of Ω .

Note that $\mathcal{D}(\Omega)$ is simply the set of smooth functions of compact support in Ω , usually denoted $C_c^\infty(\Omega)$. Here we use the symbol $\mathcal{D}(\Omega)$ because we will now endow it with a special topology which is stronger than the C^∞ topology.

The topology we would like to have on $\mathcal{D}(\Omega)$ should fulfill two requirements:

- 1) The embeddings $\mathcal{D}_K \hookrightarrow \mathcal{D}(\Omega)$ should be continuous. This means by Corollary 1.9 that the restriction of every seminorm p in $(\mathcal{D}(\Omega), \mathcal{P})$ to \mathcal{D}_K must be continuous. Here \mathcal{P} is the hypothetical family inducing a topology on $\mathcal{D}(\Omega)$. The family \mathcal{P}^∞ satisfies this requirement by Theorem 1.8(c).

However, $(\mathcal{D}(\Omega), \mathcal{P}^\infty)$ is not complete. To see this for $\Omega = \mathbb{R}$, take a $\varphi \in \mathcal{D}(\mathbb{R})$ whose support is small and concentrated near 0, and consider the sequence

$$\varphi_k(x) = \varphi(x) + 2^{-1}\varphi(x-1) + \dots + 2^{-k}\varphi(x-k), \quad k = 1, 2, \dots$$

Then (φ_k) is Cauchy in \mathcal{P}^∞ : given $r < s$,

$$p_m(\varphi_r - \varphi_s) = p_m(2^{-r-1}\varphi(x-r-1) + \dots + 2^{-s}\varphi(x-s)) = 2^{-r-1}\|\varphi\|_{C^m} \rightarrow 0$$

as r (and hence s) go to ∞ . However, the support of the limit function is not compact.

- 2) This failure indicates that \mathcal{P}^∞ has not enough seminorms to ensure that Cauchy sequences remain in a compact set. So we can add more seminorms to \mathcal{P}^∞ and hope that things get better. Having a large family of seminorms will also have the benefit that it becomes easier for a function $f : \mathcal{D}(\Omega) \rightarrow Y$ to be continuous. However, we should not enlarge \mathcal{P}^∞ too much, since we still need every restriction $p|_{\mathcal{D}_K}$ to be continuous.

These two competing requirements give rise to a unique family \mathcal{P} as follows.

Definition 1.28. We endow $\mathcal{D}(\Omega)$ with the family \mathcal{P} defined by $p \in \mathcal{P} \iff p|_{\mathcal{D}_K}$ is continuous for each compact $K \subset \Omega$.

The induced topology $\tau_{\mathcal{P}}$ on $\mathcal{D}(\Omega)$ is called the *inductive limit topology*.

Remark 1.29. Note that $\overline{\mathcal{P}} = \mathcal{P}$. Indeed, $\mathcal{P} \subseteq \overline{\mathcal{P}}$ is clear, and if $q \in \overline{\mathcal{P}}$, say $q = \max(q_{i_1}, \dots, q_{i_k})$ for some $q_{i_j} \in \mathcal{P}$, then given $K \subset \Omega$ compact, $q_{i_j}|_{\mathcal{D}_K}$ is continuous, so $\exists c_j > 0$ and $p_j \in \mathcal{P}^\infty$ such that $q_{i_j}(\varphi) \leq c_j p_j(\varphi)$ for $\varphi \in \mathcal{D}_K$. Let $c = \max(c_1, \dots, c_k)$ and $p = \max(p_1, \dots, p_k) \in \overline{\mathcal{P}^\infty} = \mathcal{P}^\infty$. Then $q(\varphi) \leq cp(\varphi)$ for $\varphi \in \mathcal{D}_K$, hence $q \in \mathcal{P}$.

Lemma 1.30. *The topology of \mathcal{D}_K is exactly the one induced by the embedding $\mathcal{D}_K \hookrightarrow \mathcal{D}(\Omega)$. In other words, the topology of \mathcal{D}_K is exactly the subspace topology inherited from $\mathcal{D}(\Omega)$.*

This shows that the topology of $\mathcal{D}(\Omega)$ is completely natural; with a different topology, the \mathcal{D}_K could have obtained two topologies: their own $\tau_{\mathcal{P}^\infty}$ topology, and a possibly different one inherited from $\mathcal{D}(\Omega)$, which is certainly inconvenient.

Proof. Let $A \subseteq \mathcal{D}(\Omega)$ be open, let $K \subset \Omega$ be compact and let $\psi \in A \cap \mathcal{D}_K$. Then $\exists \varepsilon > 0$, $p \in \mathcal{P}$ such that $B_{\psi, \varepsilon}(p) \subset A$. But $p|_{\mathcal{D}_K}$ is continuous, so $\exists p_m$ with $p \leq cp_m$ on \mathcal{D}_K , $c > 0$. Hence, if $B_{\psi, \varepsilon}(p_m; \mathcal{D}_K)$ denotes semiballs in \mathcal{D}_K , we get $B_{\psi, \varepsilon/c}(p_m, \mathcal{D}_K) \subseteq B_{\psi, \varepsilon}(p) \cap \mathcal{D}_K \subseteq A \cap \mathcal{D}_K$. Hence, $A \cap \mathcal{D}_K$ is open in \mathcal{D}_K .

Conversely, since $\{p_m\} \subset \mathcal{P}$, the semiball $B_{\varphi, \delta}(p_m)$ has a meaning, and any semiball $B_{\varphi, \delta}(p_m; \mathcal{D}_K) = B_{\varphi, \delta}(p_m) \cap \mathcal{D}_K$. It follows that any open set in \mathcal{D}_K can be written as an intersection of an open set in $\mathcal{D}(\Omega)$ with \mathcal{D}_K . \square

We now show that \mathcal{P} is large enough to solve the problem we had with \mathcal{P}^∞ :

Theorem 1.31. *A set $A \subseteq \mathcal{D}(\Omega)$ is bounded iff there is a compact $K \subset \Omega$ such that $A \subseteq \mathcal{D}_K$ and A is bounded in \mathcal{D}_K .*

Proof. Suppose A is bounded in \mathcal{D}_K for some compact $K \subset \Omega$ and let $p \in \mathcal{P}$. By continuity of $p|_{\mathcal{D}_K}$, $\exists p_m$ such that $p \leq cp_m$ on \mathcal{D}_K . But $\exists M > 0$ such that $p_m(\varphi) \leq M$ for $\varphi \in A$. Hence, $p(\varphi) \leq cM$ on A and A is bounded in $\mathcal{D}(\Omega)$.

Conversely, suppose $A \subseteq \mathcal{D}(\Omega)$ is bounded but $A \not\subseteq \mathcal{D}_K$ for any compact $K \subset \Omega$. Then there is a sequence $\{\varphi_m\} \subseteq A$ and $\{x_m\} \subset \Omega$ such that $\varphi_m(x_m) \neq 0$ and $\{x_m\}$ leaves any compact K for large m . Let

$$p(\varphi) = \sup_m \frac{m|\varphi(x_m)|}{|\varphi_m(x_m)|}, \quad \varphi \in \mathcal{D}(\Omega).$$

Then p is a seminorm and $p \in \mathcal{P}$. Indeed, if $K' \subset \Omega$ is compact, $\exists m_0$ with $x_m \notin K'$ for all $m \geq m_0$, so for any $\varphi \in \mathcal{D}_{K'}$, we have $\varphi(x_m) = 0$ for all $m \geq m_0$, so $p(\varphi) = \max_{1 \leq m \leq m_0} \frac{m|\varphi(x_m)|}{|\varphi_m(x_m)|} \leq C\|\varphi\|_{C^0}$. But $p(\varphi_j) \geq j$, so p is not bounded on A . This contradiction shows that $A \subseteq \mathcal{D}_K$ for some compact K . \square

Corollary 1.32. a) *A sequence $\{\varphi_j\}$ is Cauchy in $\mathcal{D}(\Omega)$ iff $\{\varphi_j\} \subseteq \mathcal{D}_K$ for some compact $K \subset \Omega$ and $\{\varphi_j\}$ is Cauchy in \mathcal{D}_K .*

- b) $\varphi_j \rightarrow 0$ in $\mathcal{D}(\Omega)$ iff $\{\varphi_j\} \subseteq \mathcal{D}_K$ for some compact $K \subset \Omega$ and $\varphi_j \rightarrow 0$ in \mathcal{D}_K .
 c) $\mathcal{D}(\Omega)$ is complete.

Proof. a) If $\{\varphi_j\} \subseteq \mathcal{D}_K$ is Cauchy in \mathcal{D}_K and $p \in \mathcal{P}$, then $p|_{\mathcal{D}_K}$ is continuous, so $\exists p_m$ with $p \leq cp_m$. Hence, $p(\varphi_r - \varphi_s) \leq cp_m(\varphi_r - \varphi_s)$, so $\{\varphi_j\}$ is Cauchy in $\mathcal{D}(\Omega)$.

Conversely, if $\{\varphi_j\} \subseteq \mathcal{D}(\Omega)$ is Cauchy, then $\{\varphi_j\}$ is bounded, so by Theorem 1.31, $\{\varphi_j\} \subseteq \mathcal{D}_K$ for some compact $K \subset \Omega$. Since $\{p_m\} \subset \mathcal{P}$, we also have $p_m(\varphi_r - \varphi_s) \rightarrow 0$ as $r, s \rightarrow \infty$ for each p_m .

- b) Same as a), left as an exercise.
 c) Let $\{\varphi_j\} \subseteq \mathcal{D}(\Omega)$ be Cauchy. Then by a), it is Cauchy in some \mathcal{D}_K . But \mathcal{D}_K is Fréchet, so the limit exists in \mathcal{D}_K . Using b), the sequence also converges in $\mathcal{D}(\Omega)$. \square

Theorem 1.33. *Let (Y, \mathcal{Q}) be locally convex and let $f : \mathcal{D}(\Omega) \rightarrow Y$ be linear. Then the following statements are equivalent:*

- (i) f is continuous.
 (ii) $\varphi_j \rightarrow 0$ in $\mathcal{D}(\Omega)$ implies $f(\varphi_j) \rightarrow 0$ in Y .
 (iii) For any compact $K \subset \Omega$, $f : \mathcal{D}_K \rightarrow Y$ is continuous.

Proof. (i) \implies (ii). If f is continuous, by Theorem 1.8 for any $q \in \overline{\mathcal{Q}}$ there is $p \in \mathcal{P}$ and $C > 0$ such that $q(f(\varphi)) \leq Cp(\varphi)$ for all $\varphi \in \mathcal{D}(\Omega)$. Hence, $\varphi_j \rightarrow 0$ in $\mathcal{D}(\Omega)$ implies $p(\varphi_j) \rightarrow 0$ implies $q(f(\varphi_j)) \rightarrow 0$. Since this holds for all q , we have $f(\varphi_j) \rightarrow 0$ in Y .

(ii) \implies (iii). Let $K \subset \Omega$ be compact. If (ii) holds, then using Corollary 1.32, if $\varphi_j \rightarrow 0$ in \mathcal{D}_K , then $f(\varphi_j) \rightarrow 0$ in Y . So by linearity, if $\varphi_j \rightarrow \varphi$ in \mathcal{D}_K , then $f(\varphi_j) \rightarrow f(\varphi)$. Thus, f preserves convergent sequences. But \mathcal{D}_K is metrizable. Hence, f is continuous.⁵

(iii) \implies (i) By Theorem 1.8, we must show that $\forall q \in \overline{\mathcal{Q}} \exists p \in \mathcal{P}, C > 0$ such that $q(f(\varphi)) \leq Cp(\varphi)$ on $\mathcal{D}(\Omega)$. Let $q \in \overline{\mathcal{Q}}$ and define

$$p(\varphi) := q(f(\varphi)), \quad \varphi \in \mathcal{D}(\Omega).$$

Then p is a seminorm on $\mathcal{D}(\Omega)$. Moreover, if $K \subset \Omega$ is compact, then by continuity of f on \mathcal{D}_K , we may find $c > 0$ and p_m such that $p(\varphi) = q(f(\varphi)) \leq cp_m(\varphi)$ for all $\varphi \in \mathcal{D}_K$. Hence, $p|_{\mathcal{D}_K}$ is continuous, so $p \in \mathcal{P}$. \square

Corollary 1.34. *Every differentiation $D^\alpha : \mathcal{D}(\Omega) \rightarrow \mathcal{D}(\Omega)$ is continuous.*

Proof. Let $\varphi_j \rightarrow 0$ in $\mathcal{D}(\Omega)$. Then $\{\varphi_j\} \subseteq \mathcal{D}_K$ for some compact $K \subset \Omega$ and $\|\varphi_j\|_{C^k} \rightarrow 0$ for each k . So $\{D^\alpha \varphi_j\} \subset \mathcal{D}_K$ and $\|D^\alpha \varphi_j\|_{C^k} \leq \|\varphi_j\|_{C^{k+|\alpha|}} \rightarrow 0$ for each k . So $D^\alpha \varphi_j \rightarrow 0$ in $\mathcal{D}(\Omega)$, hence D^α is continuous. \square

Lemma 1.35. *$\mathcal{D}(\Omega)$ is not metrizable.*

Using Theorem 1.15, it follows that \mathcal{P} is uncountable. Hence, we added a huge number of seminorms when we enlarged \mathcal{P}^∞ to make $\mathcal{D}(\Omega)$ complete.

5. This is general topology: if X, Y are topological spaces, X is metrizable, then $f : X \rightarrow Y$ is continuous if it preserves convergent sequences. To see this, suppose on the contrary that f is not continuous. Then there is an open $U \subset Y$ such that $f^{-1}(U)$ is not open. So $\exists x \in f^{-1}(U)$ such that $B_{x,\varepsilon}(d) \not\subseteq f^{-1}(U)$ for any $\varepsilon > 0$. So for $\varepsilon = 1/j \exists x_j \in B_{x,1/j}(d)$ but $f(x_j) \notin U$. Since $f(x) \in U$, we thus get $x_j \rightarrow x$ but $f(x_j) \not\rightarrow f(x)$.

Proof. We first note that \mathcal{D}_K , for $K \subset \Omega$, is closed in $\mathcal{D}(\Omega)$, since it is complete and its topology is just the subspace topology inherited from $\mathcal{D}(\Omega)$ by Lemma 1.30.

Next, \mathcal{D}_K has empty interior, relative to $\mathcal{D}(\Omega)$. Indeed, given $\varepsilon > 0$, $p \in \mathcal{P}$ and $\varphi \in \mathcal{D}_K$, we may choose a smooth nonzero ψ with support in $\Omega \setminus K$ (see Lemma 1.24). Let $\tilde{\psi} = \psi$ if $p(\psi) = 0$ and $\tilde{\psi} = \frac{1}{p(\psi)}\psi$ otherwise. Then, $f := \varphi + \frac{\varepsilon}{2}\tilde{\psi}$ satisfies $f \in \mathcal{D}(\Omega)$ and $p(\varphi - f) < \varepsilon$. Hence, $B_{\varphi, \varepsilon}(p) \not\subseteq \mathcal{D}_K$. Since p , ε and φ are arbitrary, \mathcal{D}_K has empty interior in $\mathcal{D}(\Omega)$.

Finally, let K_j be as in Lemma 1.20. Then $\mathcal{D}(\Omega) = \cup_j \mathcal{D}_{K_j}$. Since $\mathcal{D}(\Omega)$ is complete, it follows from the Baire category theorem that $\mathcal{D}(\Omega)$ is not metrizable.⁶ \square

Remark 1.36. In general, if $X_1 \subset X_2 \subset \dots$ are locally convex spaces, then the inductive limit topology on the union $X = \cup_j X_j$ is the finest topology that leaves the embeddings $X_j \hookrightarrow X$ continuous. If each X_j is a Fréchet space, we call X an *LF space*. The reader can find more details in [28] and [37].

For example, all the results of this section apply to the space of compactly supported L^p functions, which is defined by $L^p_{\text{comp}}(\Omega) = \cup_{K \in \Omega} L^p(K)$.

1.5 Distributions

Definition 1.37. A *distribution* (or *generalized function*) on Ω is a continuous linear functional on $\mathcal{D}(\Omega)$. The space of distributions on Ω is denoted by $\mathcal{D}'(\Omega)$.

Given $T \in \mathcal{D}'(\Omega)$, we often denote

$$\langle T, \varphi \rangle := T(\varphi), \quad \varphi \in \mathcal{D}(\Omega).$$

Lemma 1.38. A linear functional $T : \mathcal{D}(\Omega) \rightarrow \mathbb{C}$ is in $\mathcal{D}'(\Omega)$ iff any of the following conditions holds:

- a) $\varphi_j \rightarrow 0$ in $\mathcal{D}(\Omega)$ implies $T(\varphi_j) \rightarrow 0$.
- b) For any compact $K \subset \Omega$, there exist $m, C > 0$ such that

$$|T(\varphi)| \leq C \|\varphi\|_{C^m} \quad \text{for } \varphi \in \mathcal{D}_K. \quad (\dagger)$$

Proof. Follows directly from Theorem 1.33. \square

Definition 1.39. Let $T \in \mathcal{D}'(\Omega)$. If (\dagger) holds with the same m for all compact $K \subset \Omega$ (but C possibly depending on K), then the smallest such m is called the *order* of T . If no m will do for all K , we say that T has *infinite order*.

Example 1.40. 1. Define

$$\delta(\varphi) := \varphi(0) \quad \text{for } \varphi \in \mathcal{D}(\mathbb{R}^n).$$

Then δ is a distribution of order 0 since $|\delta(\varphi)| \leq \|\varphi\|_{C^0}$.

2. Given $u \in L^1_{\text{loc}}(\Omega)$, let $T_u(\varphi) := \int u\varphi$. Then T_u is a distribution of order 0 since

$$|T_u(\varphi)| \leq \|u\|_{L^1(K)} \|\varphi\|_{C^0} \quad \text{for } \varphi \in \mathcal{D}_K.$$

In particular, each $u \in C(\Omega)$ defines a distribution T_u . We show in the next chapter that $u \mapsto T_u$ embeds $L^1_{\text{loc}}(\Omega)$ in $\mathcal{D}'(\Omega)$.

6. The reader can find a different proof in [15] using sequences, which works if $\Omega = \mathbb{R}$ and more generally if $\Omega = \mathbb{R}^n$.

We now define the derivative of a distribution. To motivate this, note that if $u \in C^\infty(\mathbb{R})$ and $\varphi \in \mathcal{D}(\mathbb{R})$, then integrating by parts we have

$$\int u'(x)\varphi(x) dx = - \int u(x)\varphi'(x) dx$$

so that $T_{u'}(\varphi) = -T_u(\varphi')$. So if we think of $T_{u'}$ as the derivative of T_u , then the following definition is quite natural.

Definition 1.41 (Differentiation). Given $T \in \mathcal{D}'(\Omega)$ and $\alpha \in \mathbb{N}_0^n$, we define

$$(D^\alpha T)(\varphi) := (-1)^{|\alpha|} T(D^\alpha \varphi) \quad \text{for } \varphi \in \mathcal{D}(\Omega).$$

Lemma 1.42. Let $T \in \mathcal{D}'(\Omega)$ and $\alpha, \beta \in \mathbb{N}_0^n$. Then

1. $D^\alpha T \in \mathcal{D}'(\Omega)$.
2. $D^\alpha D^\beta T = D^{\alpha+\beta} T = D^\beta D^\alpha T$.

Proof. Let $T \in \mathcal{D}'(\Omega)$ and $K \subset \Omega$ compact, say $|T(\psi)| \leq C\|\psi\|_{C^m}$ for $\varphi \in \mathcal{D}_K$. Then

$$|(D^\alpha T)(\varphi)| = |T(D^\alpha \varphi)| \leq C\|D^\alpha \varphi\|_{C^m} \leq C\|\varphi\|_{C^{m+|\alpha|}}.$$

Hence, $D^\alpha T \in \mathcal{D}'(\Omega)$. Next,

$$\begin{aligned} (D^\alpha D^\beta T)(\varphi) &= (-1)^{|\alpha|} (D^\beta T)(D^\alpha \varphi) = (-1)^{|\alpha|+|\beta|} T(D^\beta D^\alpha \varphi) \\ &= (-1)^{|\alpha+\beta|} T(D^{\alpha+\beta} \varphi) = (D^{\alpha+\beta} T)(\varphi). \quad \square \end{aligned}$$

Example 1.43. $\delta'(\varphi) = -\delta(\varphi') = -\varphi'(0)$.

Definition 1.44 (Multiplication by functions). Let $T \in \mathcal{D}'(\Omega)$ and $f \in C^\infty(\Omega)$. We define

$$(fT)(\varphi) := T(f\varphi) \quad \text{for } \varphi \in \mathcal{D}(\Omega).$$

Lemma 1.45. If $T \in \mathcal{D}'(\Omega)$ and $f \in C^\infty(\Omega)$, then $fT \in \mathcal{D}'(\Omega)$.

Proof. Let $K \subset \Omega$ be compact and assume $\exists C, m$ such that $|T(\varphi)| \leq C\|\varphi\|_{C^m}$ on \mathcal{D}_K . Then $|(fT)(\varphi)| = |T(f\varphi)| \leq C\|f\varphi\|_{C^m}$. But $D^\alpha(f\varphi) = \sum_{\beta \leq \alpha} c_{\alpha\beta} (D^{\alpha-\beta} f)(D^\beta \varphi)$ for some $c_{\alpha\beta}$ (Leibniz formula). Hence, $\|f\varphi\|_{C^m} \leq C_f \|\varphi\|_{C^m}$ on \mathcal{D}_K . Thus, $|(fT)(\varphi)| \leq CC_f \|\varphi\|_{C^m}$ on \mathcal{D}_K and $fT \in \mathcal{D}'(\Omega)$. \square

Definition 1.46 (Sequences of distributions). We endow $\mathcal{D}'(\Omega)$ with the weak*-topology. Hence, we say that $T_j \rightarrow T$ in $\mathcal{D}'(\Omega)$ if $T_j(\varphi) \rightarrow T(\varphi)$ for all $\varphi \in \mathcal{D}(\Omega)$.

Example 1.47. Let $u_j(x) = \sin(jx)$. Then $u_j \rightarrow 0$ in $\mathcal{D}'(\mathbb{R})$ (where we abused notation and wrote u_j instead of T_{u_j}). Indeed, if $\varphi \in \mathcal{D}(\mathbb{R})$, then

$$\int \sin(jx)\varphi(x) dx = \frac{1}{j} \int \cos(jx)\varphi'(x) dx \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

Theorem 1.48. Suppose $\{T_j\} \subseteq \mathcal{D}'(\Omega)$ and $\lim_{j \rightarrow \infty} T_j(\varphi)$ exists (as a complex number) for every $\varphi \in \mathcal{D}(\Omega)$. Define $T(\varphi) := \lim_{j \rightarrow \infty} T_j(\varphi)$. Then $T \in \mathcal{D}'(\Omega)$, and for every $\alpha \in \mathbb{N}_0^n$,

$$D^\alpha T_j \rightarrow D^\alpha T \quad \text{in } \mathcal{D}'(\Omega)$$

Proof. Let $K \subset \Omega$ be compact. Then $\lim_{j \rightarrow \infty} T_j(\varphi)$ exists for every $\varphi \in \mathcal{D}_K$. Since \mathcal{D}_K is Fréchet, by a well-known application of the uniform boundedness principle (which also holds in Fréchet spaces; see [32]), we know that $T|_{\mathcal{D}_K}$ is continuous⁷. Hence, $T \in \mathcal{D}'(\Omega)$. Moreover, if $\varphi \in \mathcal{D}(\Omega)$,

$$\lim_{j \rightarrow \infty} (D^\alpha T_j)(\varphi) = (-1)^{|\alpha|} \lim_{j \rightarrow \infty} T_j(D^\alpha \varphi) = (-1)^{|\alpha|} T(D^\alpha \varphi) = (D^\alpha T)(\varphi). \quad \square$$

Remark 1.49. Let $\varphi \in \mathcal{D}(\Omega)$ and suppose $g_j \rightarrow g$ in $C^\infty(\Omega)$. Then $\varphi g_j \rightarrow \varphi g$ in $\mathcal{D}(\Omega)$. Indeed, by hypothesis, $p_{r,s}(g_j - g) \rightarrow 0$ for all r, s . Let $K = \text{supp } \varphi$. Then given $p \in \mathcal{P}$, $\exists c, p_m$ with $p \leq c p_m$, so $p(\varphi g_j - \varphi g) \leq c \|\varphi(g_j - g)\|_{C^m(K)} \leq c \|\varphi(g_j - g)\|_{C^m(K_r)}$ for any $K_r \supseteq K$. Using Leibniz, we thus get $p(\varphi g_j - \varphi g) \leq c c_\varphi \|g_j - g\|_{C^m(K_r)} = c c_\varphi p_{r,m}(g_j - g) \rightarrow 0$. As $p \in \mathcal{P}$ is arbitrary, we get $\varphi g_j \rightarrow \varphi g$.

Theorem 1.50. *If $T_j \rightarrow T$ in $\mathcal{D}'(\Omega)$ and $g_j \rightarrow g$ in $C^\infty(\Omega)$, then $g_j T_j \rightarrow gT$ in $\mathcal{D}'(\Omega)$.*

Proof. Fix $\varphi \in \mathcal{D}(\Omega)$ and define a bilinear functional B_φ on $C^\infty(\Omega) \times \mathcal{D}'(\Omega)$ by

$$B_\varphi(g, T) := (gT)(\varphi) = T(g\varphi).$$

Then B_φ is separately continuous (use Remark 1.49). By another application of the uniform boundedness principle (see [32] or [28] in the Banach case), it follows that B is jointly continuous, i.e. $B_\varphi(g_j, T_j) \rightarrow B_\varphi(g, T)$. Hence, $(g_j T_j)(\varphi) \rightarrow (gT)(\varphi)$. \square

1.6 PoU and sheaf structure

Definition 1.51. Let $T \in \mathcal{D}'(\Omega)$ and let $\omega \subset \Omega$ be open. We define $T|_\omega : \mathcal{D}(\omega) \rightarrow \mathbb{C}$ by $T|_\omega(\varphi) := T(\varphi)$ for $\varphi \in \mathcal{D}(\omega)$.

Lemma 1.52. *If $T \in \mathcal{D}'(\Omega)$ and $\omega \subset \Omega$, then $T|_\omega \in \mathcal{D}'(\omega)$.*

Proof. Let $\varphi_j \rightarrow 0$ in $\mathcal{D}(\omega)$. Then $\{\varphi_j\} \subset K$ for some compact $K \subset \omega$ and $\varphi_j \rightarrow 0$ in \mathcal{D}_K . Hence, $\varphi_j \rightarrow 0$ in $\mathcal{D}(\Omega)$. Hence, $T|_\omega(\varphi_j) = T(\varphi_j) \rightarrow 0$. Thus, $T|_\omega \in \mathcal{D}'(\omega)$. \square

Example 1.53. Let $f \in L^1_{\text{loc}}(\Omega)$. Then $T_f|_\omega = 0$ iff $f(x) = 0$ for a.e. $x \in \omega$.⁸

Restrictions allow us to discuss distributions *locally*. The aim of this section is to show that, given an open cover $\{\omega_\alpha\}$ of Ω , together with a set of “compatible” distributions $T_\alpha \in \mathcal{D}'(\omega_\alpha)$, we can define a distribution *globally* by *gluing* the T_α together. This is a widely used idea in modern mathematics, e.g. algebraic and differential geometry. We start with the following theorem.

Theorem 1.54 (Partition of unity). *Let $\Gamma = \{\omega_\alpha\}$ be a collection of open sets with $\Omega = \cup_\alpha \omega_\alpha$. Then there exists $\{\psi_i\} \subseteq \mathcal{D}(\Omega)$ with $\psi_i \geq 0$ such that*

- (a) *each ψ_i has its support in some $\omega_\alpha \in \Gamma$,*
- (b) *$\sum_{i=1}^\infty \psi_i(x) = 1$ for every $x \in \Omega$,*
- (c) *$\forall K \Subset \Omega \exists m \in \mathbb{N}$ and W open such that $W \supset K$ and*

$$\psi_1(x) + \dots + \psi_m(x) = 1 \quad \text{for all } x \in W.$$

7. This can also be found in [22, Lemma 4.9-5] in the Banach case.

8. Here we used the nontrivial fact that if $f, g \in L^1_{\text{loc}}(\omega)$, then $T_f = T_g$ iff $f = g$ a.e. on ω . This generalizes the du Bois-Reymond Lemma. We shall prove this next chapter.

We call $\{\psi_i\}$ a *locally finite partition of unity in Ω* subordinate to Γ . We say it is locally finite because, by (b) and (c), each $x \in \Omega$ has a neighborhood which intersects the supports of only finitely many ψ_i (take $K = \{x\}$, then $\forall i > m$, $\psi_i(y) = 0$ in W).

Proof. Let S be a countable dense subset of Ω . Let $\{C_1, C_2, \dots\}$ be the set of all closed balls C_i with center $p_i \in S$, radius $r_i \in \mathbb{Q}$, such that $C_i \subset \omega_\alpha$ for some $\omega_\alpha \in \Gamma$. Let $V_i \subset C_i$ be the open ball with center p_i and radius $r_i/2$. By density of S , we have $\Omega = \cup V_i$.

Using Lemma 1.24, by scaling and translating, we can find $\phi_i \in \mathcal{D}(\Omega)$ such that $0 \leq \phi_i \leq 1$, $\phi_i = 1$ on V_i and $\phi_i = 0$ outside C_i . Define $\psi_1 = \phi_1$, and inductively,

$$\psi_{i+1} = (1 - \phi_1) \cdots (1 - \phi_i) \phi_{i+1} \quad (i \geq 1). \quad (\ddagger)$$

Clearly, $\psi_i = 0$ outside C_i , so $\text{supp } \psi_i \subset C_i \subset \omega_\alpha$ and (a) is true. Next, we have

$$\psi_1 + \dots + \psi_i = 1 - (1 - \phi_1) \cdots (1 - \phi_i). \quad (\ddagger\ddagger)$$

Indeed, $(\ddagger\ddagger)$ is trivial for $i = 1$. If it holds at i , then adding (\ddagger) and $(\ddagger\ddagger)$, we obtain $(\ddagger\ddagger)$ for $i + 1$ in place of i . Hence, $(\ddagger\ddagger)$ holds for each i . Since $\phi_i = 1$ in V_i , it follows that

$$\psi_1(x) + \dots + \psi_m(x) = 1 \quad x \in V_1 \cup \dots \cup V_m.$$

This gives (b). Finally, if K is compact, $K \subset V_1 \cup \dots \cup V_m$ for some m , so (c) follows. \square

Definition 1.55. Let X be a topological space. A *presheaf of sets \mathcal{F} on X* is a rule which assigns to each open $U \subset X$ a set $\mathcal{F}(U)$, and to each inclusion $V \subset U$ a map $\rho_V^U : \mathcal{F}(U) \rightarrow \mathcal{F}(V)$ such that $\rho_U^U = \text{id}_{\mathcal{F}(U)}$, and whenever $W \subset V \subset U$, we have $\rho_W^U = \rho_W^V \circ \rho_V^U$.

The elements of $\mathcal{F}(U)$ are called *sections of \mathcal{F} over U* . We denote $s|_V := \rho_V^U(s)$ if $s \in \mathcal{F}(U)$ and $V \subset U$.

A *sheaf of sets \mathcal{F} on X* is a presheaf of sets with the following additional property: given an open cover $U = \cup_\alpha U_\alpha$ and some sections $s_\alpha \in \mathcal{F}(U_\alpha)$, if $\forall \alpha, \beta$ we have

$$s_\alpha|_{U_\alpha \cap U_\beta} = s_\beta|_{U_\alpha \cap U_\beta},$$

then there exists a unique section $s \in \mathcal{F}(U)$ such that $s|_{U_\alpha} = s_\alpha$ for each α .

Example 1.56. Let X, Y be topological spaces. Given an open $U \subset X$, let $\mathcal{F}(U) := C(U, Y)$ be the set of continuous functions from U to Y , with the usual restriction operation. Then \mathcal{F} is a sheaf. Indeed, $\rho_V^U(f) = f|_V = f$ if $f \in C(U, Y)$, we have $f|_W = (f|_V)|_W$ whenever $W \subset V \subset U$, so \mathcal{F} is a presheaf. Suppose $U = \cup U_\alpha$ is an open cover, and some $f_\alpha \in C(U_\alpha, Y)$ satisfy for each α, β that $f_\alpha|_{U_\alpha \cap U_\beta} = f_\beta|_{U_\alpha \cap U_\beta}$. Then we define $f : U \rightarrow Y$ by $f(u) := f_\alpha(u)$ if $u \in U_\alpha$. This is well defined by our assumptions. Moreover, $f|_{U_\alpha} = f_\alpha$ by definition and each f_α is continuous, so f is continuous. Finally, for uniqueness, suppose \tilde{f} also satisfies the gluing axiom. Then given $x \in U$, say $x \in U_\alpha$, we have $\tilde{f}(x) = f_\alpha(x) = f(x)$.

Theorem 1.57. *Given an open $\Omega \subseteq \mathbb{R}^n$, let $\mathcal{F}(\Omega) := \mathcal{D}'(\Omega)$. Then \mathcal{F} is a sheaf.*

Proof. Again, \mathcal{F} is clearly a presheaf. Now let $\Gamma = \{\omega_\alpha\}$ be an open cover of Ω and assume there is for each α some $T_\alpha \in \mathcal{D}'(\omega_\alpha)$, such that $T_\alpha|_{\omega_\alpha \cap \omega_\beta} = T_\beta|_{\omega_\alpha \cap \omega_\beta}$.

Let $\{\psi_i\}$ be a partition of unity subordinate to Γ , and associate to each i a set $\omega_{\alpha_i} \in \Gamma$ such that $\omega_{\alpha_i} \supset \text{supp } \psi_i$.

If $\varphi \in \mathcal{D}(\Omega)$, then $\varphi = \sum \psi_i \varphi$. Since φ has compact support, only finitely many terms in the sum are nonzero. Define

$$T(\varphi) := \sum_{i=1}^{\infty} T_{\alpha_i}(\psi_i \varphi).$$

To see that $T \in \mathcal{D}'(\Omega)$, let $\varphi_j \rightarrow 0$ in $\mathcal{D}(\Omega)$. Then $\{\varphi_j\} \subseteq \mathcal{D}_K$ for some compact $K \subset \Omega$. Choose m as in Theorem 1.54(c). Then $T(\varphi_j) = \sum_{i=1}^m T_{\alpha_i}(\psi_i \varphi_j)$. But $\psi_i \varphi_j \rightarrow 0$ in $\mathcal{D}(\omega_{\alpha_i})$ as $j \rightarrow \infty$ by Leibniz rule. Hence, $T(\varphi_j) \rightarrow 0$, so $T \in \mathcal{D}'(\Omega)$.

Next, let $\varphi \in \mathcal{D}(\omega_{\alpha})$. Then $\psi_i \varphi \in \mathcal{D}(\omega_{\alpha_i} \cap \omega_{\alpha})$, so by hypothesis $T_{\alpha_i}(\psi_i \varphi) = T_{\alpha}(\psi_i \varphi)$. Hence, $T(\varphi) = \sum T_{\alpha}(\psi_i \varphi) = T_{\alpha}(\sum \psi_i \varphi) = T_{\alpha}(\varphi)$. Thus, $T|_{\omega_{\alpha}} = T_{\alpha}$.

Thus, T exists. For uniqueness, if \tilde{T} also satisfies the requirements, then given $\varphi \in \mathcal{D}(\Omega)$ we have $\tilde{T}(\varphi) = \tilde{T}(\sum \psi_i \varphi) = \sum \tilde{T}(\psi_i \varphi) = \sum T_{\alpha_i}(\psi_i \varphi) = T(\varphi)$. \square

1.7 Distributions with compact support

Definition 1.58. The *support* of a distribution $T \in \mathcal{D}'(\Omega)$ is the set

$$\text{supp } T := \Omega \setminus \left(\bigcup \{ \omega \subseteq \Omega \text{ open} : T|_{\omega} = 0 \} \right).$$

Lemma 1.59. Let $T \in \mathcal{D}'(\Omega)$. Then $T|_{\Omega \setminus \text{supp } T} = 0$.

Proof. Let $W = \Omega \setminus \text{supp } T$, $\Gamma = \{ \omega \subseteq \Omega \text{ open} : T|_{\omega} = 0 \}$ and $\{\psi_i\}$ a partition of unity in W subordinate to Γ . Given $\varphi \in \mathcal{D}(W)$, we have $\varphi = \sum \varphi \psi_i$, and only finitely many terms in the sum are nonzero. Hence, $T(\varphi) = \sum T(\psi_i \varphi) = 0$, since each ψ_i has its support in some $\omega \in \Gamma$. \square

Example 1.60. Given $x \in \Omega$, define $\delta_x \in \mathcal{D}'(\Omega)$ by $\delta_x(\varphi) = \varphi(x)$. Then $\text{supp } \delta_x = \{x\}$.

One can show that conversely, if $\text{supp } T = \{x\}$, then $T = P(D)\delta_x$ for some differential operator $P(D)$; see e.g. [32].

Definition 1.61. Let $\mathcal{D}'_c(\Omega)$ be the subset of $\mathcal{D}'(\Omega)$ consisting of distributions whose support is a compact subset of Ω .

Lemma 1.62. For any $T \in \mathcal{D}'(\mathbb{R}^n)$ there exists $\{T_j\} \subseteq \mathcal{D}'_c(\mathbb{R}^n)$ with $T_j \rightarrow T$.

Proof. Use Lemma 1.24 to find $\phi \in C^\infty(\mathbb{R}^n)$ with $\phi(x) = 1$ if $|x| < \frac{1}{2}$ and $\phi(x) = 0$ if $|x| > 1$. Let $\phi_j(x) = \phi(x/j)$ and $T_j = \phi_j T$. Then $\text{supp } T_j \subset \overline{B}_j = \{x \in \mathbb{R}^n : |x| \leq j\}$, and given $\psi \in \mathcal{D}(\mathbb{R}^n)$, $T_j(\psi) = T(\phi_j \psi) \rightarrow T(\psi)$. \square

In the following theorem, $\mathcal{E}(\Omega)$ is the space $C^\infty(\Omega)$ endowed with the seminorms $\{p_{j,k}\}$ from Lemma 1.21.

Theorem 1.63. (1) Any $T \in \mathcal{D}'_c(\Omega)$ has finite order.

(2) $\mathcal{E}'(\Omega) = \mathcal{D}'_c(\Omega)$.

Proof. (1) Let $\psi \in \mathcal{D}(\Omega)$ such that $\psi = 1$ in a neighborhood of $\text{supp } T$. Such ψ can be obtained e.g. using Theorem 1.54. Then $T(\varphi) = T(\psi \varphi)$ for any $\varphi \in \mathcal{D}(\Omega)$. Indeed, $T(\varphi) = T(\psi \varphi + (1 - \psi)\varphi) = T(\psi \varphi) + T((1 - \psi)\varphi) = T(\psi \varphi)$ using Lemma 1.59.

Let $K = \text{supp } \psi$. Then by Lemma 1.38 $\exists m, C > 0$ such that $|T(\varphi)| \leq c_1 \|\varphi\|_{C^m}$ for all $\varphi \in \mathcal{D}_K$. Hence, for any $\varphi \in \mathcal{D}(\Omega)$, we have

$$|T(\varphi)| = |T(\psi \varphi)| \leq c_1 \|\psi \varphi\|_{C^m} \leq c_1 c_2 \|\varphi\|_{C^m}, \quad (*)$$

where we used the Leibniz formula in the last step. In particular, T has order $\leq m$.

- (2) If $T \in \mathcal{E}'(\Omega)$, then by Theorem 1.8 $\exists p_{j,m}$ and $C > 0$ such that $|T(f)| \leq Cp_{j,m}(f)$ for $f \in \mathcal{E}(\Omega)$. Thus, if $\text{supp } f \subset \Omega \setminus K_j$, then $T(f) = 0$. Hence, $\text{supp } T \subset K_j$, so $T \in \mathcal{D}'_c(\Omega)$.

Conversely, let $T \in \mathcal{D}'_c(\Omega)$ and let $\psi \in \mathcal{D}(\Omega)$ as in (1). We define $T(f) := T(\psi f)$ for $f \in \mathcal{E}(\Omega)$. Since $T(\varphi) = T(\psi\varphi)$ for $\varphi \in \mathcal{D}(\Omega)$, this definition extends T from $\mathcal{D}(\Omega)$ to $\mathcal{E}(\Omega)$. Moreover, by (*), $\exists m, C > 0$ such that $|T(f)| \leq C\|\psi f\|_{C^m}$. Let $K = \text{supp } \psi$. By the Leibniz formula $\|\psi f\|_{C^m} = \|\psi f\|_{C^m(K)} \leq c\|f\|_{C^m(K)} \leq cp_{j,m}(f)$ for any $K_j \supset K$. Hence, $|T(f)| \leq cp_{j,m}(f)$, so $T \in \mathcal{E}'(\Omega)$ by Theorem 1.8. \square

1.8 Further results and applications

We end this chapter with a collection of facts about distributions, given without proof.

We first mention that there are some representation theorems for distributions. For example, any $T \in \mathcal{D}'_c(\Omega)$ takes the form $T = \sum_{|\alpha| \leq k} D^\alpha(\mu_\alpha)$ for some complex measures μ_α . Here $\mu_\alpha(\varphi) := \int \varphi d\mu_\alpha$. The proof of this fact uses the Riesz representation theorem (the version which says that the dual of $C(\Omega)$ is a set of finite measures). For details, see [44]. We also mention a representation theorem for “tempered distributions” in Exercise 14, and there is still another such theorem for general $T \in \mathcal{D}'(\Omega)$ in [32].

One deficiency in distribution theory is that we cannot multiply two arbitrary distributions. The problem is deep; in fact Schwartz proved it is impossible to extend the product of continuous functions in a straightforward way [6]. Still, we can define a meaningful product in various special cases. One such case is when the distributions depend on different sets of variables. Thus, if $T_j \in \mathcal{D}'(X_j)$, we define the *tensor product* $T_1 \otimes T_2 \in \mathcal{D}'(X_1 \times X_2)$. In case $T_j = u_j$ are functions, this product gives the function $(u_1 \otimes u_2)(x_1, x_2) = u_1(x_1)u_2(x_2)$.

Any function $K \in C(X_1 \times X_2)$ defines an integral operator $\mathcal{K} : C_c(X_2) \rightarrow C(X_1)$ given by $(\mathcal{K}f)(x_1) = \int K(x_1, x_2)f(x_2) dx_2$. This can be extended as follows: any $K \in \mathcal{D}'(X_1 \times X_2)$ defines a continuous operator $\mathcal{K} : C_c^\infty(X_2) \rightarrow \mathcal{D}'(X_1)$ given by $(\mathcal{K}\varphi)(\psi) := K(\psi \otimes \varphi)$, where $(\psi \otimes \varphi)(x_1, x_2) = \psi(x_1)\varphi(x_2)$. The well known *Schwartz kernel theorem* says that conversely, to any continuous map $\mathcal{K} : C_c^\infty(X_2) \rightarrow \mathcal{D}'(X_1)$, there exists a unique distribution $K \in \mathcal{D}'(X_1 \times X_2)$ such that $(\mathcal{K}\varphi)(\psi) = K(\psi \otimes \varphi)$. See [21] for a proof. Another version concerning bilinear maps on the Schwartz space $\mathcal{S}(\mathbb{R}^d)$ of functions of rapid decay may be found in [28].

The main applications of distribution theory lie in the analysis of partial differential equations. We already mentioned the existence of fundamental solutions in Section 1.1 and their use in solving PDEs. Another important use is in the study of elliptic regularity. Here is an example of the results one gets: let $P(D)$ be an elliptic differential operator (e.g. $-\Delta$). If $v \in C^\infty(\Omega)$ and $u \in \mathcal{D}'(\Omega)$ is a distributional solution to $P(D)u = v$, then $u \in C^\infty(\Omega)$. Thus, u has much more smoothness than initially assumed. The reader can find the basic applications of distributions to PDE in [32]. For much more results, see [21] and [11].

1.9 Exercises

- Let p be a seminorm on a vector space X and let $M = \{x \in X : p(x) \leq 1\}$. Show that
 - M is *convex* : if $x, y \in M$ and $0 \leq t \leq 1$, then $tx + (1-t)y \in M$,
 - M is *balanced* : if $x \in M$ and $|\alpha| \leq 1$, then $\alpha x \in M$,
 - M is *absorbing* : for any $x \in X$, there exists $\alpha > 0$ such that $\alpha^{-1}x \in M$,

$$4. p(x) = \inf\{\alpha > 0 : \alpha^{-1}x \in M\}.$$

2. Let M be a convex, balanced, absorbing subset of a vector space X . Define the *Minkowski functional* of M , $\rho_M : X \rightarrow \mathbb{R}_+$ by $\rho_M(x) = \inf\{\alpha > 0 : \alpha^{-1}x \in M\}$. Show that ρ_M is a seminorm on X .

Hint: given $x, y \in X$ and $\varepsilon > 0$, let $t = \rho_M(x) + \varepsilon$ and $s = \rho_M(y) + \varepsilon$. Then x/t and $y/s \in M$. Using convexity, show that $(x+y)/(s+t) \in M$ and hence $\rho_M(x+y) \leq s+t$.

3. Show that the functions $p_i : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $p_i(x) = |x_i|$ for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ are seminorms.

4. Show that the topologies given by finite separating families of seminorms are actually given by norms.

That is, suppose $\mathcal{P} = \{p_i : i = 1, \dots, n\}$ is a separating family of seminorms on X . Show that $\|x\| := \max_{1 \leq i \leq n} p_i(x)$ is a norm on X that induces the same topology as the family \mathcal{P} .

More generally, show that for any norm N on \mathbb{R}^n , the map $\|x\|_N := N(p_1(x), \dots, p_n(x))$ is a norm on X which induces the same topology as \mathcal{P} .

5. Let X be a Hausdorff space and endow $C(X)$ with the family of seminorms $\{p_x : x \in X\}$, where $p_x(f) = |f(x)|$. Check that $f_n \rightarrow f$ is equivalent to pointwise convergence.

It can be shown that this topology is not metrizable in general⁹. In particular, one cannot find a norm on $C[0, 1]$ which induces the topology of pointwise convergence.

6. Show that the topology induced by a family of seminorms \mathcal{P} on a vector space X is the weakest topology in which every $p \in \mathcal{P}$ is continuous, and in which the operation of addition $(x, y) \mapsto x + y$ from $X \times X \rightarrow X$ is continuous. This means that for any neighborhood V of $x + y$, we can find neighborhoods V_1 of x and V_2 of y such that $V_1 + V_2 \subseteq V$. Hence, if x' is near x and y' is near y , then $x' + y'$ will be near $x + y$.

In general, if X is a set and \mathcal{F} is a family of mappings $f : X \rightarrow Y_f$, where Y_f is a topological space, the family \mathcal{F} induces a topology on X by saying that a set is open in X if it is a union of finite intersections of sets of the form $f^{-1}(V)$, with $f \in \mathcal{F}$ and V open in Y_f . We call this the *\mathcal{F} -weak topology* on X . One checks that this is the weakest topology on X which makes every $f \in \mathcal{F}$ continuous.

In our construction of locally convex spaces, we took X a vector space, $\mathcal{F} = \mathcal{P}$ a family of seminorms and $Y_f = \mathbb{R}$ for all f . Here is another example : let $\{X_\alpha\}_{\alpha \in I}$ be a collection of topological spaces, and let $X = \times_{\alpha \in I} X_\alpha$. Define $\pi_\alpha : X \rightarrow X_\alpha$ by $\pi_\alpha : (x_\beta) \mapsto x_\alpha$. Then the *product topology* on X is the weak topology generated by the projections π_α , i.e. it is the weakest topology that makes every π_α continuous.

7. Show that one can convert a seminorm into a norm by factoring out the zero directions.

More precisely, suppose p is a seminorm on a vector space E . Show that $\ker(p) = \{x \in E : p(x) = 0\}$ is a linear subspace of E . Show that the quotient space $E/\ker(p)$ is a normed space, with the norm defined by $\|[x]\| := p(x)$ for $x \in E$.¹⁰

8. Let $K \subset \mathbb{R}^n$ be a compact set. Let $C^m(K) = \{f|_K : f \in C^m(\mathbb{R}^n)\}$.

1. Show that $C(K)$ endowed with the norm $\|f\|_0 = \sup_{x \in K} |f(x)|$ is a Banach space.
2. Show that $C^m(K)$ endowed with the norm $\|f\|_m = \max_{|\alpha| \leq m} \|D^\alpha f\|_0$ is a Banach space.

9. If X is countable, $\{p_x\}$ is countable and separating, so $C(X)$ is metrizable. For uncountable sets like $[0, 1]$, it won't be metrizable.

10. You used this procedure in your undergraduate courses to construct the normed space $L^p(X, \mu)$ from the seminormed space $\mathcal{L}^p(X, \mu)$ of all measurable $f : X \rightarrow \mathbb{C}$ satisfying $(\int_X |f|^p d\mu)^{1/p} < \infty$, $1 \leq p < \infty$.

9. Let

$$\mathcal{S}(\mathbb{R}^n) = \{\varphi \in C^\infty(\mathbb{R}^n) \mid p_{\alpha,\beta}(\varphi) := \sup_{x \in \mathbb{R}^n} |x^\alpha D^\beta \varphi(x)| < \infty \text{ for all } \alpha, \beta \in \mathbb{N}_0^n\}$$

We call $\mathcal{S}(\mathbb{R}^n)$ the *Schwartz space of rapidly decreasing functions*.

Show that $\mathcal{S}(\mathbb{R}^n)$ with the topology induced by the family $\mathcal{P} = \{p_{\alpha,\beta}\}$ is a Fréchet space.

10. Let X be a vector space endowed with a topology τ . We say that X is *normable* if there exists a norm on X which induces τ .

1. Show that if X is normable, then 0 has a bounded neighborhood.

2. Let Ω be a nonempty open subset of \mathbb{R}^n . Let $K_1 \subset K_2 \subset \dots \subset \Omega$ be compact sets such that $\cup_j K_j = \Omega$. Endow $C(\Omega)$ with the family of seminorms $p_i(f) = \sup_{x \in K_i} |f(x)|$.

Show that the semiballs $B_{0,\varepsilon}(p_i) = \{f \in C(\Omega) : p_i(f) < \varepsilon\}$ are not bounded.

3. Conclude that $C(\Omega)$ is not normable.¹¹

11. The topological dual of $\mathcal{S}(\mathbb{R}^n)$ (cf. Exercise 9), denoted $\mathcal{S}'(\mathbb{R}^n)$, is called the *space of tempered distributions*.

1. Given $k \in \mathbb{N}_0$, define $\|\varphi\|_k = \max_{|\alpha+\beta| \leq k} p_{\alpha,\beta}(\varphi)$. Show that $T \in \mathcal{S}'(\mathbb{R}^n)$ iff there exists k such that $|T(\varphi)| \leq C\|\varphi\|_k$ for all $\varphi \in \mathcal{S}(\mathbb{R}^n)$.

2. Show that if $T \in \mathcal{S}'(\mathbb{R}^n)$, then $T|_{\mathcal{D}(\mathbb{R}^n)} \in \mathcal{D}'(\mathbb{R}^n)$. Hence, any tempered distribution is a distribution.

Hint: you can show for example that $\varphi_j \rightarrow 0$ in $\mathcal{D}(\mathbb{R}^n)$ implies $\varphi_j \rightarrow 0$ in $\mathcal{S}(\mathbb{R}^n)$.

12. 1. Show that $\mathcal{S}(\mathbb{R}^n) \subset \bigcap_{1 \leq p \leq \infty} L^p(\mathbb{R}^n)$, and $\|\varphi\|_{L^p} \leq (2\pi)^n \|\varphi\|_{2n}$ for any $\varphi \in \mathcal{S}(\mathbb{R}^n)$.¹²

Hint: Recall that $\int_{\mathbb{R}} \frac{1}{1+x^2} dx = \pi$.

2. Show that for any $u \in L^p(\mathbb{R}^n)$ and $\varphi \in \mathcal{S}(\mathbb{R}^n)$, one has $u\varphi \in L^1(\mathbb{R}^n)$, and $\|u\varphi\|_{L^1} \leq (2\pi)^n \|u\|_{L^p} \|\varphi\|_{2n}$.

3. Given $u \in L^p(\mathbb{R}^n)$, define $T_u(\varphi) = \int u(x)\varphi(x) dx$ for $\varphi \in \mathcal{S}(\mathbb{R}^n)$. Show that $T_u \in \mathcal{S}'(\mathbb{R}^n)$.

13. Define $PV(\frac{1}{x}) : \mathcal{S}(\mathbb{R}) \rightarrow \mathbb{C}$ by $PV(\frac{1}{x}) : \varphi \mapsto \lim_{\varepsilon \downarrow 0} \int_{|x| \geq \varepsilon} \frac{1}{x} \varphi(x) dx$. We call $PV(\frac{1}{x})$ the *Cauchy principal value*. Show that $PV(\frac{1}{x})$ is well defined on $\mathcal{S}(\mathbb{R})$, and that $|PV(\frac{1}{x})(\varphi)| \leq 2(p_{0,1}(\varphi) + p_{1,0}(\varphi))$. Conclude that $PV(\frac{1}{x}) \in \mathcal{S}'(\mathbb{R})$.

A well-known formula says that $\lim_{\varepsilon \downarrow 0} \frac{1}{x-x_0+i\varepsilon} = PV(\frac{1}{x-x_0}) - i\pi\delta(x-x_0)$. It can be shown that this holds in the sense of distributions.

14. Let $g(x) = \begin{cases} x & \text{if } x \geq 0, \\ 0 & \text{if } x \leq 0. \end{cases}$ for $x \in \mathbb{R}$. Clearly, g is continuous but not everywhere

differentiable in the classical sense.

Let $T_g(\varphi) = \int g(x)\varphi(x) dx$ for $\varphi \in \mathcal{S}(\mathbb{R})$.

1. Show that $T_g \in \mathcal{S}'(\mathbb{R})$, and that its distributional derivative $T'_g = T_H$, where H is

$$\text{the Heaviside function } H(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

11. We can also show that $C^\infty(\Omega)$ is not normable. The idea is as follows: recall that by the Arzelà-Ascoli theorem, a subset of $C(K)$ is compact iff it is closed, bounded and equicontinuous. Using Arzelà-Ascoli, one can show that any closed bounded subset of $C^\infty(\Omega)$ is compact. Recalling that the unit ball of an infinite-dimensional normed space is never compact (theorem), we conclude that $C^\infty(\Omega)$ is not normable. The same conclusion holds for \mathcal{D}_K .

12. This shows in particular that the embedding $\mathcal{S}(\mathbb{R}^n) \hookrightarrow L^p(\mathbb{R}^n)$ is continuous.

2. Show that $T'_H = \delta$.

Thus, $\delta = T''_g$, and if we identify T_g with g , we see that the nonfunction δ is the second derivative of a continuous function. This is typical of tempered distributions: any $T \in \mathcal{S}'(\mathbb{R}^d)$ takes the form $T = D^\beta T_g$ for some polynomially bounded continuous function g and some $\beta \in \mathbb{N}_0^d$. A related, more complicated result holds for $T \in \mathcal{D}'(\mathbb{R}^d)$, see [32].

15. Let (T_n) be a sequence in $\mathcal{S}'(\mathbb{R}^d)$. We say that $T_n \rightarrow T \in \mathcal{S}'(\mathbb{R}^d)$ if $T_n(\varphi) \rightarrow T(\varphi)$ for every $\varphi \in \mathcal{S}(\mathbb{R}^d)$. This is just convergence in the weak*-topology.

Define $T_n(\varphi) = \frac{n}{2} \int_{-1/n}^{1/n} \varphi(x) dx$ on $\mathcal{S}(\mathbb{R})$. Show that $T_n \rightarrow \delta$. This is an example of a *delta sequence*.

Chapter 2

Fourier analysis

2.1 Convolutions

In this chapter we shall discuss the Fourier transform and some of its applications.

We begin with some facts from measure theory, then we define convolutions and give their basic properties. Let us first take care of differentiation under integral sign.¹

Lemma 2.1. *Let (X, μ) be a measure space, $I \subseteq \mathbb{R}$ open and $f : X \times I \rightarrow \mathbb{C}$ such that $f(\cdot, t) \in L^1(\mu)$ for each $t \in I$. Let $F(t) = \int_X f(x, t) d\mu(x)$.*

- (a) *If $f(x, \cdot) \in C(I)$ for each x and there exists $g \in L^1(\mu)$ such that $|f(x, t)| \leq g(x)$ for all x, t , then $F \in C(I)$.*
- (b) *If $\partial_t f$ exists and there is a $g \in L^1(\mu)$ such that $|\partial_t f(x, t)| \leq g(x)$ for all x, t , then F is differentiable and $F'(t) = \int \partial_t f(x, t) d\mu(x)$.*

Proof. Fix $t_0 \in I$, let $\{t_n\} \subset I$ such that $t_n \rightarrow t_0$. Then $f_n(x) = f(x, t_n) \rightarrow f(x, t_0)$, so by the dominated convergence theorem, $F(t_n) \rightarrow F(t_0)$, which gives (a). Moreover, if $h_n(x) = \frac{f(x, t_n) - f(x, t_0)}{t_n - t_0}$, then $h_n(x) \rightarrow \partial_t f(x, t_0)$, so $\partial_t f$ is measurable in x . Furthermore, by the mean value theorem, $|h_n(x)| \leq \sup_{t \in I} |\partial_t f(x, t)| \leq g(x)$, so applying dominated convergence again, we get

$$F'(t_0) = \lim_{n \rightarrow \infty} \frac{F(t_n) - F(t_0)}{t_n - t_0} = \lim_{n \rightarrow \infty} \int h_n(x) d\mu(x) = \int \partial_t f(x, t) d\mu(x). \quad \square$$

The classical Minkowski inequality says that the L^p norm of a sum is bounded by the sum of L^p norms. The following is a generalization from sums to arbitrary integrals. The proof is very similar.

Lemma 2.2 (Generalized Minkowski inequality). *Let $(X, \mu), (Y, \nu)$ be σ -finite measure spaces and suppose $f : X \times Y \rightarrow \mathbb{C}$ is measurable.*

If $1 \leq p \leq \infty$, $f(\cdot, y) \in L^p(\mu)$ for a.e. y , and the function $y \mapsto \|f(\cdot, y)\|_p$ is in $L^1(\nu)$, then the map $x \mapsto \int f(x, y) d\nu(y)$ is in $L^p(\mu)$, and

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

Proof. The map $x \mapsto \int f(x, y) d\nu(y)$ is measurable by classical theorems, so it suffices to prove the inequality.

1. This chapter follows [44] and [14].

For $p = 1$, this follows from Fubini. For $p = \infty$, fix $x \in X$. Then $|f(x, y)| \leq \|f(\cdot, y)\|_\infty$ for any $y \in Y$, so $\int |f(x, y)| d\nu(y) \leq \int \|f(\cdot, y)\|_\infty d\nu(y)$. Since x is arbitrary, the assertion follows. So suppose $1 < p < \infty$, let q be its conjugate exponent (i.e. $\frac{1}{p} + \frac{1}{q} = 1$) and define $F(x) = \int_Y f(x, y) d\nu(y)$. Then for any x ,

$$|F(x)|^p = |F(x)|^{p-1} |F(x)| \leq |F(x)|^{p-1} \int_Y |f(x, y)| d\nu(y),$$

so by Fubini,

$$\begin{aligned} \|F\|_p^p &= \int_X |F(x)|^p d\mu(x) \leq \int_X \left(|F(x)|^{p-1} \int_Y |f(x, y)| d\nu(y) \right) d\mu(x) \\ &= \int_Y \left(\int_X |F(x)|^{p-1} |f(x, y)| d\mu(x) \right) d\nu(y). \end{aligned}$$

Now using Hölder and noting that $(p-1)q = p$, we get

$$\begin{aligned} \|F\|_p^p &\leq \int_Y \left(\int_X |F(x)|^{(p-1)q} d\mu(x) \right)^{1/q} \left(\int_X |f(x, y)|^p d\mu(x) \right)^{1/p} d\nu(y) \\ &= \int_Y \|F\|_p^{p/q} \|f(\cdot, y)\|_p d\nu(y). \end{aligned}$$

If $\|F\|_p = 0$, the lemma is trivially true. If $\|F\|_p \neq 0$, divide by $\|F\|_p^{p/q}$ to get $\|F\|_p = \|F\|_p^{p(1-1/q)} \leq \int_Y \|f(\cdot, y)\|_p d\nu(y)$. \square

The last measure theoretic result we need before moving on, is that translations are continuous in the L^p norm. That is, suppose $f \in L^p(\mathbb{R}^n)$, $y \in \mathbb{R}^n$ and define

$$\tau_y f(x) = f(x - y).$$

Clearly, $\|\tau_y f\|_p = \|f\|_p$ for $1 \leq p \leq \infty$. Moreover,

Lemma 2.3. *If $1 \leq p < \infty$, then translation is continuous in the L^p norm. That is, if $f \in L^p(\mathbb{R}^n)$ and $z \in \mathbb{R}^n$, then $\lim_{y \rightarrow 0} \|\tau_{y+z} f - \tau_z f\|_p = 0$.*

Proof. Since $\tau_{y+z} = \tau_y \tau_z$, by replacing f by $\tau_z f$, we may assume $z = 0$. Suppose first that $g \in C_c(\mathbb{R}^n)$. Then for $|y| \leq 1$, all $\tau_y g$ are supported on a common compact set K , so $\int |\tau_y g - g|^p \leq \max_{x \in K} |g(x - y) - g(x)|^p \text{vol}(K) \rightarrow 0$ as $y \rightarrow 0$, because g is uniformly continuous on K .

Next, if $f \in L^p(\mathbb{R}^n)$, let $\varepsilon > 0$ and choose $g \in C_c(\mathbb{R}^n)$ with $\|f - g\|_p < \varepsilon/3$. Then $\|\tau_y f - f\|_p \leq \|\tau_y(f - g)\|_p + \|\tau_y g - g\|_p + \|g - f\|_p < \frac{2\varepsilon}{3} + \|\tau_y g - g\|_p$ and we know $\|\tau_y g - g\|_p < \varepsilon/3$ for y small enough. \square

Definition 2.4. If f, g are measurable on \mathbb{R}^n , we define their *convolution* $f * g$ by

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

whenever the integral exists.

To understand convolutions, suppose $\rho : \mathbb{R}^n \rightarrow \mathbb{R}$ is measurable with $\rho \geq 0$ and $\int \rho = 1$. Then $\rho(y) dy$ is a probability measure on \mathbb{R}^n , and so is $\rho(y - x) dy$ for any $x \in \mathbb{R}^n$. Moreover, the measure $\rho(y - x) dy$ is just a translation of $\rho(y) dy$ by x . Given a function f , the map $f_\rho(x) = \int f(y) \rho(y - x) dy$, when defined, gives the average of f with respect to the measure $\rho(y - x) dy$. Intuitively, by taking averages, the function f_ρ

becomes more regular than f , because the singularities of f are averaged upon. Moreover, if ρ has a very small support, say an ε -ball around zero, f_ρ will be a good approximation of f . The convolution product $f * \rho$ is defined by $\int f(y) \rho(x - y) dy$, i.e. f_ρ instead of f_ρ , where $\tilde{\rho}(x) := \rho(-x)$. This is just for convenience. For example, one advantage is that we get

$$\rho * f = f * \rho$$

using the change of variables $z = x - y$.

Lemma 2.5. (i) (Young's Inequality). *If $f \in L^1(\mathbb{R}^n)$ and $g \in L^p(\mathbb{R}^n)$, then²*

$$f * g \in L^p(\mathbb{R}^n) \quad \text{and} \quad \|f * g\|_p \leq \|f\|_1 \|g\|_p.$$

(ii) *If $f \in L^1(\mathbb{R}^n)$, $g \in C^k(\mathbb{R}^n)$ and $D^\alpha g$ is bounded for $|\alpha| \leq k$, then*

$$f * g \in C^k(\mathbb{R}^n) \quad \text{and} \quad D^\alpha(f * g) = f * (D^\alpha g).$$

(iii) *If $\rho \in L^1(\mathbb{R}^n)$, $f \in L^p(\mathbb{R}^n)$, $h \in L^q(\mathbb{R}^n)$, where $1 \leq p \leq \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$(\rho * f, h) = (f, \rho^* * h),$$

where $(u, v) = \int u \bar{v}$ and $\rho^*(x) = \overline{\rho(-x)}$.

Proof. (i) Let $h(x, y) = f(y)g(x - y)$. Then $(f * g)(x) = \int h(x, y) dy$, so by Lemma 2.2, $f * g \in L^p(\mathbb{R}^n)$ with $\|f * g\|_p = \|\int h(\cdot, y) dy\|_p \leq \int \|h(\cdot, y)\|_p dy = \int |f(y)| \|\tau_y g\|_p dy = \|f\|_1 \|g\|_p$.

(ii) By hypothesis $\exists M > 0$ with $|f(y)D^\alpha g(x - y)| \leq Mf(y) \in L^1(\mathbb{R}^n)$. So by Lemma 2.1, $\partial_{x_j}(f * g)(x) = \partial_{x_j} \int f(y)g(x - y) dy = \int f(y)\partial_{x_j} g(x - y) dy = [f * (\partial_{x_j} g)](x)$. The result now follows by induction.

(iii) By Fubini, $(\rho * f, h) = (f * \rho, h) = (f, h * \rho^*) = (f, \rho^* * h)$. \square

Remark 2.6. Note that if K is compact and $\bar{B}_{0, \varepsilon} = \{t : |t| \leq \varepsilon\}$, then

$$\left(\text{supp } f \subseteq K \text{ and } \text{supp } g_\varepsilon \subseteq \bar{B}_{0, \varepsilon} \right) \implies \text{supp } g_\varepsilon * f \subseteq K_\varepsilon,$$

where $K_\varepsilon = \{x \in \mathbb{R}^n : \text{dist}(x, K) \leq \varepsilon\} = \{x \in \mathbb{R}^n : |x - y| \leq \varepsilon \text{ for some } y \in K\}$. Indeed, if $|x - y| > \varepsilon$ for all $y \in K$, then $g_\varepsilon(x - y)$ vanishes on K , so $(g_\varepsilon * f)(x) = (f * g_\varepsilon)(x) = \int_K f(y)g_\varepsilon(x - y) dy = 0$.

Definition 2.7. For any function ρ on \mathbb{R}^n , we denote $\rho_\varepsilon(x) := \varepsilon^{-n} \rho(x/\varepsilon)$.

If $\rho \in L^1(\mathbb{R}^n)$ and $\int \rho(x) dx = 1$, we call ρ_ε an *approximate identity*.

If $\rho \in C_c^\infty(\mathbb{R}^n)$, $\int \rho(x) dx = 1$, $\rho \geq 0$, $\rho(x) = \rho(-x)$ and $\rho(x) = 0$ if $|x| \geq 1$, we call ρ_ε a *mollifier*. We constructed such ρ in Chapter 1.

Note that if $\rho \in L^1$, then $\int \rho_\varepsilon$ is independent of ε , since $\int \rho_\varepsilon(x) dx = \int \rho(\varepsilon^{-1}x) \varepsilon^{-n} dx = \int \rho(y) dy$. The following lemma justifies the name ‘‘approximate identity’’.

Lemma 2.8. *Suppose $\rho \in L^1(\mathbb{R}^n)$ and $\int \rho(x) dx = a$. If $f \in L^p(\mathbb{R}^n)$ with $1 \leq p < \infty$, then $\rho_\varepsilon * f \rightarrow af$ in $L^p(\mathbb{R}^n)$ as $\varepsilon \rightarrow 0$.*

2. This shows that $L^1(\mathbb{R}^n)$ is a Banach algebra, if the product is taken to be the convolution.

Proof. Setting $y = \varepsilon z$, we have

$$\begin{aligned} \rho_\varepsilon * f(x) - af(x) &= \int \rho_\varepsilon(y)(f(x-y) - f(x)) \, dy \\ &= \int \rho(z)(f(x-\varepsilon z) - f(x)) \, dz = \int \rho(z)(\tau_{\varepsilon z}f(x) - f(x)) \, dz. \end{aligned}$$

Hence, by Lemma 2.2, $\|\rho_\varepsilon * f - af\|_p \leq \int |\rho(z)| \|\tau_{\varepsilon z}f - f\|_p \, dz$. But for each z , $\|\tau_{\varepsilon z}f - f\|_p$ is bounded by $2\|f\|_p$, and tends to 0 as $\varepsilon \rightarrow 0$ by Lemma 2.3. So the assertion follows by dominated convergence. \square

Corollary 2.9. *For any $1 \leq p < \infty$ and any open $\Omega \subseteq \mathbb{R}^n$, $C_c^\infty(\Omega)$ is dense in $L^p(\Omega)$.*

Proof. Since $C_c(\Omega)$ is dense in $L^p(\Omega)$, it suffices to show that $\overline{C_c^\infty(\Omega)} \supseteq C_c(\Omega)$. Let $f \in C_c(\Omega)$ and let $d = \text{dist}(\text{supp } f, \partial\Omega)$ (with $d := \infty$ if $\Omega = \mathbb{R}^n$). Let ρ_ε be a mollifier. If $\varepsilon < d/2$, we have $\rho_\varepsilon * f \in C_c(\Omega)$ by Remark 2.6. Moreover, $\rho_\varepsilon * f = f * \rho_\varepsilon \in C^\infty(\mathbb{R}^n)$ by Lemma 2.5 (ii). Finally, $\rho_\varepsilon * f \rightarrow f$ in $L^p(\Omega)$ by Lemma 2.8. \square

Corollary 2.10. *Let $f, g \in L^1_{\text{loc}}(\Omega)$. Then $T_f = T_g$ iff $f = g$ a.e. on Ω . In other words, $L^1_{\text{loc}}(\Omega)$ can be embedded in $\mathcal{D}'(\Omega)$, i.e. the map $f \mapsto T_f$ is injective.*

Proof. Let $h \in L^1_{\text{loc}}(\Omega)$ and suppose $T_h = 0$. Then $(h, \varphi) = 0$ for all $\varphi \in C_c^\infty(\Omega)$. If h was in $L^2(\Omega)$, we would deduce that $(h, h) = 0$ by density of $C_c^\infty(\Omega)$ and thus $h = 0$. Unfortunately this may not be the case, so we shall approximate h by a smooth function.

Let $U \subset \Omega$ be open with $\bar{U} \subset \Omega$ and \bar{U} compact. Then $\psi := h|_U \in L^1(U)$.

Let $V \subset U$ be open with $\bar{V} \subset U$. Let ρ_ε be a mollifier. If $\varphi \in C_c^\infty(V)$ and ε is sufficiently small, then $\rho_\varepsilon * \varphi \in C_c^\infty(U)$. So $(\psi, \rho_\varepsilon * \varphi) = 0$ and thus $(\rho_\varepsilon * \psi, \varphi) = 0$ by Lemma 2.5. But $\rho_\varepsilon * \psi$ is smooth, so $(\rho_\varepsilon * \psi)|_V \in L^2(V)$. Since $C_c^\infty(V)$ is dense in $L^2(V)$, there exists $\{\varphi_j\} \subset C_c^\infty(V)$ with $\varphi_j \rightarrow (\rho_\varepsilon * \psi)|_V$ in $L^2(V)$. Hence, $\|\rho_\varepsilon * \psi\|_{L^2(V)}^2 = \lim_j (\rho_\varepsilon * \psi, \varphi_j) = 0$ and thus $(\rho_\varepsilon * \psi)|_V = 0$ a.e. But as $\varepsilon \rightarrow 0$ we have $\rho_\varepsilon * \psi \rightarrow \psi$ in $L^1(\mathbb{R}^n)$, so $\psi|_V = 0$ a.e. Since V is arbitrary, $\psi = 0$ a.e. Since U is arbitrary, $h = 0$ a.e. on Ω . \square

2.2 The Fourier transform

Notation. Let dx be the Lebesgue measure on \mathbb{R}^n . We denote

$$dm(x) := (2\pi)^{-n/2} dx.$$

This normalization is motivated by the inversion theorem we give later on. For $x, \xi \in \mathbb{R}^n$ we set $x \cdot \xi = \sum_{i=1}^n x_i \xi_i$ and $|\xi|^2 = \sum_{i=1}^n \xi_i^2$. If $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ and $\alpha \in \mathbb{N}_0^n$, we denote

$$x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \quad D^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n} \quad \text{and} \quad |\alpha| = \alpha_1 + \cdots + \alpha_n.$$

If $f \in L^1(\mathbb{R}^n)$, we define the *Fourier transform* of f by

$$(\mathcal{F}f)(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-ix \cdot \xi} \, dm(x).$$

Given $z \in \mathbb{R}^n$ and $a \in \mathbb{R}^*$, we define the function e_z and the scaling operator S_a by

$$e_z(y) := e^{iy \cdot z} \quad \text{and} \quad (S_a f)(y) = f(y/a).$$

Lemma 2.11. *Suppose $f, g \in L^1(\mathbb{R}^n)$ and $z \in \mathbb{R}^n$ and $a > 0$. Then*

$$\begin{aligned} \text{(a)} \quad \widehat{\tau_z f} &= e_{-z} \widehat{f}, & \text{(b)} \quad \widehat{e_z f} &= \tau_z \widehat{f}, \\ \text{(c)} \quad \widehat{S_a f} &= a^n S_{1/a} \widehat{f}, & \text{(d)} \quad \widehat{f * g} &= (2\pi)^{n/2} \widehat{f} \widehat{g}. \end{aligned}$$

Proof. We leave (a), (b) and (c) as exercises to the reader. For (d), note that $f * g \in L^1(\mathbb{R}^n)$ by Lemma 2.5(i). By the absolute convergence, we may use Fubini and get

$$\begin{aligned} (2\pi)^{-n/2} \widehat{(f * g)}(\xi) &= (2\pi)^{-n/2} \int \left(\int f(y)g(x-y) dy \right) e^{-ix \cdot \xi} dm(x) \\ &= \int f(y) e^{-iy \cdot \xi} \left(\int g(x-y) e^{-i(x-y) \cdot \xi} dm(x) \right) dm(y) \\ &= \left(\int f(y) e^{-iy \cdot \xi} dm(y) \right) \left(\int g(z) e^{-iz \cdot \xi} dm(z) \right) = \widehat{f}(\xi) \widehat{g}(\xi). \quad \square \end{aligned}$$

Definition 2.12. Let $C_b(\mathbb{R}^n)$ be the space of bounded continuous functions on \mathbb{R}^n . We define the space $C_0(\mathbb{R}^n) \subset C_b(\mathbb{R}^n)$ of continuous functions vanishing at infinity by

$$C_0(\mathbb{R}^n) = \{h \in C(\mathbb{R}^n) : h(\xi) \rightarrow 0 \text{ as } |\xi| \rightarrow \infty\}.$$

Theorem 2.13. (a) *If $f \in L^1(\mathbb{R}^n)$, then $\widehat{f} \in C_b(\mathbb{R}^n)$ and $\|\widehat{f}\|_\infty \leq \|f\|_1$.*

(b) *If $x^\alpha f \in L^1(\mathbb{R}^n)$ for $|\alpha| \leq k$, then $\widehat{f} \in C^k(\mathbb{R}^n)$, and $D^\alpha \widehat{f} = \mathcal{F}[(-ix)^\alpha f]$.*

(c) *If $f \in C_c^k(\mathbb{R}^n)$, then for $|\alpha| \leq k$, $(\widehat{D^\alpha f})(\xi) = (i\xi)^\alpha \widehat{f}(\xi)$.*

(d) *If $f \in C_c^\infty(\mathbb{R}^n)$, then $\widehat{f} \in C_0(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ for every $p \geq 1$.*

(e) (Riemann-Lebesgue Lemma). *If $f \in L^1(\mathbb{R}^n)$, then $\widehat{f} \in C_0(\mathbb{R}^n)$.*

Proof. If $f \in L^1(\mathbb{R}^n)$, then $|\widehat{f}(\xi)| \leq \int |f| = \|f\|_1$, so $\|\widehat{f}\|_\infty \leq \|f\|_1$. Moreover, if $\{\xi_n\} \subset \mathbb{R}^n$ with $\xi_n \rightarrow \xi$, then $e^{-ix \cdot \xi_n} \rightarrow e^{-ix \cdot \xi}$, so $\widehat{f}(\xi_n) \rightarrow \widehat{f}(\xi)$ by dominated convergence. Hence, \widehat{f} is continuous, so $\widehat{f} \in C_b(\mathbb{R}^n)$.

The case $k = 0$ of (b) is given in (a). For $k > 0$, we have by Lemma 2.1 and induction on $|\alpha| \leq k$ that $\widehat{f} \in C^k(\mathbb{R}^n)$, with

$$D^\alpha \widehat{f}(\xi) = D_\xi^\alpha \int f(x) e^{-ix \cdot \xi} dm(x) = \int f(x) (-ix)^\alpha e^{-ix \cdot \xi} dm(x),$$

so (b) holds. For (c), integrating by parts,

$$\widehat{\partial_{x_j} f}(\xi) = \int \partial_{x_j} f(x) e^{-ix \cdot \xi} dm(x) = - \int f(x) (-i\xi_j) e^{-ix \cdot \xi} dm(x) = i\xi_j \widehat{f}(\xi),$$

so (c) follows by induction. For (d), let $f \in C_c^\infty(\mathbb{R}^n)$. Then by (c), $(\widehat{D^\alpha f})(\xi) = (i\xi)^\alpha \widehat{f}(\xi)$ for all α , so using (a), $\xi^\alpha \widehat{f}(\xi)$ is bounded all α . Hence, $[\prod_1^n (1 + \xi_j^2)] \widehat{f}(\xi)$ is bounded. In particular, $\widehat{f}(\xi) \rightarrow 0$ as $|\xi| \rightarrow \infty$, i.e. $\widehat{f} \in C_0(\mathbb{R}^n)$, and $\widehat{f} \in L^1(\mathbb{R}^n)$ because $\int_{\mathbb{R}} \frac{1}{1+\xi_j^2} d\xi_j = \pi$ converges. But $\widehat{f} \in C_0(\mathbb{R}^n)$ implies $\|\widehat{f}\|_p^p \leq \sup_\xi |\widehat{f}(\xi)|^{p-1} \|\widehat{f}\|_1 < \infty$, so $\widehat{f} \in L^p(\mathbb{R}^n)$.

For (e), we know by (a) that $\widehat{f} \in C_b(\mathbb{R}^n)$. Choose $\varphi \in C_c^\infty(\mathbb{R}^n)$ such that $\|f - \varphi\|_1 < \varepsilon/2$ (use Corollary 2.9). Then $|\widehat{f}(\xi)| \leq |\widehat{(f - \varphi)}(\xi)| + |\widehat{\varphi}(\xi)| \leq \|f - \varphi\|_1 + |\widehat{\varphi}(\xi)| < \varepsilon/2 + |\widehat{\varphi}(\xi)|$. But $\widehat{\varphi}(\xi) \rightarrow 0$ as $|\xi| \rightarrow \infty$ by (d). So $|\widehat{f}(\xi)| < \varepsilon$ for $|\xi|$ sufficiently large. \square

We now calculate the Fourier transform of a Gaussian.

Lemma 2.14. *Let $\phi(x) = e^{-|x|^2/2}$ for $x \in \mathbb{R}^n$. Then $\widehat{\phi} = \phi$.*

Proof. First suppose $n = 1$. Since $\phi'(x) = -x\phi(x)$, we have by Theorem 2.13(b),(c) that $\frac{d}{d\xi}\widehat{\phi}(\xi) = \widehat{(-ix\phi)}(\xi) = \widehat{(i\phi')}\xi = i(i\xi)\widehat{\phi}(\xi) = -\xi\widehat{\phi}(\xi)$. Thus, both ϕ and $\widehat{\phi}$ solve the equation $y' = -xy$, so $\widehat{\phi} = c\phi$. But $\widehat{\phi}(0) = \int_{\mathbb{R}} e^{-x^2/2} dm(x) = 1$.³ Thus $c\phi(0) = 1$, i.e. $c = 1$, so $\widehat{\phi} = \phi$.

For general n , we have by Fubini, denoting $dm_i = (2\pi)^{-1/2} dx_i$,

$$\begin{aligned}\widehat{\phi}(\xi) &= \int_{\mathbb{R}^n} e^{-|x|^2/2} e^{-ix\cdot\xi} dm(x) = \int_{\mathbb{R}} e^{-x_1^2/2} e^{-ix_1\xi_1} dm_1 \cdots \int_{\mathbb{R}} e^{-x_n^2/2} e^{-ix_n\xi_n} dm_n \\ &= e^{-\xi_1^2/2} \cdots e^{-\xi_n^2/2} = e^{-|\xi|^2/2}.\end{aligned}\quad \square$$

Definition 2.15. Given $f \in L^1(\mathbb{R}^n)$, we define

$$(\mathcal{F}^* f)(x) = \check{f}(x) = \int_{\mathbb{R}^n} f(\xi) e^{ix\cdot\xi} dm(\xi) = \widehat{f}(-x).$$

The notation \mathcal{F}^* is motivated by the following lemma.

Lemma 2.16. *If $f, g \in L^1(\mathbb{R}^n)$, then $\int \widehat{f}g = \int f\widehat{g}$ and $(\widehat{f}, g) = (f, \check{g})$.*

Proof. By Fubini,

$$\begin{aligned}\int \widehat{f}(\xi)g(\xi) &= \int \int f(x)e^{-ix\cdot\xi}g(\xi) = \int f(x) \int g(\xi)e^{-ix\cdot\xi} = \int f(x)\widehat{g}(x), \\ (\widehat{f}, g) &= \int \widehat{f}(\xi)\overline{g(\xi)} = \int \int f(x)e^{-ix\cdot\xi}\overline{g(\xi)} = \int f(x) \int \overline{g(\xi)}e^{ix\cdot\xi} = (f, \check{g}).\end{aligned}\quad \square$$

We claim that $\mathcal{F}^*\mathcal{F}f = f$. A simple appeal to Fubini's theorem fails because the integrand in $(\mathcal{F}^*\mathcal{F}f)(x) = \int \int f(y)e^{-iy\cdot\xi}e^{ix\cdot\xi}dm(y)dm(\xi)$ is not in $L^1(\mathbb{R}^n \times \mathbb{R}^n)$. The trick is to introduce a convergence factor as follows.

Theorem 2.17 (Fourier Inversion Theorem). *If $f \in L^1(\mathbb{R}^n)$ and $\widehat{f} \in L^1(\mathbb{R}^n)$, then f agrees a.e. with a continuous function f_c , and $\mathcal{F}^*\mathcal{F}f = \mathcal{F}\mathcal{F}^*f = f_c$.*

Proof. Let $\phi(x) = e^{-|x|^2/2}$. Given $\varepsilon > 0$ and $x \in \mathbb{R}^n$, we have by Lemma 2.11(a),

$$\int e^{-\varepsilon^2|\xi|^2/2} e^{ix\cdot\xi} \widehat{f}(\xi) d\xi = \int \phi(\varepsilon\xi) e_x(\xi) \widehat{f}(\xi) d\xi = \int S_{\varepsilon^{-1}}\phi(\xi) \widehat{\tau_{-x}f}(\xi) d\xi,$$

so using Lemma 2.16 and Lemma 2.11(c),

$$\int e^{-\varepsilon^2|\xi|^2/2} e^{ix\cdot\xi} \widehat{f}(\xi) d\xi = \int \widehat{S_{\varepsilon^{-1}}\phi(\xi)} \tau_{-x}f(\xi) d\xi = \varepsilon^{-n} \int S_{\varepsilon}\widehat{\phi}(\xi) f(x + \xi) d\xi.$$

But $\widehat{\phi} = \phi$ by Lemma 2.14. So

$$\int e^{-\varepsilon^2|\xi|^2/2} e^{ix\cdot\xi} \widehat{f}(\xi) d\xi = \varepsilon^{-n} \int S_{\varepsilon}\phi(y) f(x + y) dy = \int \phi_{\varepsilon}(y) f(x - y) dy = \phi_{\varepsilon} * f(x),$$

where we changed the variable y to $-y$ and used the fact that $\phi(-y) = \phi(y)$. Now $\int \phi(x) dx = (2\pi)^{n/2}$, so by Lemma 2.8, $\phi_{\varepsilon} * f \rightarrow (2\pi)^{n/2} f$ in L^1 as $\varepsilon \rightarrow 0$. But since $\widehat{f} \in L^1$, then by dominated convergence, the LHS tends to $\int e^{ix\cdot\xi} \widehat{f}(\xi) d\xi = (2\pi)^{n/2} (\mathcal{F}^*\widehat{f})(x)$. Hence, $f = \mathcal{F}^*\mathcal{F}f$ a.e.⁴ Similarly, $\mathcal{F}\mathcal{F}^*f = f$ a.e. Since $\mathcal{F}^*\mathcal{F}f$ and $\mathcal{F}\mathcal{F}^*f$ are continuous, being the Fourier transforms of L^1 functions, the proof is complete. \square

3. Because $\left(\int_{-\infty}^{\infty} e^{-x^2/2} dx\right)^2 = \int_{\mathbb{R}^2} e^{-(x^2+y^2)/2} dx dy = \int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2/2} r dr = 2\pi$.

4. If $f_m \rightarrow f$ in L^1 and $f_m \rightarrow g$ pointwise, then $f = g$ a.e. Indeed, if $f_m \rightarrow f$ in L^1 , then some subsequence f_{m_k} converges pointwise to f (see e.g. [14]), so $f = g$ by uniqueness of the pointwise limit.

We finally give the important

Theorem 2.18 (Plancherel). *The map $\mathcal{F} : L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n) \rightarrow C_0(\mathbb{R}^n)$ extends uniquely to a unitary isomorphism $\mathcal{F} : L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$.*

In other words, the unique extension of \mathcal{F} to $L^2(\mathbb{R}^n)$ is linear, bijective, and satisfies

$$\forall f, g \in L^2(\mathbb{R}^n) : (\mathcal{F}f, \mathcal{F}g) = (f, g) \quad (\text{Parseval's identity}).$$

Proof. We first note that if $f, g \in C_c^\infty(\mathbb{R}^n)$, then $\widehat{f}, \widehat{g} \in L^1(\mathbb{R}^n)$ by Lemma 2.13(c), so by Lemma 2.16 and Theorem 2.17,

$$(\mathcal{F}f, \mathcal{F}g) = (f, \mathcal{F}^* \mathcal{F}g) = (f, g).$$

Hence, $\|\mathcal{F}f\|_2 = \|f\|_2$ for every $f \in C_c^\infty(\mathbb{R}^n)$. So $\mathcal{F}|_{C_c^\infty(\mathbb{R}^n)}$ extends uniquely to an isometry $\widetilde{\mathcal{F}} : L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$. Moreover, $\widetilde{\mathcal{F}}(f) = \widehat{f}$ on $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$. Indeed, given $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, use Corollary 2.9 (and its proof) to find $f_j \in C_c^\infty(\mathbb{R}^n)$ such that $f_j \rightarrow f$ in both $L^1(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$. Then $\|\widetilde{\mathcal{F}}f_j - \widetilde{\mathcal{F}}f\|_2 = \|f_j - f\|_2 \rightarrow 0$ and $\|\widehat{f_j} - \widehat{f}\|_\infty \leq \|f_j - f\|_1 \rightarrow 0$ by Lemma 2.13(a). But $\widetilde{\mathcal{F}}f_j = \widehat{f_j}$. Hence, $\widetilde{\mathcal{F}}f = \widehat{f}$.

It remains to show that $\widetilde{\mathcal{F}}$ is surjective. Let $h \in C_c^\infty(\mathbb{R}^n)$. Then $\mathcal{F}^*h(x) = \widehat{h}(-x) \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ by Lemma 2.13(c), so $\widetilde{\mathcal{F}}(\mathcal{F}^*h) = \widehat{\mathcal{F}^*h} = h$ by Theorem 2.17. Hence, $h \in \text{Ran } \widetilde{\mathcal{F}}$, so $C_c^\infty(\mathbb{R}^n) \subseteq \text{Ran } \widetilde{\mathcal{F}}$. Since $\widetilde{\mathcal{F}}$ is an isometry, then $\text{Ran } \widetilde{\mathcal{F}}$ is closed.⁵ Hence, by Corollary 2.9, $L^2(\mathbb{R}^n) = \overline{C_c^\infty(\mathbb{R}^n)} \subseteq \text{Ran } \widetilde{\mathcal{F}}$, so $\widetilde{\mathcal{F}}$ is surjective. \square

2.3 Sobolev spaces and embeddings

Definition 2.19. Let $f, h \in L^1_{\text{loc}}(\Omega)$. We say that h is the *weak α -th derivative* of f , and denote $h = D_w^\alpha f$, if $T_h = D^\alpha T_f$, i.e. if $(h, \varphi) = (-1)^{|\alpha|} (f, D^\alpha \varphi)$ for all $\varphi \in C_c^\infty(\Omega)$.

Remark 2.20. By Corollary 2.10, if a weak α -th derivative exists, it is unique (up to sets of measure 0). In particular, if $f \in C^\infty(\Omega)$, then $D^\alpha f = D_w^\alpha f$ for all α .

Definition 2.21. Let $1 \leq p < \infty$. We define the *Sobolev space* $W^{k,p}(\Omega)$ by

$$W^{k,p}(\Omega) = \{f \in L^p(\Omega) : D_w^\alpha f \text{ exists for all } |\alpha| \leq k \text{ and } D_w^\alpha f \in L^p(\Omega)\}$$

and endow it with the norm

$$\|f\|_{k,p} = \left(\sum_{|\alpha| \leq k} \|D_w^\alpha f\|_p^p \right)^{1/p}.$$

Sometimes we endow $W^{k,p}(\Omega)$ with the equivalent norm $N_{k,p}(f) = \sum_{|\alpha| \leq k} \|D_w^\alpha f\|_p$.

Remark 2.22. If $f, f_j \in W^{k,p}(\Omega)$, then

$$f_j \rightarrow f \text{ in } W^{k,p}(\Omega) \iff D_w^\alpha f_j \rightarrow D_w^\alpha f \text{ in } L^p(\Omega) \text{ for each } |\alpha| \leq k.$$

Indeed, if $f_j \rightarrow f$ in $W^{k,p}$, then $\|D_w^\alpha f_j - D_w^\alpha f\|_p \leq \|f_j - f\|_{k,p} \rightarrow 0$, so $D_w^\alpha f_j \rightarrow D_w^\alpha f$ in L^p . Conversely, if each $D_w^\alpha f_j \rightarrow D_w^\alpha f$ in L^p , then $\|f_j - f\|_{k,p} = \left(\sum_{|\alpha| \leq k} \|D_w^\alpha f_j - D_w^\alpha f\|_p^p \right)^{1/p} \rightarrow 0$. Similarly, (f_j) is Cauchy in $W^{k,p}$ iff each $(D_w^\alpha f_j)$ is Cauchy in L^p .

⁵ If $\widetilde{\mathcal{F}}f_j \rightarrow g$, then $\|f_j - f_k\|_2 = \|\widetilde{\mathcal{F}}f_j - \widetilde{\mathcal{F}}f_k\|_2 \rightarrow 0$ as $j, k \rightarrow \infty$, so by completeness of L^2 , $f_j \rightarrow h$ for some $h \in L^2$, so $\widetilde{\mathcal{F}}f_j \rightarrow \widetilde{\mathcal{F}}h$, so $g = \widetilde{\mathcal{F}}h$ and the range is closed.

Lemma 2.23. $W^{k,p}(\Omega)$ is a Banach space. $W^{k,2}(\Omega)$ is a Hilbert space.

Proof. Let (f_j) be Cauchy in $W^{k,p}(\Omega)$. Then $(D_w^\alpha f_j)$ is Cauchy in $L^p(\Omega)$ for every $|\alpha| \leq k$, so by completeness of $L^p(\Omega)$, there exists $f_\alpha \in L^p(\Omega)$ with $D_w^\alpha f_j \rightarrow f_\alpha$. Let $f := f_0$ and $\varphi \in C_c^\infty(\Omega)$. Then $(f, D^\alpha \varphi) = \lim_j (f_j, D^\alpha \varphi) = \lim_j (-1)^{|\alpha|} (D_w^\alpha f_j, \varphi) = (-1)^{|\alpha|} (f_\alpha, \varphi)$. Hence, $D_w^\alpha f = f_\alpha$. We thus proved that $D_w^\alpha f_j \rightarrow D_w^\alpha f$ in $L^p(\Omega)$ for each $|\alpha| \leq k$, so $f_j \rightarrow f$ in $W^{k,p}(\Omega)$.

When $p = 2$, $(f, g)_k := \sum_{|\alpha| \leq k} (D_w^\alpha f, D_w^\alpha g)$ defines an inner product on $W^{k,2}(\Omega)$ which induces the norm $\|\cdot\|_{k,2}$, so $W^{k,2}(\Omega)$ is a Hilbert space. \square

Lemma 2.24. (a) If $f \in W^{k,p}(\Omega)$ and $|\alpha| + |\beta| \leq k$, then $D_w^\alpha (D_w^\beta f) = D_w^{\alpha+\beta} f$.

(b) If $f \in W^{k,p}(\Omega)$ and $\varphi \in C_c^\infty(\Omega)$, then $f\varphi \in W^{k,p}(\Omega)$ and $\partial_j^w (f\varphi) = (\partial_j^w f)\varphi + f\partial_j \varphi$, where ∂_j^w is the weak j -th partial derivative.

(c) If $\rho \in C_c^\infty(\mathbb{R}^n)$ and $f \in W^{k,p}(\mathbb{R}^n)$, then $\rho * f \in C^\infty(\mathbb{R}^n) \cap W^{k,p}(\mathbb{R}^n)$ and $D^\alpha (\rho * f) = \rho * D_w^\alpha f$ for all $|\alpha| \leq k$.

Proof. (a) Exercise.

(b) Let $\psi \in C_c^\infty(\Omega)$. Then

$$\begin{aligned} (f\varphi, \partial_j \psi) &= (f, \overline{\varphi} \partial_j \psi) = (f, \partial_j (\overline{\varphi} \psi)) - (f, \psi \partial_j \overline{\varphi}) \\ &= -(\partial_j^w f, \overline{\varphi} \psi) - (f \partial_j \varphi, \psi) = -((\partial_j^w f)\varphi + f\partial_j \varphi, \psi). \end{aligned}$$

Hence, $\partial_j^w (f\varphi) = (\partial_j^w f)\varphi + f\partial_j \varphi \in L^p(\Omega)$, and (b) follows by induction.

(c) We have $\rho * f = f * \rho \in C^\infty(\mathbb{R}^n)$ by Lemma 2.5(ii), and $\rho * D_w^\alpha f \in L^p(\mathbb{R}^n)$ by Lemma 2.5(i). So it suffices to see that $D^\alpha (\rho * f) = \rho * D_w^\alpha f$. Let $\varphi \in C_c^\infty(\mathbb{R}^n)$. Then

$$\begin{aligned} (\rho * f, D^\alpha \varphi) &= (f, \rho^* * D^\alpha \varphi) && \text{(Lemma 2.5(iii))} \\ &= (f, D^\alpha (\rho^* * \varphi)) && \text{(Lemma 2.5(ii))} \\ &= (-1)^{|\alpha|} (D_w^\alpha f, \rho^* * \varphi) && \text{(since } \rho^* * \varphi \in C_c^\infty(\mathbb{R}^n)) \\ &= (-1)^{|\alpha|} (\rho * D_w^\alpha f, \varphi). && \square \end{aligned}$$

Definition 2.25. If $\Omega \subseteq \mathbb{R}^n$ is open, we define $W_0^{k,p}(\Omega)$ as the closure of $C_c^\infty(\Omega)$ in $W^{k,p}(\Omega)$. Thus, $f \in W_0^{k,p}(\Omega)$ iff there exists $\{\varphi_j\} \subset C_c^\infty(\Omega)$ such that $\|f - \varphi_j\|_{k,p} \rightarrow 0$.

Roughly speaking, $W_0^{k,p}(\Omega)$ consists of functions in $W^{k,p}(\Omega)$ which vanish on $\partial\Omega$. This can be made precise using *traces*, but we will not discuss this here. Note that $W_0^{0,p}(\Omega) = L^p(\Omega)$ by Corollary 2.9.

Lemma 2.26. $W_0^{k,p}(\mathbb{R}^n) = W^{k,p}(\mathbb{R}^n)$. In other words, $C_c^\infty(\mathbb{R}^n)$ is dense in $W^{k,p}(\mathbb{R}^n)$.

Since $C_c^\infty(\mathbb{R}^n) \subset C^\infty(\mathbb{R}^n) \cap W^{k,p}(\mathbb{R}^n)$, then $C^\infty(\mathbb{R}^n) \cap W^{k,p}(\mathbb{R}^n)$ is in particular dense in $W^{k,p}(\mathbb{R}^n)$. Actually $C^\infty(\Omega) \cap W^{k,p}(\Omega)$ is dense in $W^{k,p}(\Omega)$ for any open set $\Omega \subseteq \mathbb{R}^n$. This is known as the Meyers-Serrin theorem. For a proof, see e.g. [23, Theorem 10.15].

Proof. Let $\psi \in C_c^\infty(\mathbb{R}^n)$ such that $\psi(x) = 1$ if $|x| \leq 1$ and $\psi(x) = 0$ if $|x| \geq 2$. Given $f \in W^{k,p}(\mathbb{R}^n)$, let $f_R(x) = f(x)\psi(\frac{x}{R})$. Then $f_R(x) = f(x)$ if $|x| \leq R$, $f_R(x) = 0$ if $|x| \geq 2R$, and $f_R \in W^{k,p}(\mathbb{R}^n)$ by Lemma 2.24(b). Assuming $|D^\gamma \psi(x)| \leq M$ for $|\gamma| \leq k$, we get $|D^\gamma \psi(\frac{x}{R})| \leq \frac{M}{R^{|\gamma|}} \leq M$ for $R \geq 1$, hence $|D_w^\alpha f_R(x)| \leq M' \sum_{\beta \leq \alpha} |D_w^\beta f(x)|$ by Leibniz. Thus, $\|D_w^\alpha f_R - D_w^\alpha f\|_p = (\int_{|x| > R} |D_w^\alpha f_R - D_w^\alpha f|^p)^{1/p} \leq M'' \sum_{\beta \leq \alpha} (\int_{|x| > R} |D_w^\beta f|^p)^{1/p} \rightarrow 0$ as $R \rightarrow \infty$, because $f \in W^{k,p}(\mathbb{R}^n)$. Let ρ_δ be a mollifier. Then $\rho_\delta * f_R \in C_c^\infty(\mathbb{R}^n)$

by Lemma 2.24(c) and Remark 2.6. Moreover, $D^\alpha(\rho_\delta * f_R) = \rho_\delta * (D_w^\alpha f_R) \rightarrow D_w^\alpha f_R$ in $L^p(\mathbb{R}^n)$ for each $|\alpha| \leq k$ as $\delta \rightarrow 0$. Thus, given $\varepsilon > 0$, we may choose δ small enough so that $\|\rho_\delta * f_R - f_R\|_{k,p} < \varepsilon/2$ and R so large that $\|f_R - f\|_{k,p} < \varepsilon/2$, which yields $\|\rho_\delta * f_R - f\|_{k,p} < \varepsilon$. Hence, $C_c^\infty(\mathbb{R}^n)$ is dense in $W^{k,p}(\mathbb{R}^n)$. \square

Corollary 2.27. *If $f \in W^{k,2}(\mathbb{R}^n)$, then for $|\alpha| \leq k$ we have*

$$(\widehat{D_w^\alpha f})(\xi) = (i\xi)^\alpha \widehat{f}(\xi).$$

In particular, $\xi^\alpha \widehat{f} \in L^2(\mathbb{R}^n)$ for all $|\alpha| \leq k$.

Proof. Use Lemma 2.26 to find $\{\varphi_j\} \subset C_c^\infty(\mathbb{R}^n)$ with $\varphi_j \rightarrow f$ in $W^{k,2}(\mathbb{R}^n)$. Then $D^\alpha \varphi_j \rightarrow D_w^\alpha f$ in $L^2(\mathbb{R}^n)$. Since \mathcal{F} is unitary, hence continuous on L^2 , we get by Theorem 2.13(c), $\widehat{D_w^\alpha f} = \lim \widehat{D^\alpha \varphi_j} = \lim (i\xi)^\alpha \widehat{\varphi_j} = (i\xi)^\alpha \widehat{f}$. \square

We are finally ready to illustrate the power of the Fourier transform in understanding the structure of Sobolev spaces.

Definition 2.28. We define

$$C_0^k(\mathbb{R}^n) = \{f \in C^k(\mathbb{R}^n) : D^\alpha f \in C_0(\mathbb{R}^n) \text{ for } |\alpha| \leq k\}.$$

Then $C_0^k(\mathbb{R}^n)$ is a Banach space when endowed with $\|f\| = \sum_{|\alpha| \leq k} \|D^\alpha f\|_{C^0}$.

Theorem 2.29 (Sobolev embedding theorem). *Suppose $k > r + \frac{n}{2}$.*

- (1) *If $f \in W^{k,2}(\mathbb{R}^n)$, then for $|\alpha| \leq r$, we have $\widehat{D_w^\alpha f} \in L^1(\mathbb{R}^n)$ with $\|\widehat{D_w^\alpha f}\|_1 \leq C \|f\|_{k,2}$ for some C independent of f .*
- (2) *Every $f \in W^{k,2}(\mathbb{R}^n)$ agrees a.e. with an $f_c \in C_0^r(\mathbb{R}^n)$, and the embedding $W^{k,2}(\mathbb{R}^n) \hookrightarrow C_0^r(\mathbb{R}^n)$ which maps $f \mapsto f_c$ is continuous.*

We usually express (2) less precisely by saying that $W^{k,2}(\mathbb{R}^n) \subset C_0^r(\mathbb{R}^n)$ and that the inclusion map is bounded.

Proof. By Corollary 2.27, $\|\widehat{D_w^\alpha f}\|_1 = \|\xi^\alpha \widehat{f}\|_1$. Let $h_1 = (1 + |\xi|^k) \widehat{f}$ and $h_2 = (1 + |\xi|^k)^{-1} \xi^\alpha$. Then $\xi^\alpha \widehat{f} = h_1 h_2$. We first show that $h_1, h_2 \in L^2(\mathbb{R}^n)$, then apply Cauchy-Schwarz.

For h_1 , we have $\|h_1\|_2 \leq \|f\|_2 + \|\xi|^k \widehat{f}\|_2$. For the first term, we have by Plancherel, $\|\widehat{f}\|_2 = \|f\|_2 \leq \|f\|_{k,2}$. Next, note that $|\xi|^{2k} = (\sum_{i=1}^n \xi_i^2)^k = \sum_{|\beta| \leq k} c_\beta \xi^{2\beta}$ for some $c_\beta \geq 0$. So using Corollary 2.27 and Plancherel again,

$$\begin{aligned} \|\xi|^k \widehat{f}\|_2^2 &= \int |\xi|^{2k} |\widehat{f}(\xi)|^2 d\xi = \sum c_\beta \int |\xi^\beta \widehat{f}(\xi)|^2 d\xi = \sum c_\beta \|\xi^\beta \widehat{f}\|_2^2 \\ &= \sum c_\beta \|\widehat{D_w^\beta f}\|_2^2 = \sum c_\beta \|D_w^\beta f\|_2^2 \leq c \|f\|_{k,2}^2. \end{aligned}$$

Hence,

$$\|h_1\|_2 \leq \|\widehat{f}\|_2 + \|\xi|^k \widehat{f}\|_2 \leq c' \|f\|_{k,2}.$$

For h_2 , we have $\|h_2\|_2^2 = \int_{|\xi| \leq 1} |h_2(\xi)|^2 d\xi + \int_{|\xi| \geq 1} |h_2(\xi)|^2 d\xi$. Since h_2 is continuous, the first integral is finite. For the second integral, since $|\xi_j| \leq |\xi|$, we have $|\xi^\alpha| \leq |\xi|^{|\alpha|}$. But for $|\xi| \geq 1$ we have $|\xi|^{|\alpha|} \leq |\xi|^r$ if $|\alpha| \leq r$. Thus, $\int_{|\xi| \geq 1} |h_2(\xi)|^2 d\xi \leq \int_{|\xi| \geq 1} (1 + |\xi|^k)^{-2} |\xi|^{2r} d\xi = \int_{R \geq 1} (1 + R^k)^{-2} R^{2r} R^{n-1} dR \leq \int_{R \geq 1} R^{-2k} R^{2r+n-1} dR < \infty$ because $2r + n - 1 - 2k < -1$, since $r + \frac{n}{2} < k$.

We thus showed that $\|\widehat{D_w^\alpha f}\|_1 = \|\xi^\alpha \widehat{f}\|_1 = \|h_1 h_2\|_1 \leq \|h_1\|_2 \|h_2\|_2 \leq C \|f\|_{k,2}$.

For (2), we know from (1) and Corollary 2.27 that for each $|\alpha| \leq r$, we have $(i\xi)^\alpha \widehat{f} = \widehat{D_w^\alpha f} \in L^1(\mathbb{R}^n)$, so by Theorem 2.13(b),(e), we have $\mathcal{F}(\widehat{f}) \in C_0^r(\mathbb{R}^n)$. Hence, $\mathcal{F}^* \mathcal{F} f(\xi) = \mathcal{F}(\widehat{f})(-\xi) \in C_0^r(\mathbb{R}^n)$. So by the inversion formula, f agrees a.e. with an $f_c \in C_0^r(\mathbb{R}^n)$. Moreover, by Theorem 2.13(b), $\|D^\alpha f_c\|_\infty = \|D^\alpha \mathcal{F}(\widehat{f})\|_\infty = \|\mathcal{F}[(-i\xi)^\alpha \widehat{f}]\|_\infty$, and by Theorem 2.13(a), $\|\mathcal{F}[(-i\xi)^\alpha \widehat{f}]\|_\infty \leq \|(-i\xi)^\alpha \widehat{f}\|_1$. But by Corollary 2.27 and (1), $\|(-i\xi)^\alpha \widehat{f}\|_1 = \|\widehat{D_w^\alpha f}\|_1 \leq C \|f\|_{k,2}$. Hence, $\|f_c\|_{C_0^r} = \sum_{|\alpha| \leq r} \|D^\alpha f_c\|_\infty \leq C' \|f\|_{k,2}$, so the embedding is bounded, hence continuous. \square

Let us partly extend the previous theorem to open subsets.

Lemma 2.30. *Let $f \in W_0^{k,p}(\Omega)$ and let \widetilde{f} be the extension of f by zero to \mathbb{R}^n , i.e. $\widetilde{f}(x) = f(x)$ if $x \in \Omega$ and $\widetilde{f}(x) = 0$ if $x \notin \Omega$. Then $\widetilde{f} \in W^{k,p}(\mathbb{R}^n)$, $D_w^\alpha \widetilde{f} = \widehat{D_w^\alpha f}$, and the map $f \mapsto \widetilde{f}$ is isometric.*

Proof. Let $\{\phi_j\} \subset C_c^\infty(\Omega)$ converge to f in $W_0^{k,p}(\Omega)$. Then $D^\alpha \phi_j \rightarrow D_w^\alpha f$ in $L^p(\Omega)$ for all $|\alpha| \leq k$. Denote $(f, g)_E = \int_E f \overline{g}$. Then given $\psi \in C_c^\infty(\mathbb{R}^n)$, we have

$$\begin{aligned} (\widetilde{f}, D^\alpha \psi)_{\mathbb{R}^n} &= (f, D^\alpha \psi)_\Omega = \lim_{j \rightarrow \infty} (\phi_j, D^\alpha \psi)_\Omega = \lim_{j \rightarrow \infty} (-1)^{|\alpha|} (D^\alpha \phi_j, \psi)_\Omega \\ &= (-1)^{|\alpha|} (D_w^\alpha f, \psi)_\Omega = (-1)^{|\alpha|} (\widehat{D_w^\alpha f}, \psi)_{\mathbb{R}^n}. \end{aligned}$$

Thus, $D_w^\alpha \widetilde{f} = \widehat{D_w^\alpha f} \in L^p(\mathbb{R}^n)$. So $\widetilde{f} \in W^{k,p}(\mathbb{R}^n)$ and $\|\widetilde{f}\|_{W^{k,p}(\mathbb{R}^n)} = \|f\|_{W_0^{k,p}(\Omega)}$. \square

Corollary 2.31. *Let $\Omega \subseteq \mathbb{R}^n$ be open and $k > r + \frac{n}{2}$. Then*

- (a) $W^{k,2}(\Omega) \hookrightarrow C^r(\Omega)$.
- (b) $W_0^{k,2}(\Omega) \hookrightarrow C_b^r(\Omega)$ and the embedding is continuous.

The continuity of the embedding $W^{k,2}(\Omega) \hookrightarrow C^r(\Omega)$ is delicate for general Ω , and depends on the nature of $\partial\Omega$.

Proof. For (a), let $U \subset \Omega$, with $\overline{U} \subset \Omega$ and \overline{U} compact. Let $\phi \in C_c^\infty(\Omega)$ with $\phi = 1$ on U . Then for $f \in W^{k,2}(\Omega)$, we have $\phi f \in W^{k,2}(\mathbb{R}^n)$ by Lemma 2.24, so we can apply Theorem 2.29 to deduce that ϕf agrees a.e. with some $g_U \in C^r(\Omega)$. Hence, $f|_U$ agrees a.e. with g_U , and since U is arbitrary, f can be identified with some $f_c \in C^r(\Omega)$.

For (b), the maps $f \mapsto \widetilde{f} \mapsto \widetilde{f}_c \mapsto \widetilde{f}_c|_\Omega$ from $W_0^{k,2}(\Omega) \rightarrow W^{k,2}(\mathbb{R}^n) \hookrightarrow C_0^r(\mathbb{R}^n) \rightarrow C_b^r(\Omega)$ are all bounded by Theorem 2.29 and Lemma 2.30, so the embedding is continuous. \square

We now prove the Rellich theorem, which is of great importance in applications.

Definition 2.32. Let X, Y be normed spaces. We say that $T : X \rightarrow Y$ is *compact* if $\overline{T(B)}$ is compact in Y for every bounded set $B \subseteq X$.

Equivalently, T is compact if any bounded sequence (x_n) in X has a subsequence (x_{n_k}) such that (Tx_{n_k}) converges (Exercise).

For example, if T has a finite rank, i.e. $\dim \text{Ran}(T) < \infty$, then T is compact.

Theorem 2.33 (Rellich embedding theorem). *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set.*

- (a) *For any $k \geq 1$, the inclusion $W_0^{k,2}(\Omega) \hookrightarrow W_0^{k-1,2}(\Omega)$ is a compact operator.*
- (b) *If $k > r + \frac{n}{2} + 1$, then the embedding $W_0^{k,2}(\Omega) \hookrightarrow C_b^r(\Omega)$ is a compact operator.*

In particular, the identity map $W_0^{1,2}(\Omega) \rightarrow L^2(\Omega)$ is compact (recall that $W_0^{0,2}(\Omega) = L^2(\Omega)$). It is also true that the identity map $W^{1,2}(\Omega) \rightarrow L^2(\Omega)$ is compact, but $\partial\Omega$ should be continuous in this case (see e.g. [27]). In fact the following proof works without change for $W^{1,2}(\Omega)$ if Ω is an *extension domain*, i.e. there exists a continuous linear operator $E : W^{1,2}(\Omega) \rightarrow W^{1,2}(\mathbb{R}^n)$. Domains with regular boundaries (e.g. Lipschitz continuous) are extension domains; see e.g. [23].

Proof. (b) follows from (a) and Corollary 2.31, so we only need to prove (a).

First assume $k = 1$ and let (f_j) be bounded in $W_0^{1,2}(\Omega)$. By Lemma 2.30, identifying f_j with \widehat{f}_j , we may regard (f_j) as a bounded sequence in $W^{1,2}(\mathbb{R}^n)$. So applying Plancherel,

$$\|f_p - f_q\|_2^2 = \|\widehat{f}_p - \widehat{f}_q\|_2^2 = \int_{|\xi| \leq R} |\widehat{f}_p - \widehat{f}_q|^2 + \int_{|\xi| \geq R} |\widehat{f}_p - \widehat{f}_q|^2$$

for any $R > 0$. Since for each i , $(\partial_i^w f_j)$ is bounded in $L^2(\mathbb{R}^n)$, the sequence $(\xi_i \widehat{f}_j)$ is bounded in $L^2(\mathbb{R}^n)$ by Plancherel and Corollary 2.27. Hence, $(|\xi| \widehat{f}_j)$ is bounded, say $\| |\xi| \widehat{f}_j \|_2 \leq C$ for all j . Hence,

$$\int_{|\xi| \geq R} |\widehat{f}_p - \widehat{f}_q|^2 d\xi \leq \int_{|\xi| \geq R} \frac{|\xi|^2}{R^2} |\widehat{f}_p - \widehat{f}_q|^2 d\xi \leq \frac{4C^2}{R^2}.$$

Given $\varepsilon > 0$, we may thus choose R so that $\int_{|\xi| \geq R} |\widehat{f}_p - \widehat{f}_q|^2 < \varepsilon/2$ for all p, q . We now turn to the integral over $\{|\xi| \leq R\}$. Since (f_j) is bounded in $L^2(\Omega)$, by passing to a subsequence, we may assume (f_j) is weakly convergent.⁶ Since Ω is bounded, we have $e_\xi \in L^2(\Omega)$, so $\widehat{f}_j(\xi) = (2\pi)^{-n/2} (f_j, e_\xi)$ is convergent and in particular Cauchy. That is, $\widehat{f}_p(\xi) - \widehat{f}_q(\xi) \rightarrow 0$ pointwise as $p, q \rightarrow \infty$. But since Ω is bounded and (f_j) is bounded in $L^2(\Omega)$, then (f_j) is bounded in $L^1(\Omega)$, because $\|f_j\|_1 = \|f_j \cdot 1\|_1 \leq \|f_j\|_2 \cdot \text{vol}(\Omega)^{1/2}$ by Cauchy-Schwarz. So by Theorem 2.13(a), (\widehat{f}_j) is uniformly bounded in $C_b(\mathbb{R}^n)$, say $\|\widehat{f}_j\|_\infty \leq M$. Since M is integrable on $\{|\xi| \leq R\}$, then by dominated convergence, we get $\int_{|\xi| \leq R} |\widehat{f}_p(\xi) - \widehat{f}_q(\xi)|^2 \rightarrow 0$ as $p, q \rightarrow \infty$. We thus showed that for p, q large enough, $\|f_p - f_q\|_2^2 < \varepsilon$, so (f_j) is Cauchy in $L^2(\Omega)$, hence convergent. This proves (a) for $k = 1$.

Now let (f_j) be bounded in $W_0^{k,2}(\Omega)$, $k > 1$. Then (f_j) is bounded in $W_0^{1,2}(\Omega)$, so by the case $k = 1$, (f_j) has a convergent subsequence in $L^2(\Omega)$, say $(f_{\varphi_0(j)})$, where $\varphi_0 : \mathbb{N} \rightarrow \mathbb{N}$ is strictly increasing. Next, $(\partial_{x_1} f_{\varphi_0(j)})$ is bounded in $W_0^{1,2}(\Omega)$, so it has a convergent subsequence in $L^2(\Omega)$, say $(\partial_{x_1} f_{\varphi_0 \circ \varphi_1(j)})$. Note that $(f_{\varphi_0 \circ \varphi_1(j)})$ also converges, since it is a subsequence of $(f_{\varphi_0(j)})$. By induction, we find a subsequence $(f_{\varphi_0 \circ \dots \circ \varphi_n(j)})$ such that $(D^\alpha f_{\varphi_0 \circ \dots \circ \varphi_n(j)})$ converges for all $|\alpha| \leq 1$. Take $\psi_1 = \varphi_0 \circ \dots \circ \varphi_n$. By induction, we find a subsequence $(f_{\psi_1 \circ \dots \circ \psi_{k-1}(j)})$ such that $(D^\alpha f_{\psi_1 \circ \dots \circ \psi_{k-1}(j)})$ converges for all $|\alpha| \leq k-1$. Thus, taking $\chi = \psi_1 \circ \dots \circ \psi_{k-1}$, it follows that the subsequence $(f_{\chi(j)})$ converges in $W_0^{k-1,2}(\Omega)$. \square

Here is an important application of the Rellich embedding. One can prove that the Laplace operator $-\Delta : C_c^\infty(\Omega) \rightarrow L^2(\Omega)$ has a self-adjoint extension in $L^2(\Omega)$, the *Dirichlet Laplacian*. One can prove that its resolvent $R_\lambda = (-\Delta - \lambda)^{-1}$ is a bounded map from $L^2(\Omega) \rightarrow W_0^{1,2}(\Omega)$. Since the embedding $W_0^{1,2}(\Omega) \hookrightarrow L^2(\Omega)$ is compact, then the resolvent $R_\lambda : L^2(\Omega) \rightarrow L^2(\Omega)$ is a compact operator. It follows that $-\Delta_\Omega$ has a discrete spectrum, with eigenvalues tending to ∞ . This generalizes the fact that the Dirichlet eigenvalues of $-\Delta$ on $]-a, a[$ are $E_n = (\frac{n\pi}{2a})^2$, and is very convenient because we can add a bounded potential V without any effort. Hopefully, we'll prove the details later in Chapter 4.

6. Any bounded sequence in a Hilbert space has a weakly convergent subsequence. This can be proved directly (see e.g. [30, Chapter 16]), or using the Banach-Alaoglu theorem we shall prove next chapter.

2.4 Paley-Wiener Theorem

The Paley-Wiener theorems relate the decay properties of functions with the analyticity of their Fourier transform. There are several versions of this theorem, some apply to square integrable functions [31], others to distributions [32]. In this section we present one for smooth functions, following [40]. We denote $\overline{B}_r = \{x \in \mathbb{R}^n : |x| \leq r\}$.

Theorem 2.34 (Paley-Wiener). *A function f on \mathbb{R}^n is the Fourier transform of a function in $C_c^\infty(\mathbb{R}^n)$ with support in \overline{B}_r iff f can be extended to \mathbb{C}^n as an entire function, and there are constants C_N such that*

$$|f(\zeta)| \leq \frac{C_N e^{r|\operatorname{Im} \zeta|}}{(1 + |\zeta|)^N} \quad \text{for all } \zeta \in \mathbb{C}^n \text{ and } N = 0, 1, 2, \dots \quad (\star)$$

Proof. Suppose $f(\lambda) = \widehat{\varphi}(\lambda)$ for some $\varphi \in C_c^\infty(\mathbb{R}^n)$ with $\operatorname{supp} \varphi \subseteq \overline{B}_r$. Define $f(\zeta) = \int_{\mathbb{R}^n} \varphi(x) e^{-ix \cdot \zeta} dm(x)$ for $\zeta \in \mathbb{C}^n$. This is well-defined because $|e^{-ix \cdot \zeta}| \leq e^{r|\operatorname{Im} \zeta|}$ for $x \in \overline{B}_r$, so $e_{-\zeta} \varphi \in L^1(\mathbb{R}^n)$ for each $\zeta \in \mathbb{C}^n$. Moreover, the power series of $e^{-ix \cdot \zeta}$ converges uniformly on \overline{B}_r , so $f(\zeta) = \int_{\overline{B}_r} \varphi(x) \sum_{\alpha \in \mathbb{N}_0^n} \frac{(-ix)^\alpha \zeta^\alpha}{\alpha!} dm(x) = \sum_{\alpha \in \mathbb{N}_0^n} \left(\int_{\overline{B}_r} \varphi(x) (-ix)^\alpha dm(x) \frac{\zeta^\alpha}{\alpha!} \right)$. Since $|\int_{\overline{B}_r} \varphi(x) (-ix)^\alpha dm(x)| \leq r^{|\alpha|} \|\varphi\|_1$, the series of $f(\zeta)$ converges absolutely for all $\zeta \in \mathbb{C}^n$. Hence, f is entire. To see (\star) , integrate by parts to get $(i\zeta)^\alpha f(\zeta) = \int \varphi(x) (i\zeta)^\alpha e^{-i\zeta \cdot x} dm(x) = \int (D^\alpha \varphi)(x) e^{-i\zeta \cdot x} dm(x)$ and thus $|\zeta^\alpha| |f(\zeta)| \leq \|D^\alpha \varphi\|_1 e^{r|\operatorname{Im} \zeta|}$. Hence, $(1 + |\zeta|)^N |f(\zeta)| \leq (1 + |\zeta_1| + \dots + |\zeta_n|)^N |f(\zeta)| \leq C_N e^{r|\operatorname{Im} \zeta|}$ for some C_N .

Conversely, suppose f is entire and satisfies (\star) . Then $\lambda \mapsto (1 + |\lambda|)^N f(\lambda)$ is bounded on \mathbb{R}^n for every N , so we may define

$$\varphi(x) = \int_{\mathbb{R}^n} f(\lambda) e^{ix \cdot \lambda} dm(\lambda) = \widetilde{f}(x).$$

Then $\varphi \in C^\infty(\mathbb{R}^n)$ by Theorem 2.13(b) (or Lemma 2.1(b)). We claim that $\operatorname{supp} \varphi \subseteq \overline{B}_r$. To prove this, first note that for any $\eta \in \mathbb{R}^n$, we have

$$\varphi(x) = \int_{\mathbb{R}^n} f(\lambda + i\eta) e^{ix \cdot (\lambda + i\eta)} dm(\lambda). \quad (\dagger)$$

Indeed, fix η_k for $k > 1$ and consider

$$I(\eta_1) = \int_{\mathbb{R}} f(\lambda_1 + i\eta_1, \dots, \lambda_n + i\eta_n) e^{ix \cdot [(\lambda_1 + i\eta_1) + \dots + (\lambda_n + i\eta_n)]} d\lambda_1.$$

Let Γ be a rectangular contour in \mathbb{C} with vertices at $\pm R$ and $\pm R + i\eta_1$, where $R > 0$, and denote $\zeta = \lambda + i\eta$. Then by Cauchy's theorem, $\int_{\Gamma} f(\zeta) e^{ix \cdot \zeta} d\zeta_1 = 0$, since the integrand is analytic in ζ_1 . For points $\zeta = (\pm R + is, \zeta_2, \dots, \zeta_n)$ along the vertical sides of Γ , where $|s| \leq |\eta_1|$, we have by (\star)

$$|f(\zeta) e^{ix \cdot \zeta}| \leq \frac{C_N e^{r|\operatorname{Im} \zeta|} e^{-x \cdot \operatorname{Im} \zeta}}{(1 + |\zeta|)^N} \rightarrow 0$$

as $R \rightarrow \infty$. Hence, the part of the contour integral along the vertical sides tends to 0 as $R \rightarrow \infty$, so $0 = \lim_{R \rightarrow \infty} \int_{\Gamma} f(\zeta) e^{ix \cdot \zeta} d\zeta_1 = I(0) - I(\eta_1)$. Hence, $I(\eta_1) = I(0)$. Repeating this for each η_j and using Fubini, we get (\dagger) . Hence, using (\star) ,

$$|\varphi(x)| \leq \int \frac{C_N e^{r|\eta|}}{(1 + |\lambda + i\eta|)^N} e^{-x \cdot \eta} dm(\lambda) \leq e^{r|\eta| - x \cdot \eta} \int \frac{C_N}{(1 + |\lambda|)^N} dm(\lambda),$$

where N is chosen large enough so that the integral in the RHS converges. Set $\eta = \lambda x$ with $\lambda > 0$. Then $e^{r|\eta| - x \cdot \eta} = e^{-\lambda|x|(|x| - r)}$, so taking $\lambda \rightarrow \infty$, we conclude that $|\varphi(x)| = 0$ if $|x| > r$. Hence, $\operatorname{supp} \varphi \subseteq \overline{B}_r$.

It follows from the inversion formula that $f(\lambda) = \widehat{\varphi}(\lambda)$ for $\lambda \in \mathbb{R}^n$. \square

2.5 Further results

Many parts of this chapter can be generalized quite easily to distributions. For example, if $\varphi \in \mathcal{D}(\mathbb{R}^n)$ and $T \in \mathcal{D}'(\mathbb{R}^n)$, we define the convolution $(T * \varphi)(x) := T(\tau_x \tilde{\varphi})$, where $\tilde{\varphi}(x) = \varphi(-x)$. Note that for distributions T_f we have $T_f(\tau_x \tilde{\varphi}) = \int f(y) \tau_x \tilde{\varphi}(y) dy = \int f(y) \varphi(x-y) dy$, so this generalizes the convolution of functions. One can prove expected results; for example $T * \varphi \in C^\infty(\mathbb{R}^n)$, with $D^\alpha(T * \varphi) = T * (D^\alpha \varphi)$. Convolutions can similarly be defined for $T \in \mathcal{E}'(\mathbb{R}^n)$ and $\varphi \in \mathcal{E}(\mathbb{R}^n)$. The Fourier transform can also be extended to tempered distributions. The reader will find out how in Exercise 12. For details on all these generalizations, see e.g. [32].

Sobolev spaces are widely used in spectral theory, because they arise in the domains of self-adjointness of differential operators. There are many embedding theorems for these spaces, and many important inequalities which are used on a regular basis. For example, the Sobolev embedding can be generalized to $W^{k,p}(\Omega) \hookrightarrow C_b^r(\Omega)$, when Ω is an open set with a sufficiently regular boundary (for example Lipschitz continuous). The condition becomes $k > r + \frac{n}{p}$. Similarly, the Rellich embedding generalizes to the Rellich-Kondrachov theorem. A classic reference here is [1], but the basic results can be found in many books, e.g. [27], [4], [12] or [24].

One generally learns as an undergraduate that the Laplace transform can be helpful in solving ordinary differential equations. For example, to solve $f''(t) + 4f(t) = \sin(2t)$ with initial conditions $f(0) = f'(0) = 0$, take the Laplace transform of both sides. This gives $s^2 \mathcal{L}\{f(t)\} - sf(0) - f'(0) + 4\mathcal{L}\{f(t)\} = \mathcal{L}\{\sin(2t)\}$. Since $\mathcal{L}\{\sin(2t)\} = \frac{2}{s^2+4}$, we deduce that $\mathcal{L}\{f(t)\} = \frac{2}{(s^2+4)^2}$, so taking the inverse Laplace transform we get $f(t) = \frac{1}{8} \sin(2t) - \frac{t}{4} \cos(2t)$. The bottom line is that the differential equation was transformed into an algebraic one, which was easy to solve, and we deduced the differential solution from the algebraic one using the inverse Laplace. This idea is also exploited in partial differential equations, using the Fourier transform. For example, if we have the PDE $P(-i\partial_x, -i\partial_t)u(x, t) = 0$ and we take the Fourier transform, we get $P(\xi, -i\partial_t)\hat{u}(\xi, t) = 0$. This is a PDE in t only; which is a lot easier to solve. We then get the solution of the original PDE using the inverse Fourier transform. For details, see e.g. [11].

There are many directions to pursue from where we stopped. One can for example continue with harmonic analysis; see [8] for an elementary treatment. Another route is the theory of pseudo-differential operators and microlocal/semiclassical analysis. Let us mention an example of such operators. As we know, if $f \in C_c^\infty(\mathbb{R}^n)$, then $(\widehat{D^\alpha f})(\xi) = (i\xi)^\alpha \hat{f}(\xi)$. So denoting $D_\alpha = (i)^{-|\alpha|} D^\alpha$, we get $(\widehat{D_\alpha f})(\xi) = \xi^\alpha \hat{f}(\xi)$. Using the inversion formula, this reads $(D_\alpha f)(x) = \int \xi^\alpha \hat{f}(\xi) e^{ix \cdot \xi} dm(\xi)$. More generally, given a partial differential operator $P(x, D) = \sum c_\alpha(x) D_\alpha$, we get $[P(x, D)f](x) = \int p(x, \xi) \hat{f}(\xi) e^{ix \cdot \xi} dm(\xi)$, where $p(x, \xi) := \sum c_\alpha(x) \xi^\alpha$ is called the *symbol* of P . However, this last operator has a meaning even if $p(x, \xi)$ is not a polynomial in ξ . We obtain *pseudo-differential operators* by using different $p(x, \xi)$. For example, if $\lambda^s(\xi) = (1 + |\xi|^2)^{s/2}$, we may define $[\lambda^s(D)f](x) = \int \lambda^s(\xi) e^{ix \cdot \xi} \hat{f}(\xi) dm(\xi)$ for arbitrary $s > 0$. Differential operators are thus thought of as functions $p(x, \xi)$ on the *phase space* $\mathbb{R}_x^n \times \mathbb{R}_\xi^n$. To work *microlocally* means to work on the phase space rather than on \mathbb{R}^n . This leads to a more coherent theory. A classic treatise here is the 4 volumes [21]. Shorter accounts can be found in [11], [43], [33] and [45].

2.6 Exercises

1. Compute the following limit and justify the calculation:

$$\lim_{t \rightarrow 0} \int_0^\infty \frac{\sin(xt)}{t} \cdot \frac{1}{x(1+x^2)} dx.$$

2. Let $F(t) = \int_0^5 \frac{e^{xt}}{\sqrt{x}} dx$ for $t \in \mathbb{R}$. Compute $F'(0)$ and justify your answer.
3. Let (X, μ) be a σ -finite measure space and $K \in L^p(X \times X, \mu \times \mu)$, $1 \leq p \leq \infty$. Define $(T_K f)(x) = \int_X f(y)K(x, y) d\mu(y)$ for $f \in L^q(X)$, where $\frac{1}{p} + \frac{1}{q} = 1$. Show that T_K is a bounded linear operator from $L^q(X) \rightarrow L^p(X)$, with $\|T_K\| \leq \|K\|_{L^p(X \times X)}$.
4. Assuming all integrals in question exist, prove that

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

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

(c) For $z \in \mathbb{R}^n$, $\tau_z(f * g) = (\tau_z f) * g = f * (\tau_z g)$.

(d) $\text{supp}(f * g) \subseteq \overline{\text{supp}(f) + \text{supp}(g)}$, where $A + B = \{a + b : a \in A, b \in B\}$.

5. Let $f(x) = e^{-x^2}$ and $g(x) = x$ on \mathbb{R} . Compute $(f * g)(x)$.

6. Let p, q be conjugate exponents, $1 \leq p \leq \infty$. Let $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$.

(1) Show that $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$.

(2) Show that $f * g$ is uniformly continuous.

*Hint: you must show that $\lim_{y \rightarrow 0} \|\tau_y(f * g) - f * g\|_\infty = 0$. For this, use (1), Exercise 4(c) and continuity of translations in L^p norm.*

(3) Show that if $1 < p < \infty$, so that $1 < q < \infty$, then $f * g \in C_0(\mathbb{R}^n)$.

*Hint: choose $\{f_n\}, \{g_n\}$ in $C_c(\mathbb{R}^n)$ such that $\|f_n - f\|_p \rightarrow 0$ and $\|g_n - g\|_q \rightarrow 0$. Show that $f_n * g_n \in C_c(\mathbb{R}^n)$, $\|f_n * g_n - f * g\|_\infty \rightarrow 0$ and deduce the result.*

7. Let $\rho \in L^1(\mathbb{R}^n)$ with $\int \rho(x) dx = a$. If f is bounded and uniformly continuous, show that $\rho_\varepsilon * f \rightarrow af$ uniformly as $\varepsilon \rightarrow 0$.

8. Suppose $f, g \in L^1(\mathbb{R}^n)$ and $z \in \mathbb{R}^n$ and $a > 0$. Prove that

$$(a) \widehat{\tau_z f} = e_{-z} \widehat{f}, \quad (b) \widehat{e_z f} = \tau_z \widehat{f}, \quad (c) \widehat{S_a f} = a^n S_{1/a} \widehat{f}.$$

9. Using a change of variables, show that

$$\widehat{f}(\xi) = \int (\tau_{\frac{\pi\xi}{|\xi|^2}} f)(x) e^{-ix \cdot \xi} e^{\pi i} dm(x).$$

Deduce that

$$2\widehat{f}(\xi) = \int (f - \tau_{\frac{\pi\xi}{|\xi|^2}} f)(x) e^{-ix \cdot \xi} dm(x).$$

Conclude that $|\widehat{f}(\xi)| \rightarrow 0$ as $|\xi| \rightarrow \infty$ using continuity of translations in the L^1 norm. This gives another proof of the Riemann-Lebesgue lemma.

10. Given $a < b$, let $g = \chi_{[a,b]}$ on \mathbb{R} . Compute $\widehat{g}(\xi)$.

11. Recall the *Schwartz space* studied in the exercises of Chapter 1:

$$\mathcal{S}(\mathbb{R}^n) = \{\varphi \in C^\infty(\mathbb{R}^n) \mid p_{\alpha,\beta}(\varphi) := \sup_{x \in \mathbb{R}^n} |x^\alpha D^\beta \varphi(x)| < \infty \text{ for all } \alpha, \beta \in \mathbb{N}_0^n\}.$$

The reader proved in particular that $\mathcal{S}(\mathbb{R}^n)$ is a Fréchet space when endowed with $\mathcal{P} = \{p_{\alpha,\beta}\}$, and that $\mathcal{S}(\mathbb{R}^n) \subset \bigcap_{1 \leq p \leq \infty} L^p(\mathbb{R}^n)$, with $\|\varphi\|_{L^p} \leq (2\pi)^n \|\varphi\|_{2n}$ for $\varphi \in \mathcal{S}(\mathbb{R}^n)$. Here $\|\varphi\|_k = \max_{|\alpha+\beta| \leq k} p_{\alpha,\beta}(\varphi)$.

- (a) (Continuity of differentiation). Let $\alpha \in \mathbb{N}_0^n$ and $\varphi \in \mathcal{S}(\mathbb{R}^n)$. Show that $D^\alpha \varphi \in \mathcal{S}(\mathbb{R}^n)$ with $\|D^\alpha \varphi\|_k \leq \|\varphi\|_{k+|\alpha|}$ for all $k \in \mathbb{N}_0$.
- (b) (Continuity of multiplication). Let $\psi \in C^\infty(\mathbb{R}^n)$ have polynomial growth. Show that there are C_k, N_k such that $\varphi \in \mathcal{S}(\mathbb{R}^n) \implies \psi\varphi \in \mathcal{S}(\mathbb{R}^n)$, with $\|\psi\varphi\|_k \leq C_k \|\varphi\|_{k+2N_k}$. In particular, if $\psi(x) = x^\alpha$, show that $\|x^\alpha \varphi\|_k \leq 2^k (\alpha!) \|\varphi\|_{k+|\alpha|}$.
- (c) (Continuity of Fourier transform). Show that if $\varphi \in \mathcal{S}(\mathbb{R}^n)$, then $\widehat{\varphi} \in \mathcal{S}(\mathbb{R}^n)$, and $\|\widehat{\varphi}\|_k \leq (8\pi)^n (k+1)! \|\varphi\|_{2n+k}$.
- (d) Show that $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \rightarrow \mathcal{S}(\mathbb{R}^n)$ is an isomorphism.
- 12.** Recall that a *tempered distribution* is a continuous linear functional on $\mathcal{S}(\mathbb{R}^n)$. If $T \in \mathcal{S}'(\mathbb{R}^n)$, we define $\widehat{T} : \mathcal{S}(\mathbb{R}^n) \rightarrow \mathbb{C}$ by $\widehat{T}(\varphi) = T(\widehat{\varphi})$.
- (a) Check that $\widehat{T} \in \mathcal{S}'(\mathbb{R}^n)$.
- (b) Check that if $g \in L^1(\mathbb{R}^n)$, then $\widehat{T}_g = T_{\widehat{g}}$.
- (c) Compute $\widehat{\delta}_b$ and $\widehat{\delta}'_b$ for $b \in \mathbb{R}^n$.
- (d) Check that \mathcal{F} is continuous on $\mathcal{S}'(\mathbb{R}^n)$, in the sense that $T_n \rightarrow T$ in $\mathcal{S}'(\mathbb{R}^n)$ implies $\widehat{T}_n \rightarrow \widehat{T}$ in $\mathcal{S}'(\mathbb{R}^n)$.
- (e) Show that $(ix)^\alpha \widehat{T} = \widehat{D^\alpha T}$ and $D^\beta \widehat{T} = \mathcal{F}[(-ix)^\beta T]$.
- 13.** Show that if $f, g \in \mathcal{S}(\mathbb{R}^n)$, then $f * g \in \mathcal{S}(\mathbb{R}^n)$.
*Hint: show that $f * g = (2\pi)^{n/2} \mathcal{F}^{-1}(\widehat{f\widehat{g}})$.*
- 14.** Show that $\|f\|_{k,p} := (\sum_{|\alpha| \leq k} \|D_w^\alpha f\|_p^p)^{1/p}$ and $N_{k,p}(f) := \sum_{|\alpha| \leq k} \|D_w^\alpha f\|_p$ are equivalent norms on $W^{k,p}(\mathbb{R}^n)$.
- 15.** Given $s > 0$, let $\lambda^s(\xi) = (1 + |\xi|^2)^{s/2}$. We define⁷

$$H^s(\mathbb{R}^n) = \{u \in L^2(\mathbb{R}^n) : \lambda^s \widehat{u} \in L^2(\mathbb{R}^n)\}$$

and endow it with $\|u\|_{(s)}^2 = \int |\lambda^s(\xi) \widehat{u}(\xi)|^2 dm(\xi)$.

(1) Show that $|\lambda^{s+1}(\xi) \widehat{u}(\xi)|^2 = |\lambda^s(\xi) \widehat{u}(\xi)|^2 + \sum_{j=1}^n |\lambda^s(\xi) \widehat{\partial_j^w u}(\xi)|^2$.

(2) Deduce that $u \in H^{s+1}(\mathbb{R}^n) \iff u, \partial_1^w u, \dots, \partial_n^w u \in H^s(\mathbb{R}^n)$, with

$$\|u\|_{(s+1)}^2 = \|u\|_{(s)}^2 + \sum_{|\alpha|=1} \|D_w^\alpha u\|_{(s)}^2.$$

(3) Deduce that $H^k(\mathbb{R}^n) = W^{k,2}(\mathbb{R}^n)$ for $k \in \mathbb{N}$.

16. (Heisenberg uncertainty principle). Let $f \in C_c^1(\mathbb{R})$.

1. Show that $\|f\|_{L^2}^2 = -2 \operatorname{Re} \int_{\mathbb{R}} x f \widehat{f}'$.

2. Deduce that $\|f\|_{L^2}^2 \leq 2 \|x f\|_{L^2} \|\xi \widehat{f}\|_{L^2}$.

3. Deduce that for any $x_0, \xi_0 \in \mathbb{R}$, $\|f\|_{L^2}^2 \leq 2 \|(x - x_0) f\|_{L^2} \|(\xi - \xi_0) \widehat{f}\|_{L^2}$.

Hint: consider $g(x) = e^{-i\xi_0 x} f(x + x_0)$.

Thus, f and \widehat{f} cannot both be sharply localized about single points x_0, ξ_0 . This inequality can be generalized to all $f \in L^2(\mathbb{R})$ (but the RHS may be infinite).

7. Actually, $H^s(\mathbb{R}^n)$ is defined for all $s \in \mathbb{R}$, not just $s > 0$, but in this case, we consider $u \in \mathcal{S}'(\mathbb{R}^n)$.

17. (Fourier series). Let $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$ be the one-dimensional torus, which is obtained from \mathbb{R} by identifying points which differ by $2\pi k$, for some $k \in \mathbb{Z}$. We denote $L^p(\mathbb{T}) = L^p[0, 2\pi]$ and regard functions in $L^p(\mathbb{T})$ as maps defined on the circle \mathbb{T} .

It is well known that if $e_k(x) = \frac{1}{\sqrt{2\pi}}e^{ikx}$ on $[0, 2\pi]$, then $\{e_k\}_{k \in \mathbb{Z}}$ is an orthonormal basis of $L^2(\mathbb{T})$. It follows from Hilbert space theory that any $f \in L^2(\mathbb{T})$ can be written as $f = \sum_{k \in \mathbb{Z}} \langle f, e_k \rangle e_k$, and that $\langle f, g \rangle = \sum_{k \in \mathbb{Z}} \langle f, e_k \rangle \overline{\langle g, e_k \rangle}$ for any $f, g \in L^2(\mathbb{T})$. Now if for $f \in L^2(\mathbb{T})$ and $k \in \mathbb{Z}$ we define

$$(\mathcal{F}f)_k = \hat{f}_k = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) e^{-ikx} dx,$$

we have $\langle f, e_k \rangle = \hat{f}_k$. Hence, $f = \sum_{k \in \mathbb{Z}} \hat{f}_k e_k$. Moreover, since $\sum_{k \in \mathbb{Z}} |\hat{f}_k|^2$ converges to $\|f\|_{L^2(\mathbb{T})}^2$, it follows that $\mathcal{F} : L^2(\mathbb{T}) \rightarrow \ell^2(\mathbb{Z})$, and

$$\langle f, g \rangle_{L^2(\mathbb{T})} = \langle \hat{f}, \hat{g} \rangle_{\ell^2(\mathbb{Z})} \quad (\text{Parseval's identity}).$$

- Let $C^1(\mathbb{T}) = \{f \in C^1[0, 2\pi] : f(0) = f(2\pi)\}$. Show that for $f \in C^1(\mathbb{T})$, we have $(\hat{f}')_k = ik\hat{f}_k$.
- Let $\lambda^s(k) = (1 + k^2)^{s/2}$ for $k \in \mathbb{Z}$. Given $s > 0$, we define the Sobolev space

$$H^s(\mathbb{T}) = \left\{ f \in L^2(\mathbb{T}) : \sum_{k \in \mathbb{Z}} \lambda^{2s}(k) |\hat{f}_k|^2 < \infty \right\}$$

and endow it with the norm $\|f\|_{(s)}^2 = \sum_{k \in \mathbb{Z}} \lambda^{2s}(k) |\hat{f}_k|^2$. Note that $\|f\|_{(0)} = \|f\|_{L^2(\mathbb{T})}$ by Parseval.

We say that $h \in L^2(\mathbb{T})$ is a *weak derivative* of $f \in L^2(\mathbb{T})$ if $\langle h, \varphi \rangle = -\langle f, \varphi' \rangle$ for all $\varphi \in C^1(\mathbb{T})$. In this case, we denote $h = f'_w$. Show that

$$H^1(\mathbb{T}) = \{f \in L^2(\mathbb{T}) : f'_w \text{ exists and } f'_w \in L^2(\mathbb{T})\},$$

and that $\|f\|_{(1)}^2 = \|f\|_{L^2(\mathbb{T})}^2 + \|f'_w\|_{L^2(\mathbb{T})}^2$.

- Let $S_N = \sum_{k=-N}^N \hat{f}_k e_k$. Show that if $s > 1/2$, there exists C_s such that for all $f \in H^s(\mathbb{T})$, we have $\|S_N - f\|_\infty \leq \frac{C_s}{N^{s-1/2}} \|f\|_{(s)}$.

Hence, in this case, the Fourier series S_N converges not only in $L^2(\mathbb{T})$ (which is known by Hilbert space theory), but also uniformly.

(Hint: $\|S_N - f\|_\infty \leq \frac{1}{\sqrt{2\pi}} \sum_{|k| > N} |\hat{f}_k| \leq \frac{1}{\sqrt{2\pi}} \|f\|_{(s)} (\int_N^\infty r^{-2s} dr)^{1/2}$.)

- Deduce the following Sobolev embedding: if $s > 1/2$, then $H^s(\mathbb{T}) \hookrightarrow C(\mathbb{T})$. (Hint: $S_N \in C(\mathbb{T})$ for each N .)

- Let $f \in C^1(\mathbb{R})$ and suppose there exist $C, \epsilon > 0$ such that

$$|(1 + x^2)^{1/2+\epsilon} f(x)| \leq C \quad \text{and} \quad |(1 + x^2)^{1/2+\epsilon} f'(x)| \leq C$$

for all $x \in \mathbb{R}$. Show that the *Poisson summation formula* holds:

$$\sum_{k \in \mathbb{Z}} f(x + 2\pi k) = \sum_{k \in \mathbb{Z}} \hat{f}(k) e_k(x).$$

Here \hat{f} is the Fourier transform of $L^1(\mathbb{R})$ functions, i.e. $\hat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ixk} dx$.

In particular, for $x = 0$ we get $\sum_{k \in \mathbb{Z}} f(2\pi k) = \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} \hat{f}(k)$.

(Hint: let $g(x) = \sum_{k \in \mathbb{Z}} f(x + 2\pi k)$ for $x \in [0, 2\pi]$. The hypotheses imply that this series converge uniformly, so $g \in C^1(\mathbb{T}) \subset H^1(\mathbb{T})$. Hence, $g = \sum \hat{g}_k e_k$. By (3), the Fourier series of g converges uniformly, so we may calculate \hat{g}_k by integrating termwise.

6. Deduce that $\coth(\pi a) = \frac{1 + e^{-2\pi a}}{1 - e^{-2\pi a}} = \frac{1}{\pi} \sum_{k \in \mathbb{Z}} \frac{a}{a^2 + k^2}$.

(Hint: Apply Poisson to $f(x) = e^{-a|x|}$.)

Chapter 3

Some Important Theorems

This chapter is a potpourri of classical results which are regularly used in analysis, and generally assumed to be well known. From now on, X is a vector space over $\mathbb{K} = \mathbb{R}$ or \mathbb{C} , and if X is locally convex, we denote its topological dual by X^* instead of X' . Thus, $X^* = \{\lambda : X \rightarrow \mathbb{K} \mid \lambda \text{ is linear and continuous}\}$.

3.1 Measure theory

We start by giving some fundamental results of measure theory, which are not always covered in undergraduate courses. Here we follow [31].

Definition 3.1. A *positive measure* is a countably additive function defined on a σ -algebra, with range in $[0, \infty]$. A *complex measure* is a complex-valued countably additive function defined on a σ -algebra.

Hence, a positive measure is just an ordinary measure with ∞ as an admissible value. On the other hand, when we talk about complex measures, $\mu(E)$ is always a complex number. So a positive measure is not necessarily a complex measure.

Definition 3.2. Let μ be a complex measure on a σ -algebra \mathfrak{M} . We define the *total variation* of μ as the function $|\mu| : \mathfrak{M} \rightarrow [0, \infty]$ given by $|\mu|(E) := \sup \sum_{i=1}^{\infty} |\mu(E_i)|$, where the supremum is taken over all measurable partitions $\{E_i\}_{i=1}^{\infty}$ of E .

Lemma 3.3. $|\mu|$ is a positive measure on \mathfrak{M} , and $|\mu|(X) < \infty$.

Proof. See the Appendix, Section 3.7. □

Definition 3.4. Let μ be a positive measure on a σ -algebra \mathfrak{M} and let ν be a positive or complex measure on \mathfrak{M} . We say that ν is *absolutely continuous* with respect to μ , and denote $\nu \ll \mu$, if for each $E \in \mathfrak{M} : \mu(E) = 0$ implies $\nu(E) = 0$.

We say that ν is *concentrated* on a set $A \in \mathfrak{M}$ if $\nu(E) = \nu(A \cap E)$ for every $E \in \mathfrak{M}$.

We say that two measures ν_1 and ν_2 on \mathfrak{M} are *mutually singular*, and denote $\nu_1 \perp \nu_2$, if there are disjoint A, B such that ν_1 is concentrated on A and ν_2 is concentrated on B .

The following result is “probably the most important theorem in measure theory”, according to [31].

Theorem 3.5 (Lebesgue-Radon-Nikodym). *Let (X, \mathfrak{M}) be a measurable space. Let μ be a σ -finite positive measure on \mathfrak{M} and ν a complex measure on \mathfrak{M} .*

- (a) *There is a unique pair of complex measures ν_a and ν_s on \mathfrak{M} such that $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$ and $\nu_s \perp \mu$.*
- (b) *There is a unique $h \in L^1(\mu)$ such that $\nu_a(E) = \int_E h d\mu$ for every $E \in \mathfrak{M}$.*

Proof. See the Appendix, Section 3.7. □

This theorem has many applications. For instance, any complex measure gets a *polar decomposition* $d\mu = h d|\mu|$ for some measurable h with $|h(x)| = 1$. Consequently, we can define integration w.r.t μ by $\int f d\mu := \int fh d|\mu|$. One also obtains the *Hahn decomposition* for signed measures μ . Namely, X can be written as a disjoint union of two measurable sets A, B such that μ is positive on A and negative on B . Finally, we obtain the following important consequence:

Theorem 3.6. *Let (X, \mathfrak{M}, μ) be a measure space with μ σ -finite. Let $1 \leq p < \infty$ and q its conjugate exponent (i.e. $\frac{1}{p} + \frac{1}{q} = 1$). Define $\Phi : L^q(\mu) \rightarrow L^p(\mu)^*$ by $g \mapsto \lambda_g$, where $\lambda_g(f) = \int fg$. Then Φ is an isometric isomorphism.*

In particular, the dual of ℓ^p is ℓ^q (take $X = \mathbb{N}$ with the counting measure).

Proof. See the Appendix, Section 3.7. □

We conclude this section by discussing the dual of $C(X)$.

Definition 3.7. Let X be a locally compact Hausdorff space. We say that μ is a *Borel measure* on X if μ is defined on the σ -algebra of Borel sets.

If μ is positive, we say it is *regular* if for every Borel set $E \subseteq X$, we have

$$\mu(E) = \inf_{V \supseteq E, V \text{ open}} \mu(V) = \sup_{K \subseteq E, K \text{ compact}} \mu(K).$$

If μ is complex, we say it is regular if $|\mu|$ is regular. We denote

$$\begin{aligned} \mathcal{M}(X) &= \{ \text{regular complex Borel measures on } X \}, \\ \mathcal{M}_{+,1}(X) &= \{ \text{regular Borel probability measures on } X \}, \end{aligned}$$

and endow $\mathcal{M}(X)$ and $\mathcal{M}_{+,1}(X)$ with the total variation norm $\|\mu\| = |\mu|(X)$.

Theorem 3.8 (Riesz-Markov). *Let X be a locally compact Hausdorff space and define $\Phi : \mathcal{M}(X) \rightarrow C_0(X)^*$ by $\mu \mapsto \lambda_\mu$, where $\lambda_\mu(f) = \int f d\mu$. Then Φ is an isometric isomorphism.*

*If X is compact, the map $\mu \mapsto \lambda_\mu$ gives an isometric isomorphism of $\mathcal{M}_{+,1}(X)$ with $C(X)^*_{+,1} = \{\lambda \in C(X)^* : \|\lambda\| = 1 \text{ and } \lambda(f) \geq 0 \text{ whenever } f \geq 0\}$.*

This result is also known as the “Riesz representation theorem”.

Proof. See the Appendix, Section 3.7 for a sketch. □

Before moving on, note that by Theorem 3.6, if $p > 1$, then L^p is reflexive, i.e. $(L^p)^{**} = (L^q)^* = L^p$. However, $(L^1)^* = L^\infty$ but $(L^\infty)^* \neq L^1$. It turns out that $L^\infty(\mu)^*$ is the space of bounded finitely additive signed measures ν which are absolutely continuous with respect to μ . This is known as the *Kantorovitch representation theorem*, and is somehow reminiscent of Riesz-Markov. For a proof of this theorem, see [30, Section 19.3].

3.2 Separation of convex sets

In this section we follow [32]. Let us first recall the Hahn-Banach theorem:

Theorem 3.9 (Hahn-Banach). *Let X be a real vector space and suppose $p : X \rightarrow \mathbb{R}$ satisfies $p(x + y) \leq p(x) + p(y)$ and $p(tx) = tp(x)$ for $x, y \in X$ and $t \geq 0$. Let Y be a subspace of X . If $\lambda : Y \rightarrow \mathbb{R}$ is linear and $\lambda(x) \leq p(x)$ on Y , then λ has a linear extension $L : X \rightarrow \mathbb{R}$ such that $L(x) = \lambda(x)$ on Y and $L(x) \leq p(x)$ on X .*

Although this theorem is sometimes stated in the special case of normed spaces, the proof of the above statement is exactly the same. Let us also recall some properties of the Minkowski functional defined in the Exercise 2 of Chapter 1.

If C is a convex absorbing set in a vector space X , we define the Minkowski functional $\rho_C : X \rightarrow \mathbb{R}$ by $\rho_C(x) = \inf\{t > 0 : t^{-1}x \in C\}$. The reader was asked to prove that $\rho_C(x + y) \leq \rho_C(x) + \rho_C(y)$ and $\rho_C(tx) = t\rho_C(x)$ if $t \geq 0$. If C is moreover open, then $C = \{x \in X : \rho_C(x) < 1\}$. Indeed, if $x \in C$, then $(1 + \varepsilon)x \in C$ for $\varepsilon > 0$ small enough, so $\rho_C(x) \leq \frac{1}{1+\varepsilon} < 1$. Conversely, if $\rho_C(x) < 1$, then $\exists t \in (0, 1)$ with $t^{-1}x \in C$, so $x = t(t^{-1}x) + (1 - t)0 \in C$.

We now prove the following geometric version of the Hahn-Banach theorem.

Theorem 3.10. *Suppose X is a locally convex space defined by a separating family of seminorms. Let $A, B \subset X$ be disjoint nonempty convex sets.*

- (a) *If A is open, there exists $\lambda \in X^*$ and $\gamma \in \mathbb{R}$ such that $\operatorname{Re} \lambda(a) < \gamma \leq \operatorname{Re} \lambda(b)$ for every $a \in A$ and $b \in B$.*
- (b) *If A is compact and B is closed, there exists $\lambda \in X^*$ and $\gamma_1, \gamma_2 \in \mathbb{R}$ such that $\operatorname{Re} \lambda(a) < \gamma_1 < \gamma_2 < \operatorname{Re} \lambda(b)$ for every $a \in A$ and $b \in B$.*

This is sometimes called the *separating hyperplane theorem*. That's because, if we define $H_\gamma = \{y \in X : \lambda(y) = \gamma\}$, then this theorem says that A can be separated from B by the hyperplane H_γ . This is certainly not true if A or B is not convex, even in \mathbb{R}^2 .

Proof. It suffices to prove the theorem for real scalars. Indeed, if the field is \mathbb{C} and the real case is proved, then there is a continuous real-linear λ_1 on X that gives the required separation. Define $\lambda(x) = \lambda_1(x) - i\lambda_1(ix)$. Then λ is complex-linear and continuous, i.e. $\lambda \in X^*$, and $\operatorname{Re} \lambda = \lambda_1$. Hence, we assume the field is \mathbb{R} .

For (a), let $-x_0 \in A - B$ and let $C = A - B + \{x_0\}$. Then $0 \in C$, C is open, thus absorbing, and convex. Let ρ_C be the Minkowski functional of C . Since $A \cap B = \emptyset$, then $x_0 \notin C$, so $\rho_C(x_0) \geq 1$. Define λ on $Y = \{tx : t \in \mathbb{R}\}$ by $\lambda(tx_0) = t$. Then for $t \geq 0$, we have $\lambda(tx_0) = t \leq t\rho_C(x_0) = \rho_C(tx_0)$, and for $t < 0$, $\lambda(tx_0) < 0 \leq \rho_C(tx_0)$. Thus, $\lambda(y) \leq \rho_C(y)$ on Y . So by Hahn-Banach, we may extend λ linearly to all X and have $\lambda(x) \leq \rho_C(x)$ on X .

Now if $y \in C \cap (-C)$, then $\rho_C(\pm y) \leq 1$, so $\lambda(y) \leq 1$ and $-\lambda(y) = \lambda(-y) \leq 1$. Thus, $|\lambda(y)| \leq 1$ on $V = C \cap (-C)$, so λ is continuous. Indeed, since V is open and $0 \in V$, then given $\varepsilon > 0$, we may choose $\delta > 0$ and $p \in \overline{V}$ such that $B_{0,\delta}(p) \subset \varepsilon V$. Then $z \in B_{0,\delta}(p)$ implies $z = \varepsilon y$, $y \in V$, implies $|\lambda(z)| = \varepsilon|\lambda(y)| \leq \varepsilon$, so λ is continuous at 0, so by linearity, λ is continuous on X .

Finally, if $a \in A$ and $b \in B$, then $\lambda(a) - \lambda(b) + 1 = \lambda(a - b + x_0) \leq \rho_C(a - b + x_0) < 1$, since $\lambda(x_0) = 1$, $a - b + x_0 \in C$ and C is open. Thus, $\lambda(a) < \lambda(b)$. It follows that $\lambda(A)$ and $\lambda(B)$ are disjoint convex subsets of \mathbb{R} . Moreover, $\lambda(A)$ is open. Indeed, if $\lambda(x) \in \lambda(A)$, there is $\varepsilon > 0$ with $B_{x,\varepsilon}(p) \subseteq A$ for some $p \in \overline{A}$, since A is open. Taking a small $\varepsilon' > 0$

we have $x \pm \varepsilon'x_0 \in B_{x,\varepsilon}(p)$. Hence, $\lambda(x \pm \varepsilon'x_0) \in \lambda(A)$, i.e. $\lambda(x) \pm \varepsilon' \in \lambda(A)$, so $\lambda(A)$ is open. Hence, if we let $\gamma = \sup_{a \in A} \lambda(a)$, we get (a).¹

For (b), let V be a convex neighborhood of 0 in X such that $(A + V) \cap B = \emptyset$.² Then applying (a) to $A + V$ in place of A , there exists $\lambda \in X^*$ such that $\lambda(x) < \gamma \leq \lambda(b)$ for every $x \in A + V$ and $b \in B$. Let $\gamma_0 = \sup \lambda(A) = \lambda(a_0)$, since A is compact. Then $\lambda(a) \leq \gamma_0 = \lambda(a_0) < \gamma \leq \lambda(b)$ for any $a \in A$ and $b \in B$, so taking γ_1, γ_2 such that $\gamma_0 < \gamma_1 < \gamma_2 < \gamma$, we get (b). \square

Corollary 3.11. *If X is a locally convex space defined by a separating family of seminorms, then X^* separates points.*

Proof. If $x_1, x_2 \in X$, $x_1 \neq x_2$, apply Theorem 3.10(b) to $A = \{x_1\}$ and $B = \{x_2\}$. \square

3.3 Banach-Alaoglu

Definition 3.12. Let X be a locally convex space defined by a separating family of seminorms.

- (a) The *weak topology* on X is the topology induced by the family $\{p_\lambda : \lambda \in X^*\}$, where $p_\lambda(x) := |\lambda(x)|$. This family of seminorms is separating by Corollary 3.11.
- (b) The *weak* topology* on X^* is the topology on X^* defined by the family of seminorms $\{p_x : x \in X\}$, where $p_x(\lambda) := |\lambda(x)|$. This family is separating by definition.

- Remarks 3.13.**
- (i) The weak* topology is weaker than the weak topology on X^* , since the latter would be induced by $\{p_\eta : \eta \in X^{**}\}$, and any p_x is in this family, by taking $\eta(\lambda) := \lambda(x)$ for $\lambda \in X^*$, so that $\eta \in X^{**}$, and $p_\eta(\lambda) = |\eta(\lambda)| = |\lambda(x)| = p_x(\lambda)$.
 - (ii) A sequence $\{x_n\} \subset X$ converges weakly to x if $p_\lambda(x_n - x) \rightarrow 0$ for every $\lambda \in X^*$, i.e. if $\lambda(x_n) \rightarrow \lambda(x)$ for every $\lambda \in X^*$. This is of course weaker than the convergence of x_n to x in norm, since $|\lambda(x_n) - \lambda(x)| \leq \|\lambda\| \|x_n - x\|$.
 - (iii) A sequence $\{\lambda_n\} \subset X^*$ is weak* convergent to λ if $p_x(\lambda_n - \lambda) \rightarrow 0$ for every $x \in X$, i.e. if $\lambda_n(x) \rightarrow \lambda(x)$ for every $x \in X$.
 - (iv) If X is a Hilbert space, then by the Riesz representation theorem, $\{x_n\}$ converges weakly to x iff $\langle x_n - x, z \rangle \rightarrow 0$ for each $z \in X$.

Example 3.14. If $\{x_n\}$ is an orthonormal sequence in an infinite dimensional Hilbert space, then $\{x_n\}$ does not converge strongly, since $\|x_n - x_m\|^2 = 2$ for $n \neq m$. However, $\{x_n\}$ converges weakly to 0, because by Bessel's inequality, $\sum_{n=1}^{\infty} |\langle x_n, x \rangle|^2$ converges for every x , so $\langle x_n, x \rangle \rightarrow 0$ for every x . This shows that weak convergence is strictly weaker than convergence in norm.

Similarly, in ℓ^p with $1 < p < \infty$, the sequence $e_1 = (1, 0, 0, \dots)$, $e_2 = (0, 1, 0, 0, \dots)$, etc converges weakly to zero, but not in norm. Indeed, if $\frac{1}{p} + \frac{1}{q} = 1$, then $(\ell^p)^* = \ell^q$ via the identification $\ell^q \ni u \leftrightarrow f_u \in (\ell^p)^*$, where $f_u(x) = \sum x_i u_i$ for $x \in \ell^p$. Since $\sum |u_i|^q < \infty$ for every $u \in \ell^q$, then $f_u(e_i) = u_i \rightarrow 0$, so e_i converges weakly to 0.

In $X = C[a, b]$, if a sequence $\{f_n\} \subset X$ converges weakly to $f \in X$, then f_n converges pointwise to f . Indeed, $\lambda(f_n) \rightarrow \lambda(f)$ for any $\lambda \in X^*$, so fixing $t \in [a, b]$ and taking $\lambda_t(f) := f(t)$, we have $\lambda_t \in X^*$, so $f_n(t) \rightarrow f(t)$.

1. Since $\lambda(A)$ is open, then $\gamma \notin \lambda(A)$, for otherwise, there would exist $\varepsilon > 0$ with $\gamma + \varepsilon \in \lambda(A)$, which contradicts that $\gamma = \sup \lambda(A)$. Hence, $\lambda(a) < \gamma$ for all a .

2. If X is a normed space, just take a sufficiently small ball. In general, use [32, Theorem 1.10].

We now proceed to the most useful property of the weak* topology. We denote by X_1^* the closed unit ball of X^* , i.e. $X_1^* = \{\lambda \in X^* : \|\lambda\| \leq 1\}$. The following proof will involve the product topology. We remind the reader (see Exercise 6 in Chapter 1), that if $\{Y_\alpha\}_{\alpha \in I}$ is a collection of topological spaces and $Y = \prod_{\alpha \in I} Y_\alpha$, then the product topology on Y is the weak topology generated by the projections $\pi_\alpha : Y \rightarrow Y_\alpha$ which map $(x_\beta) \mapsto x_\alpha$. In other words, it is the weakest topology on Y which makes every π_α continuous.

Theorem 3.15 (Banach-Alaoglu). *For every normed space X , the ball X_1^* is compact in the weak* topology.*

Proof. Given $x \in X$, let $B_x = \{c \in \mathbb{K} : |c| \leq \|x\|\}$. Then B_x is compact, so $K := \prod_{x \in X} B_x$ is compact in the product topology, by Tychonoff's theorem. Define $i : X_1^* \rightarrow K$ by $\lambda \mapsto (\lambda(x))_{x \in X}$. This is well-defined: if $\lambda \in X_1^*$, then $\|\lambda\| \leq 1$ and thus $\lambda(x) \in B_x$. Clearly, i is injective. Moreover, if we let $K_{x,y,a,b} = \{\kappa \in K : \kappa_{ax+by} = a\kappa_x + b\kappa_y\}$ for $x, y \in X$ and $a, b \in \mathbb{K}$, then each $K_{x,y,a,b}$ is closed³ in K , and $i(X_1^*) = \bigcap_{x,y,a,b} K_{x,y,a,b}$. Thus, $i(X_1^*)$ is closed in K , hence compact. Finally, if $B_{\lambda,\varepsilon}(p_{x_1}, \dots, p_{x_k})$ is a semiball in X_1^* , then $i(B_{\lambda,\varepsilon}(p_{x_1}, \dots, p_{x_k})) = \{(\eta(x))_{x \in X} : |\lambda(x_1) - \eta(x_1)| < \varepsilon \wedge \dots \wedge |\lambda(x_k) - \eta(x_k)| < \varepsilon\} = \bigcap_{i=1}^k \pi_{x_i}^{-1}(\lambda(x_i) - \varepsilon, \lambda(x_i) + \varepsilon)$, which is open in K since all the π_x are continuous. Hence, the image of any open set under i is open, so i^{-1} is continuous on $i(X_1^*)$. Hence, $i^{-1}(i(X_1^*)) = X_1^*$ is compact (because i^{-1} maps compact sets to compact sets). \square

Theorem 3.16 (Sequential Banach-Alaoglu). *For every separable normed space X , the ball X_1^* is sequentially compact in the weak* topology.*

This theorem follows from Banach-Alaoglu because if X is separable, then X_1^* is metrizable, so compactness is equivalent to sequential compactness. However, we shall give an alternative proof which does not rely on Tychonoff, and only uses a “diagonal trick”.

Proof. Let $\{x_k\}_{k=1}^\infty$ be dense in X . The sequence $\{f_n(x_1)\}_{n \in \mathbb{N}}$ is bounded in \mathbb{C} , since $|f_n(x_1)| \leq \|f_n\| \|x_1\| \leq \|x_1\|$ for all n . Hence, $\{f_n(x_1)\}$ has a convergent subsequence $\{f_{\varphi_1(n)}(x_1)\}_{n \in \mathbb{N}}$ by Bolzano-Weierstrass. The sequence $\{f_{\varphi_1(n)}(x_2)\}$ is also bounded, so it has a convergent subsequence $\{f_{\varphi_1 \circ \varphi_2(n)}(x_2)\}_{n \in \mathbb{N}}$. And so on. Let $\psi_k = \varphi_1 \circ \dots \circ \varphi_k$. Then $\{f_{\psi_k(n)}(x_k)\}_{n \in \mathbb{N}}$ converges for each k .

Let $g_n = f_{\psi_n(n)}$. Then for any $k \in \mathbb{N}$, the sequence $\{g_n(x_k)\}_{n \in \mathbb{N}}$ converges, because $n > k$ implies that $\{g_n(x_k)\}$ is a subsequence of the convergent sequence $\{f_{\psi_k(n)}(x_k)\}_{n \in \mathbb{N}}$. Define $g(x_k) := \lim_{n \rightarrow \infty} g_n(x_k)$. Then $|g(x_k)| = \lim_n |g_n(x_k)| \leq \liminf \|g_n\| \|x_k\| \leq \|x_k\|$. So g is a bounded linear map on the dense set $\{x_k\}$, so g has a unique bounded linear extension \tilde{g} defined on all X . Let $x \in X$, say $x = \lim_k x_{\varphi(k)}$. Then

$$\begin{aligned} |\tilde{g}(x) - g_n(x)| &\leq |\tilde{g}(x - x_{\varphi(k)})| + |\tilde{g}(x_{\varphi(k)}) - g_n(x_{\varphi(k)})| + |g_n(x - x_{\varphi(k)})| \\ &\leq (\|\tilde{g}\| + \sup_n \|g_n\|) \|x - x_{\varphi(k)}\| + |\tilde{g}(x_{\varphi(k)}) - g_n(x_{\varphi(k)})|, \end{aligned}$$

and $\sup_n \|g_n\| \leq 1$. Given $\varepsilon > 0$, choose k large enough so that the first term is smaller than $\varepsilon/2$. Then $|\tilde{g}(x) - g_n(x)| \leq \varepsilon/2 + |\tilde{g}(x_{\varphi(k)}) - g_n(x_{\varphi(k)})|$. But $g_n(x_{\varphi(k)}) \rightarrow g(x_{\varphi(k)}) = \tilde{g}(x_{\varphi(k)})$ as $n \rightarrow \infty$. Hence, $g_n(x) \rightarrow \tilde{g}(x)$. Since x is arbitrary, the subsequence (g_n) of (f_n) is weak* convergent to \tilde{g} . \square

Corollary 3.17. *Any bounded sequence in a Hilbert space has a weakly convergent subsequence.*

3. as it is the preimage of the weak* closed set $\{0\}$ under the map $\kappa \mapsto \kappa_{ax+by} - a\kappa_x - b\kappa_y$, which is continuous in the product topology.

Proof. Let $\{x_n\}$ be bounded in \mathcal{H} and let $\mathcal{H}_0 = \overline{\text{sp}\{x_n\}}$. Then \mathcal{H}_0 is separable. Assume $\|x_n\| \leq M$ for all n and let $f_n(y) = \langle y, \frac{x_n}{M} \rangle$ for $y \in \mathcal{H}_0$. Then $\|f_n\| = \|\frac{x_n}{M}\| \leq 1$, so f_n is a bounded sequence in $(\mathcal{H}_0)_1^*$, and has a weak* convergent subsequence by Theorem 3.16, say $\{f_{n_k}(y)\}$ converges for all $y \in \mathcal{H}_0$. Hence, $\langle y, x_{n_k} \rangle$ converges for all $y \in \mathcal{H}_0$. Let P be the orthogonal from $\mathcal{H} \rightarrow \mathcal{H}_0$ and let $Q = I - P$. Then given $x \in \mathcal{H}$, $\langle Qx, x_{n_k} \rangle = 0$, since $Qx \perp \mathcal{H}_0$. Hence, $\langle x, x_{n_k} \rangle = \langle Px, x_{n_k} \rangle + \langle Qx, x_{n_k} \rangle = \langle Px, x_{n_k} \rangle$ converges. As $x \in \mathcal{H}$ is arbitrary, $\{x_{n_k}\}$ converges weakly. \square

This result can be strengthened as follows: a Banach space is reflexive iff every bounded sequence has a weakly convergent subsequence. This is known as the *Eberlein-Šmulian theorem*. For a proof, see [10, Chapter 9.8].

Example 3.18. Let $\mu_0 \in \mathcal{M}_{+,1}([0, 1])$ be the probability measure supported on $\{0\}$, and $\mu_n \in \mathcal{M}_{+,1}([0, 1])$ be the normalized Lebesgue measure supported on $[0, 1/n]$. By Theorem 3.8, we may identify $\mathcal{M}_{+,1}([0, 1])$ with a subspace of $C[0, 1]^*$. Choose $f_n \in C[0, 1]$ such that $\|f_n\| = 1$, $\int f_n d\mu_n = 1/2$ and $f_n(0) = 0$.⁴ Then $\int f_n d\mu_0 = f_n(0) = 0$. Hence, $\|\mu_n - \mu_0\| = \sup_{\|f\|=1} |\int f d(\mu_n - \mu_0)| \geq 1/2$, so μ_n does not converge to μ_0 in norm. But μ_n converges to μ_0 in the weak* topology. Indeed, if $f \in C[0, 1]$, then given $\varepsilon > 0$, we may choose n large enough so that $|f(x) - f(0)| < \varepsilon$ on $[0, 1/n]$ (by continuity of f at 0). Then $|\int f d\mu_n - \int f d\mu_0| = |\int (f(x) - f(0)) d\mu_n(x)| < \varepsilon$.

Corollary 3.19. *If X is a compact Hausdorff space, then $\mathcal{M}_{1,+}(X)$ is compact in the weak* topology.*

Proof. Let $C(X)_{1,+}^* = \bigcap_{f \geq 0} \{\lambda \in C(X)_1^* : \lambda(f) \geq 0\}$. For each f , $\{\lambda : \lambda(f) \geq 0\}$ is closed in the weak* topology. Thus, $C(X)_{1,+}^*$ is closed in $C(X)_1^*$, hence compact by Banach-Alaoglu. Thus, $\mathcal{M}_{1,+}(X)$ is compact by Theorem 3.8. \square

Our last application of Banach-Alaoglu gives an interesting property of $C(K)$.

Theorem 3.20 (Universality of $C(K)$). *Every normed space X is isometrically embedded into $C(K)$ for some compact topological space K .*

Proof. Let $K = X_1^*$, endowed with the weak* topology. Then K is compact by Banach-Alaoglu. Define $\varphi : X \rightarrow C(K)$ by $\varphi : x \mapsto E_x$, where $E_x(f) = f(x)$. Note that E_x is indeed continuous by definition of the weak* topology. Clearly φ is linear, and it is an isometric embedding because $\|E_x\| = \sup_{f \in K} |E_x(f)| = \sup_{f \in X_1^*} |f(x)| = \|x\|$, where we used the Hahn-Banach theorem in the last equality. \square

Theorem 3.21 (Banach-Mazur). *Every separable normed space X is isometrically embedded into $C[0, 1]$.*

Proof. (Sketch). When X is separable, the weak* topology on X_1^* becomes metrizable, so X can be embedded in $C(K)$ for a compact metric space K , and this essentially finishes the proof. For the details, see [25, Theorem 3.12] or [5, Corollary 12.14]. \square

3.4 Markov-Kakutani and Haar

Definition 3.22. Let $T : X \rightarrow X$ be a map on a set X . We say that $x \in X$ is a *fixed point* of T if $Tx = x$.

4. e.g. $f_n(x) = nx$ if $x \in [0, 1/n]$, $f_n(x) = 2 - nx$ if $x \in [1/n, 2/n]$ and $f_n(x) = 0$ on $[2/n, 1]$.

The reader probably learned the Banach fixed point theorem as an undergraduate. Namely, if X is a complete metric space and $T : X \rightarrow X$ is a strict contraction (meaning $d(Tx, Ty) \leq kd(x, y)$ with $0 \leq k < 1$), then T has a unique fixed point.

Here is another fixed point theorem:

Theorem 3.23 (Leray-Schauder-Tychonoff). *Let X be a locally convex space defined by a separating family of seminorms. Let K be a nonempty compact convex subset of X and let $T : K \rightarrow K$ be continuous. Then T has a fixed point.*

This theorem is much more difficult to prove; see [9, Chapter V.10]. It generalizes the Brouwer fixed point theorem, which says that if \overline{B}_1 is the closed unit ball of \mathbb{R}^n and $T : \overline{B}_1 \rightarrow \overline{B}_1$ is continuous, then T has a fixed point; a result which is already quite deep.

Following [7, 44], we shall prove a stronger result by more elementary means, but in the special case where the maps are linear in some sense.

Definition 3.24. Let X, Y be vector spaces, C a convex subset of X . We say that $T : C \rightarrow Y$ is *affine* if $T(\sum \alpha_j x_j) = \sum \alpha_j T(x_j)$ whenever $x_j \in C$, $\alpha_j \geq 0$ and $\sum \alpha_j = 1$.

Theorem 3.25 (Markov-Kakutani). *Let X be a locally convex space defined by a separating family of seminorms. Let K be a nonempty compact convex subset of X and let \mathcal{F} be a family of commuting continuous affine maps of K into itself. Then there is an $x_0 \in K$ such that $T(x_0) = x_0$ for all $T \in \mathcal{F}$.*

Proof. If $T \in \mathcal{F}$ and $n \geq 1$, define $T^{(n)} = \frac{1}{n} \sum_{k=0}^{n-1} T^k$. Then $T^{(n)} : K \rightarrow K$, since K is convex. Let $\mathcal{K} = \{T^{(n)}(K) : T \in \mathcal{F}, n \geq 1\}$. If $T_1, \dots, T_p \in \mathcal{F}$ and $n_1, \dots, n_p \geq 1$, then the commutativity of \mathcal{F} implies that $\emptyset \neq T_1^{(n_1)} \dots T_p^{(n_p)}(K) \subseteq \bigcap_{j=1}^p T_j^{(n_j)}(K)$. Thus, \mathcal{K} has the finite intersection property. Since \mathcal{K} is a collection of compact sets, it follows that there is an x_0 in $\bigcap \{B : B \in \mathcal{K}\}$. We claim that x_0 is the required common fixed point.

Let $T \in \mathcal{F}$ and $n \geq 1$. Then $x_0 \in T^{(n)}(K)$, so $\exists x \in K$ such that $x_0 = T^{(n)}(x) = \frac{1}{n}(x + T(x) + \dots + T^{n-1}(x))$. Hence, $T(x_0) - x_0 = \frac{1}{n}(T(x) + \dots + T^n(x)) - \frac{1}{n}(x + \dots + T^{n-1}(x)) = \frac{1}{n}(T^n(x) - x)$. Let p be one of the seminorms defining the topology. Then for each n we have $p(T(x_0) - x_0) \leq \frac{2B}{n}$, where $B = \sup\{p(y) : y \in K\} < \infty$ since K is compact. Since this holds for each n , it follows that $p(T(x_0) - x_0) = 0$. Since p is arbitrary and \mathcal{P} is separating, we thus get $T(x_0) = x_0$. \square

Definition 3.26. We say that G is a *compact abelian group* if G is a compact topological space which is also an abelian group.

We say that a measure μ on G is *invariant* if $\int f d\mu = \int f_g d\mu$ for any $g \in G$ and $f \in C(G)$, where f_g is the translate of f , namely $f_g(h) = f(h - g)$.

Corollary 3.27 (Haar measure). *There exists an invariant measure on any compact abelian group G . This measure is unique up to positive scalar multiple. We call it the Haar measure on G .*

This corollary can be generalized to locally compact groups. Moreover, the group can be non-abelian, but in this case one gets left and right Haar measures.

Proof. We only prove existence; uniqueness is proved by different arguments.

Let $\mu \in \mathcal{M}_{1,+}(G)$ and define $T_g \mu$ by $T_g \mu(f) = \mu(f_g)$. Then T_g maps $\mathcal{M}_{1,+}(G) \rightarrow \mathcal{M}_{1,+}(G)$ and is continuous in the vague (weak*) topology. Indeed, given $\varepsilon > 0$ and $f \in C(G)$, we have $p_f(\mu) = |\mu(f)|$, so $|p_f(T_g \mu - T_g \nu)| = |p_{f_g}(\mu - \nu)|$. Choosing $\delta = \varepsilon$, we have $|p_{f_g}(\mu - \nu)| < \delta \implies |p_f(T_g \mu - T_g \nu)| < \varepsilon$. Hence, T_g is continuous.

Clearly, $\mathcal{M}_{1,+}(G)$ is convex, and $\mathcal{M}_{1,+}(G)$ is compact by Corollary 3.19. Finally, the various T_g commute since G is abelian, and they are affine. It follows from Markov-Kakutani that there exists a common fixed point μ_{Haar} with the desired invariance property. \square

- Example 3.28.** (a) If $G = \mathbb{R}^n$, the Haar measure is the Lebesgue measure.
 (b) If $G = S^1$ is the unit circle, the Haar measure is arc length. For the torus $\mathbb{T}^n = (S^1)^n$, the Haar measure is the product of the arc length measures on the factors.
 (c) If G is discrete, the Haar measure is the counting measure.

3.5 Krein-Milman

The Krein-Milman theorem describes compact convex sets in terms of their extreme points. We shall follow [44]. If X is a vector space and $x, y \in X$, we denote

$$[x, y] = \{tx + (1-t)y : t \in [0, 1]\} \quad \text{and} \quad (x, y) = \{tx + (1-t)y : t \in (0, 1)\}.$$

Definition 3.29. Let X be a vector space and let $A \subseteq X$ be convex.

If $x \in A$, we say that x is an *extreme point* of A if

$$x \in [y, z] \text{ for some } y, z \in A \implies x = y \text{ or } x = z.$$

We let $\text{ex}(A)$ be the set of extreme points of A .

If $\emptyset \neq B \subseteq A$ is convex, we say that B is an *extreme set* in A if

$$y, z \in A \text{ and } (y, z) \cap B \neq \emptyset \implies [y, z] \subseteq B.$$

Remark 3.30. x is an extreme point of A iff $\{x\}$ is an extreme set in A . Indeed, if x is an extreme point of A and $\{x\} \cap (y, z) \neq \emptyset$, then $x \in (y, z)$, so $x \in [y, z]$ but $x \neq y$ and $x \neq z$, a contradiction. Hence, $\{x\} \cap (y, z) = \emptyset$ for any $y, z \in A$, so $\{x\}$ is trivially an extreme set. Conversely, if $\{x\}$ is an extreme set and $x \in [y, z]$, then $x = y$ or $x = z$ or $y < x < z$. The last case is impossible, since it implies $\{x\} \cap (y, z) \neq \emptyset$, and hence $[y, z] \subseteq \{x\}$, i.e. $y = z = x$, a contradiction. Hence, x is an extreme point.

Example 3.31. (a) In \mathbb{R}^n , the extreme points of a closed convex polyhedron are just the vertices. Every edge, face, etc. is an extreme set. Half a face is not an extreme set.

- (b) The set of extreme points of a closed euclidean ball in \mathbb{R}^n is the boundary sphere.
 (c) If X is a compact Hausdorff space, then $\text{ex}(\mathcal{M}_{+,1}(X)) = \{\delta_x : x \in X\}$, where δ_x is the probability measure supported on $\{x\}$. Indeed, if $C \subset X$ has $0 < \mu(C) < 1$ and $\mu_C(B) := \mu(C)^{-1}\mu(B \cap C)$, $\mu_{X \setminus C}(B) := \mu(X \setminus C)^{-1}\mu(B \setminus C)$, then taking $\theta = \mu(C)$ we get $\mu = \theta\mu_C + (1-\theta)\mu_{X \setminus C}$, so μ is not an extreme point.

If μ has the property that $\mu(A)$ is 0 or 1 for each $A \subseteq X$, and if $x \neq y$ are both in $\text{supp } \mu$, then we can find disjoint open sets B, C with $x \in B$ and $y \in C$. Hence, $1 = \mu(B \cup C) = \mu(B) + \mu(C)$, so $\mu(B) = 0$ or $\mu(C) = 0$ by the 0-1 hypothesis. This means that x and y cannot be both in $\text{supp } \mu$, so $\mu = \delta_x$ for some x .

Thus, $\text{ex}(\mathcal{M}_{+,1}(X)) \subseteq \{\delta_x : x \in X\}$. Suppose $\delta_x = a\mu + b\nu$ with $a + b = 1$. If $x \in A$, then $1 = \delta_x(A) = a\mu(A) + b\nu(A)$. Now $\mu(A) \leq 1$ and $\nu(A) \leq 1$. If $\mu(A) < 1$ or $\nu(A) < 1$, then $a\mu(A) + b\nu(A) < a + b = 1$, a contradiction. Hence, $\mu(A) = \nu(A) = 1$. Similarly, if $x \notin A$, then since $\mu(A) \geq 0$ and $\nu(A) \geq 0$, we get $\mu(A) = \nu(A) = 0$. Hence, $\mu = \nu = \delta_x$, so $\delta_x \in \text{ex}(\mathcal{M}_{+,1}(X))$. Thus, $\text{ex}(\mathcal{M}_{+,1}(X)) = \{\delta_x : x \in X\}$.

Definition 3.32. Let X be locally convex and let $A \subseteq X$. We define the *closed convex hull* of A , denoted $\overline{\text{co}}(A)$, the smallest closed convex set containing A . This set exists since intersections of convex/closed sets are convex/closed.

Theorem 3.33 (Krein-Milman). *Let X be a locally convex space defined by a separating family of seminorms. If $A \subseteq X$ is compact and convex, then $A = \overline{\text{co}}(\text{ex}(A))$.*

Proof. Let B be a closed extreme subset of A . We first prove that B contains an extreme point of A . Let \mathfrak{F} be the family of closed subsets of B which are extreme in A . Partially order \mathfrak{F} by $F_1 \geq F_2$ if $F_1 \subseteq F_2$. Then every chain $\{F_\alpha\}_{\alpha \in I}$ has an upper bound, namely $\bigcap_{\alpha \in I} F_\alpha$. This is closed, nonempty (Cantor's intersection theorem) and extreme in A : if $(y, z) \cap (\bigcap_{\alpha \in I} F_\alpha) \neq \emptyset$, then $(y, z) \cap F_\alpha \neq \emptyset$ for each α , so $[y, z] \subseteq F_\alpha$ for each α , so $[y, z] \subseteq \bigcap_{\alpha \in I} F_\alpha$. It follows from Zorn's lemma that \mathfrak{F} has a maximal element (i.e. a minimal closed subset of B which is extreme in A), call it F . If $x, y \in F$, $x \neq y$, then by Theorem 3.10(b), there is an $f \in X^*$ with $\text{Re } f(x) < \text{Re } f(y)$. Let $M = \max_{z \in F} \text{Re } f(z)$, which exists since F is compact, and let $D = \{w \in F : \text{Re } f(w) = M\}$. Then D is extreme in A ,⁵ $D \subset F$ and $D \neq F$ since $x \notin D$. This contradicts the minimality of F . Thus, F is a singleton $\{x_0\}$. We thus showed that B contains an extreme point of A (namely, x_0).

Since $\text{ex}(A) \subset A$ and A is closed and convex, then $C := \overline{\text{co}}(\text{ex}(A)) \subseteq A$. Suppose $C \neq A$. Then there is an $x \in A \setminus C$. Since C is closed and convex, by Theorem 3.10(b), we may find $\lambda \in X^*$ such that $\text{Re } \lambda(y) < \alpha < \text{Re } \lambda(x)$ for all $y \in C$. Let $M = \max_{z \in A} \text{Re } \lambda(z)$ and $B = \{z \in A : \text{Re } \lambda(z) = M\}$. Then B is extreme in A , and since $M > \alpha$, we have $B \cap C = \emptyset$. But by the preceding paragraph, B contains an extreme point of A , say $x_0 \in B \cap C$. This contradiction shows that $C = A$. \square

The Krein-Milman theorem has many applications in different fields. For example, it can be combined with Markov-Kakutani to prove the existence of ergodic measures on compact spaces (see [44, 36]). A stronger Krein-Milman theorem can be found in [32, 36]. This states that for A as in Theorem 3.33, any $p \in A$ is the barycenter of a probability measure μ on $\text{ex}(A)$, that is, $f(p) = \int_{z \in \text{ex}(A)} f(x) d\mu(x)$ for all $f \in X^*$. This gives a better representation than the convex series of extreme points given by the usual Krein-Milman. Let us mention, however, that the mere existence of extreme points in compact convex sets is powerful. In fact, this is all we'll need in the following section to prove the Stone-Weierstrass theorem. The Krein-Milman theorem is also used in game theory to solve matrix games (see [16]). As a final connection with the idea of extreme points, we mention the *Banach-Stone theorem*: let X and Y be compact Hausdorff spaces such that $C(X)$ and $C(Y)$ are linearly isometric. Then X and Y are homeomorphic. For a proof, see [9, Chapter V.8].

3.6 Stone-Weierstrass vs Monotone Class

In this section we follow [7, 2, 13].

The Stone-Weierstrass theorem is an important extension of the Weierstrass approximation theorem of $C[0, 1]$. To motivate the result, let us give some of its consequences:

Example 3.34. (a) Suppose $X \subset \mathbb{R}^n$ is compact. Let $P(X) \subset C(X)$ be the restrictions of polynomial functions on \mathbb{R}^n to X . Then $P(X)$ is dense in $C(X)$.

5. Indeed, if $y, z \in A$ and $(y, z) \cap D \neq \emptyset$, say $v = ty + (1-t)z \in D$, then $(y, z) \cap F \neq \emptyset$, so $[y, z] \subseteq F$, so $\text{Re } f(y) \leq M$ and $\text{Re } f(z) \leq M$. But $\text{Re } f(v) = M$. If $\text{Re } f(y) < M$ or $\text{Re } f(z) < M$, we would get $\text{Re } f(v) = t \text{Re } f(y) + (1-t) \text{Re } f(z) < M$, a contradiction. Hence, $\text{Re } f(y) = \text{Re } f(z) = M$. By linearity, if $w \in [y, z]$, then $\text{Re } f(w) = M$. Hence, $[y, z] \subseteq D$.

- (b) Let $S^1 \subset \mathbb{C}$ be the unit circle. Let $P(S^1)$ be the set of functions of the form $f(z) = \sum_{n=-N}^N a_n z^n$, for arbitrary $N \in \mathbb{N}$. Then $P(S^1)$ is dense in $C(S^1)$.
- (c) Let X be a compact Hausdorff space and let $\mathcal{A} \subset C(X \times X)$ be the set of all finite linear combinations of functions of the form $(x, y) \mapsto f(x)g(y)$, where $f, g \in C(X)$.⁶ Then \mathcal{A} is dense in $C(X \times X)$.

Theorem 3.35 (Stone-Weierstrass). *Let K be a compact Hausdorff space and $\mathcal{A} \subseteq C(K)$ a linear subspace. Suppose that*

- (i) \mathcal{A} is a subalgebra, i.e. is closed under multiplication.
- (ii) The constant function $1 \in \mathcal{A}$.
- (iii) \mathcal{A} separates points, i.e. if $x, y \in K$ and $x \neq y$, there is some $f \in \mathcal{A}$ such that $f(x) \neq f(y)$.
- (iv) If $\mathbb{K} = \mathbb{C}$, then $f \in \mathcal{A}$ implies $\bar{f} \in \mathcal{A}$.

Then \mathcal{A} is dense in $C(K)$.

Proof. (de Branges). Let $\mathcal{A}^\perp = \{\mu \in C(K)^* : \int g d\mu = 0 \forall g \in \mathcal{A}\}$. It suffices to show that $\mathcal{A}^\perp = \{0\}$.⁷ Suppose $\mathcal{A}^\perp \neq \{0\}$. By Banach-Alaoglu, \mathcal{A}^\perp is weak* compact, so by Krein-Milman, there is an extreme point $\mu \in \mathcal{A}^\perp$. Then $\mu \neq 0$, since $0 = \frac{1}{2}\nu + \frac{1}{2}(-\nu)$ for $\nu \in \mathcal{A}^\perp \setminus \{0\}$ shows that 0 is not extreme. Thus, $0 < \|\mu\| \leq 1$. If $\|\mu\| < 1$, then $\mu = \|\mu\|(\mu/\|\mu\|) + (1 - \|\mu\|)0$ implies $\mu = \mu/\|\mu\| = 0$, a contradiction. Hence, $\|\mu\| = 1$. In particular, if $C = \text{supp } \mu$, then $C \neq \emptyset$. Fix $x_0 \in C$. We show that $C = \{x_0\}$.

Let $x \in C$, $x \neq x_0$. By (iii), there is an $f_1 \in \mathcal{A}$ with $f_1(x_0) \neq f_1(x) =: \beta$. By (ii), the constant function $\beta \in \mathcal{A}$. Hence, $f_2 = f_1 - \beta \in \mathcal{A}$, $f_2(x_0) \neq 0 = f_2(x)$. By (iv), $f_3 = |f_2|^2 = f_2 \bar{f}_2 \in \mathcal{A}$. Also, $f_3(x) = 0 < f_3(x_0)$ and $f_3 \geq 0$. Put $f = (\|f_3\| + 1)^{-1} f_3$. Then $f \in \mathcal{A}$, $f(x) = 0$, $f(x_0) > 0$ and $0 \leq f < 1$. By (i), gf and $g(1 - f) \in \mathcal{A}$ for every $g \in \mathcal{A}$. Since $\mu \in \mathcal{A}^\perp$, we get $0 = \int gf d\mu = \int g(1 - f) d\mu$ for every $g \in \mathcal{A}$. Hence, $f\mu$ and $(1 - f)\mu \in \mathcal{A}^\perp$. (Here $h\mu$ denotes the measure $E \mapsto \int_E h d\mu$).

Put $\alpha = \|f\mu\| = \int f d|\mu|$. Since $f(x_0) > 0$, there is an open neighborhood U of x_0 and an $\varepsilon > 0$ such that $f(y) > \varepsilon$ for $y \in U$. Thus, $\alpha = \int f d|\mu| \geq \int_U f d|\mu| \geq \varepsilon|\mu|(U) > 0$, since $U \cap C \neq \emptyset$. Similarly, $1 - f > 0$ (as $f < 1$) implies $\alpha < 1$. Hence, $0 < \alpha < 1$. Moreover, $1 - \alpha = \int (1 - f) d|\mu| = \|(1 - f)\mu\|$. Hence, $\mu = \alpha(f\mu/\|f\mu\|) + (1 - \alpha)[(1 - f)\mu/\|(1 - f)\mu\|]$, which implies $\mu = f\mu/\|f\mu\|$. But the only way the measures μ and $\alpha^{-1}f\mu = f\mu/\|f\mu\|$ can be equal, is that $\alpha^{-1}f = 1$ μ -a.e. Since f is continuous, this implies $f \equiv \alpha$ on C . In particular, $f(x_0) = f(x) = \alpha$. But $f(x) = 0 \neq \alpha$. This contradiction shows that $C = \{x_0\}$. Hence, $\mu = \gamma\delta_{x_0}$ for some $|\gamma| = 1$. But $\mu \in \mathcal{A}^\perp$ and $1 \in \mathcal{A}$, so $0 = \int 1 d\mu = \gamma$, a contradiction. Thus, $\mathcal{A}^\perp = \{0\}$, so $\mathcal{A} = C(K)$. \square

Using Banach-Alaoglu and Krein-Milman to prove Stone-Weierstrass is perhaps overkill. A more standard proof can be found in [28]. The previous theorem can be extended to $C_0(X)$ with X locally compact without difficulty; see [7, Corollary 8.3].

We now give an analog of the Stone-Weierstrass theorem which holds in the measurable world. This is frequently used in probability theory.

Theorem 3.36 (Functional Monotone Class). *Let \mathcal{V} be a family of bounded real valued-functions on a set Ω such that*

- (i) \mathcal{V} is a real vector space.

6. This is just the algebraic tensor product $C(X) \otimes C(X)$.

7. If $\bar{\mathcal{A}} \neq C(K)$, take $f \in C(K) \setminus \bar{\mathcal{A}}$. By Hahn-Banach [22, Lemma 4.6-7], there is a $\lambda \in C(K)^*$ with $\lambda(f) = 1$ and $\lambda(g) = 0$ for all $g \in \mathcal{A}$. Hence, $\lambda \in \mathcal{A}^\perp$ is nonzero, i.e. $\mathcal{A}^\perp \neq \{0\}$.

(ii) The constant function $1 \in \mathcal{V}$.

(iii) If $f_n \in \mathcal{V}$ are nonnegative, increasing and $\sup_n \|f_n\|_\infty < \infty$, then $f := \lim_n f_n \in \mathcal{V}$.

Suppose C is a subset of \mathcal{V} which is closed under multiplication. Then \mathcal{V} contains all bounded $\sigma(C)$ -measurable functions.

Proof. We first prove that \mathcal{V} is closed under uniform convergence. So let $\{f_n\} \subset \mathcal{V}$ with $\|f_n - f\|_\infty \rightarrow 0$. By passing to a subsequence, we may assume $\|f_{n+1} - f_n\|_\infty \leq 2^{-n}$. Let $g_n = f_n - f_1 - 2^{1-n} + 1$. Then $g_n \in \mathcal{V}$ since \mathcal{V} is a vector space containing the constant functions. Moreover, $\|g_n\|_\infty \leq \|f_n\|_\infty + \|f_1\|_\infty + 2$, so g_n is uniformly bounded. Finally, $g_{n+1} - g_n = f_{n+1} - f_n + 2^{-n} \geq 0$ for $n \geq 1$ and $g_1 = 0$. It follows from (iii) that $\lim g_n \in \mathcal{V}$. Hence, $f = (\lim g_n) + f_1 - 1 \in \mathcal{V}$.

Now let $\mathcal{L} = \{D \in \sigma(C) : 1_D \in \mathcal{V}\}$. By (i), (ii), (iii), \mathcal{L} is a λ -system⁸ on Ω . We will show that \mathcal{L} contains a π -system \mathcal{P} generating $\sigma(C)$. By the Dynkin π - λ theorem, it will follow that $\mathcal{L} \supseteq \sigma(\mathcal{P}) = \sigma(C)$, hence $\mathcal{L} = \sigma(C)$. Consequently, \mathcal{V} will contain the indicator function 1_D of every $D \in \sigma(C)$, and this will complete the proof since every bounded $\sigma(C)$ -measurable function is the pointwise limit of an increasing sequence of simple functions, and \mathcal{V} is closed under such limits by (iii).

Let $f_1, \dots, f_n \in C$, let $M = \max_{1 \leq i \leq n} \|f_i\|_\infty$ and let $\Phi : \mathbb{C}^n \rightarrow \mathbb{C}$ be continuous. Then by Stone-Weierstrass, Φ is the uniform limit on $K = \{|x| \leq M\}$ of a sequence (R_m) of polynomial functions in n variables. By hypothesis on C and \mathcal{V} , $R_m \circ (f_1, \dots, f_n)$ belongs to \mathcal{V} . Take $\Phi(x_1, \dots, x_n) = \prod_{i=1}^n \min(1, r(x_i - a_i)_+)$ for $r > 0$ and $a_1, \dots, a_n \in \mathbb{R}$. The (increasing) limit of $\Phi(x_1, \dots, x_n)$ as $r \rightarrow \infty$ is $\prod_{i=1}^n 1_{(x_i > a_i)}$. Hence, if $f_1, \dots, f_n \in C$, the function $\prod_{i=1}^n 1_{(f_i > a_i)}$ belongs to \mathcal{V} . Hence, $B = \bigcap_{i=1}^n (f_i > a_i) \in \mathcal{L}$ for all $a_1, \dots, a_n \in \mathbb{R}$. It follows that \mathcal{L} contains the π -system \mathcal{P} consisting of finite intersections of sets of the form $f^{-1}(I)$, where $f \in C$ and $I \subset \mathbb{R}$ is an open halfline. Thus, $\sigma(C) = \sigma(\mathcal{P}) \subseteq \mathcal{L}$. \square

Example 3.37. 1. Let (Ω, \mathbb{P}) be a probability space, let X, Y be random variables and suppose $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)g(X)]$ for all real $f, g \in C_c^\infty(\mathbb{R})$. Then $\mathbb{P}[X = Y] = 1$. Indeed, let \mathcal{V} be the set of all bounded real $\mathcal{B}(\mathbb{R}^2)$ -measurable $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that $\mathbb{E}[h(X, Y)] = \mathbb{E}[h(X, X)]$. Then \mathcal{V} satisfies (i)-(iii). Let C be the set of functions $(x, y) \mapsto f(x)g(y)$ with $f, g \in C_c^\infty(\mathbb{R})$. By hypothesis, $C \subset \mathcal{V}$, so by monotone class, \mathcal{V} contains all bounded $\sigma(C)$ -measurable functions. But $\mathcal{B}(\mathbb{R}^2)$ is generated by projections $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$, which are pointwise limits of $f_n(x)g_n(y)$ and $g_n(x)f_n(y)$, respectively, where f_n, g_n are smooth, $f_n(t) = t$ on $[-n, n]$, $g_n = 1$ on $[-n, n]$, $f_n = g_n = 0$ outside $[-n-1, n+1]$. Since f_n, g_n are $\sigma(C)$ -measurable, then π_1, π_2 are $\sigma(C)$ -measurable. Hence, $\sigma(C)$ contains $\mathcal{B}(\mathbb{R}^2)$. Thus, \mathcal{V} contains all Borel functions. Choosing $h = 1_\Delta$, where $\Delta = \{(x, x) : x \in \mathbb{R}\}$, we get $\mathbb{P}[X = Y] = 1$.

2. Let $\{\mu_\omega\}_{\omega \in \Omega}$ be a family of measures on \mathbb{R} . The following statements are equivalent:

(a) For all $t \in \mathbb{R}$, the map $\omega \mapsto \int \chi_{(-\infty, t]}(x) d\mu_\omega(x)$ is measurable.

(b) For all bounded Borel $f : \mathbb{R} \rightarrow \mathbb{C}$, the map $\omega \mapsto \int f(x) d\mu_\omega(x)$ is measurable.

Indeed, (b) trivially implies (a). For the converse, let \mathcal{V} be the set of all bounded measurable $f : \mathbb{R} \rightarrow \mathbb{R}$ with $\omega \mapsto \int f(x) d\mu_\omega(x)$ measurable. Then \mathcal{V} satisfies (i)-(iii). Let $C = \{\chi_{(-\infty, t]} : t \in \mathbb{R}\}$. By hypothesis, $C \subset \mathcal{V}$. By monotone class, \mathcal{V} contains all $\sigma(C)$ -measurable functions, i.e. all Borel measurable functions $f : \mathbb{R} \rightarrow \mathbb{R}$. By linearity, for all Borel functions $f : \mathbb{R} \rightarrow \mathbb{C}$, the map $\omega \mapsto \int f(x) d\mu_\omega(x)$ is measurable.

8. Recall that a π -system \mathcal{P} on a set Ω is a collection of subsets of Ω which is closed under finite intersections (e.g. $\Omega = \mathbb{R}$ and $\mathcal{P} = \{(a, b] : -\infty \leq a < b < \infty\}$). A λ -system on Ω is a collection \mathcal{L} of subsets of Ω such that (i) $\Omega \in \mathcal{L}$, (ii) if $A \in \mathcal{L}$, then $A^c \in \mathcal{L}$, (iii) if $A_n \in \mathcal{L}$, $A_n \cap A_m = \emptyset$ for $n \neq m$, then $\bigcup A_n \in \mathcal{L}$. The Dynkin π - λ theorem says that if \mathcal{L} is a λ -system containing a π -system \mathcal{P} , then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$. For a proof, see [3, Theorem 1.1].

3. Endow Ω with a σ -algebra \mathcal{F} , let $\{\mu_\omega\}_{\omega \in \Omega}$ be a family of measures on \mathbb{R} and suppose $\omega \mapsto \mu_\omega(B)$ is \mathcal{F} -measurable for each $B \in \mathcal{B}(\mathbb{R})$. If $W \subseteq \Omega \times \mathbb{R}$ is $\mathcal{F} \otimes \mathcal{B}(\mathbb{R})$ -measurable, define the ω -section $W_\omega = \{t \in \mathbb{R} : (\omega, t) \in W\}$ for each fixed $\omega \in \Omega$. Then the map $\omega \mapsto \mu_\omega(W_\omega)$ is \mathcal{F} -measurable. Indeed, if $W = A \times B$ for some $A \in \mathcal{F}$ and $B \in \mathcal{B}(\mathbb{R})$, then $\mu_\omega(W_\omega) = 1_A(\omega)\mu_\omega(B)$ and the measurability of this mapping is clear. The general case now follows from a monotone class argument.

3.7 Appendix

In this section we prove the results of Section 3.1.

Lemma 3.38. $|\mu|$ is a positive measure on \mathfrak{M} , and $|\mu|(X) < \infty$.

Proof. Clearly, $|\mu|(B) \geq 0$. If $B = \cup_1^\infty B_n$ with $\{B_n\}_1^\infty \subset \mathfrak{M}$ pairwise disjoint, then given $\varepsilon > 0$, choose a measurable partition $\{E_i\}_1^\infty$ of B such that $|\mu|(B) - \varepsilon < \sum_1^\infty |\mu(E_i)|$. Then $|\mu|(B) - \varepsilon \leq \sum_{i=1}^\infty |\sum_{n=1}^\infty \mu(E_i \cap B_n)| \leq \sum_{n=1}^\infty \sum_{i=1}^\infty |\mu(E_i \cap B_n)|$ (we can always change the order of summation when the terms are positive). But $\{E_i \cap B_n\}_{i=1}^\infty$ is a partition of B_n . Hence, $|\mu|(B) - \varepsilon \leq \sum_1^\infty |\mu|(B_n)$. Thus, $|\mu|(B) \leq \sum_1^\infty |\mu|(B_n)$. If $|\mu|(B) = \infty$, we get $\sum_1^\infty |\mu|(B_n) = \infty$, so $|\mu|(B) = \sum_1^\infty |\mu|(B_n)$. If $|\mu|(B) < \infty$, then $|\mu|(B_n) < \infty$ for each n . Given $\varepsilon > 0$, let $\{E_i^{(n)}\}_{i=1}^\infty$ be a partition of B_n such that $\sum_i |\mu(E_i^{(n)})| > |\mu|(B_n) - \frac{\varepsilon}{2^n}$. Then $\sum_{n=1}^N |\mu|(B_n) < \sum_{n=1}^N (\sum_i |\mu(E_i^{(n)})| + \frac{\varepsilon}{2^n}) \leq \varepsilon + \sum_{n=1}^N \sum_i |\mu(E_i^{(n)})| \leq \varepsilon + |\mu|(B)$. Letting $N \rightarrow \infty$ and $\varepsilon \rightarrow 0$ gives $\sum_1^\infty |\mu|(B_n) \leq |\mu|(B)$. Thus, $|\mu|(B) = \sum_1^\infty |\mu|(B_n)$.

To prove that $|\mu|$ is finite, we use a somehow reversed triangle inequality:

If $z_1, \dots, z_N \in \mathbb{C}$, then $\exists S \subseteq \{1, \dots, N\}$ such that $|\sum_{k \in S} z_k| \geq \frac{1}{\pi} \sum_{k=1}^N |z_k|$.

To see this, write $z_k = |z_k|e^{i\alpha_k}$. For any $S \subseteq \{1, \dots, N\}$ and $\theta \in [-\pi, \pi]$, we have $|\sum_S z_k| = |\sum_S e^{-i\theta} z_k| \geq \operatorname{Re} \sum_S e^{-i\theta} z_k = \sum_S |z_k| \cos(\alpha_k - \theta)$. Let $S_\theta = \{1 \leq k \leq N : \cos(\alpha_k - \theta) > 0\}$. Then $|\sum_{S(\theta)} z_k| \geq \sum_{S(\theta)} |z_k| \cos(\alpha_k - \theta) = \sum_{k=1}^N |z_k| \cos^+(\alpha_k - \theta)$. Now $\int_{-\pi}^\pi \cos^+(\alpha - \theta) d\theta = 2$ for any $\alpha \in [0, 2\pi[$. Indeed, $\int_{-\pi}^\pi = \int_{-\pi}^{\alpha-\pi} + \int_{\alpha-\pi}^\pi$, and $\int_{-\pi}^{\alpha-\pi} \cos^+(\alpha - \theta) = \int_{-\pi}^{\alpha-\pi} \cos^+(\alpha - \theta - 2\pi) = \int_{\pi}^{\pi+\alpha} \cos^+(\alpha - \theta)$. Hence, $\int_{-\pi}^{\alpha-\pi} + \int_{\alpha-\pi}^\pi = \int_{\pi}^{\pi+\alpha} + \int_{\alpha-\pi}^\pi = \int_{\alpha-\pi}^{\alpha+\pi}$. Thus, $\int_{-\pi}^\pi \cos^+(\alpha - \theta) = \int_{\alpha-\pi}^{\alpha+\pi} \cos^+(\alpha - \theta) = \int_{-\pi}^\pi \cos^+ \theta = \int_{-\pi/2}^{\pi/2} \cos \theta = 2$. We thus showed that $2 \sum_{k=1}^N |z_k| = \int_{-\pi}^\pi \sum_{k=1}^N |z_k| \cos^+(\alpha_k - \theta) \leq \int_{-\pi}^\pi |\sum_{S(\theta)} z_k| \leq 2\pi |\sum_{S(\theta_0)} z_k|$, where $|\sum_{S(\theta_0)} z_k| = \max_{-\pi \leq \theta \leq \pi} |\sum_{S(\theta)} z_k|$ is attained since the function is continuous. Choosing $S = S_{\theta_0}$ gives the claim.

Back to the lemma. Suppose first that some $E \in \mathfrak{M}$ has $|\mu|(E) = \infty$. Put $t = \pi(1 + |\mu(E)|)$. Since $|\mu|(E) > t$, there is a partition $\{E_i\}$ of E such that $\sum_1^N |\mu(E_i)| > t$ for some N . Using the preceding paragraph with $z_i = \mu(E_i)$, we find a set $A \subseteq E$ (a union of some E_i) such that $|\mu(A)| > \frac{t}{\pi} > 1$. Set $B = E \setminus A$. Then $|\mu(B)| = |\mu(E) - \mu(A)| \geq |\mu(A)| - |\mu(E)| > \frac{t}{\pi} - |\mu(E)| = 1$. We have thus split E into disjoint sets A and B such that $|\mu(A)| > 1$ and $|\mu(B)| > 1$. Since $|\mu|(E) = \infty$, either $|\mu|(A)$ or $|\mu|(B)$ is infinite (because we showed that $|\mu|$ is a measure).

Now suppose $|\mu|(X) = \infty$. Split X into A_1 and B_1 as above with $|\mu(A_1)| > 1$ and $|\mu(B_1)| = \infty$. Split B_1 into A_2 and B_2 with $|\mu(A_2)| > 1$ and $|\mu(B_2)| = \infty$. Continuing this way, we get a countably infinite collection $\{A_i\}$ with $|\mu(A_i)| > 1$ for each i . Since μ is countably additive, we have $\sum_i \mu(A_i) = \mu(\cup_i A_i) \in \mathbb{C}$. But the series cannot converge since $\mu(A_i)$ does not tend to 0 as $i \rightarrow \infty$. This contradiction shows that $|\mu|(X) < \infty$. \square

We will need some technical lemmas before we move on.

Lemma 3.39. Let (X, \mathfrak{M}, μ) be a measure space, $E \in \mathfrak{M}$ and $f : E \rightarrow [0, \infty]$ measurable.

- (i) (Markov inequality). For any $\alpha > 0$, $\mu\{x \in E : f(x) \geq \alpha\} \leq \frac{1}{\alpha} \int_E f \, d\mu$.
(ii) If $\int_E f \, d\mu = 0$, then $f = 0$ for μ -a.e. $x \in E$.
(iii) If $f > 1$ on E and $\mu(E) \neq 0$, then $\int_E f \, d\mu > \mu(E)$.

Proof. (i) Since $f \geq \alpha \chi_{\{f \geq \alpha\}}$, we have $\int_E f \, d\mu \geq \int_E \alpha \chi_{\{f \geq \alpha\}} \, d\mu = \alpha \mu\{x \in E : f(x) \geq \alpha\}$. For (ii), let $F = \{x \in E : f(x) > 0\}$, $F_n = \{x \in E : f(x) \geq \frac{1}{n}\}$ and note that $F = \bigcup_1^\infty F_n$. By (i), $\mu(F_n) \leq n \int_E f \, d\mu = 0$. Hence, $\mu(F) \leq \sum \mu(F_n) = 0$. Finally, if $f - 1 > 0$ on E and $\mu(E) \neq 0$, then by contraposition of (ii), we must have $\int_E (f - 1) \, d\mu \neq 0$. But $\int_E (f - 1) \, d\mu \geq 0$ since $f > 1$. Hence, $\int_E (f - 1) \, d\mu > 0$, i.e. $\int_E f \, d\mu > \mu(E)$. \square

Lemma 3.40. Let (X, \mathfrak{M}, μ) be a measure space with μ positive and finite. Let $f \in L^1(\mu)$, $\mathfrak{M}_+ = \{E \in \mathfrak{M} : \mu(E) > 0\}$ and define the average $A_E(f) = \frac{1}{\mu(E)} \int_E f \, d\mu$ for $E \in \mathfrak{M}_+$. If $S \subseteq \mathbb{C}$ is closed and $A_E(f) \in S$ for all $E \in \mathfrak{M}_+$, then $f(x) \in S$ for μ -a.e. $x \in X$.

Proof. Let Δ be a disc in S^c with center α and radius $r > 0$ and put $E = \{x : f(x) \in \Delta\}$. Since S^c is a countable union of discs, it suffices to prove that $\mu(E) = 0$. Suppose $\mu(E) > 0$. Then $|A_E(f) - \alpha| = \frac{1}{\mu(E)} |\int_E (f - \alpha) \, d\mu| \leq \frac{1}{\mu(E)} \int_E |f - \alpha| \, d\mu \leq r$. But this is impossible since $A_E(f) \in S$. Hence, $\mu(E) = 0$. \square

Lemma 3.41. Let (X, \mathfrak{M}, μ) be a measure space with μ positive and σ -finite. Then there exists $w \in L^1(\mu)$ such that $0 < w(x) \leq 1$ for every $x \in X$.

We will use this lemma to replace μ by the finite measure $\tilde{\mu} = w\mu$ given by $\tilde{\mu}(E) = \int_E w \, d\mu$, which has exactly the same sets of measure 0 as μ by Lemma 3.39.

Proof. Let $X = \bigcup_1^\infty E_n$ with $E_n \in \mathfrak{M}$ and $\mu(E_n)$ finite. Put $w_n(x) = 2^{-n}/(1 + \mu(E_n))$ if $x \in E_n$ and $w_n(x) = 0$ if $x \in E_n^c$. Let $w = \sum_1^\infty w_n$. If $x \in X$, say $x \in E_n$, then $0 < w_n(x) \leq w(x) \leq \sum 2^{-n} = 1$ and $\sum_1^N w_n \uparrow w$, so by monotone convergence, $\|w\|_1 = \int w = \lim_N \int \sum_1^N w_n = \lim_N \sum_1^N \int w_n = \lim_N \sum_1^N (2^{-n} \frac{\mu(E_n)}{1 + \mu(E_n)}) \leq 1$. \square

Theorem 3.42 (Lebesgue-Radon-Nikodym). Let (X, \mathfrak{M}) be a measurable space. Let μ be a σ -finite positive measure on \mathfrak{M} and ν a complex measure on \mathfrak{M} .

- (a) There is a unique pair of complex measures ν_a and ν_s on \mathfrak{M} such that $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$ and $\nu_s \perp \mu$.
(b) There is a unique $h \in L^1(\mu)$ such that $\nu_a(E) = \int_E h \, d\mu$ for every $E \in \mathfrak{M}$.

Proof. (von Neumann). First assume ν is positive and finite. Associate w to μ as in Lemma 3.41 and let $\lambda = w\mu + \nu$. Consider the functional $F(f) = \int f \, d\nu$ on $L^2(\lambda)$. By Cauchy-Schwarz, $|F(f)| \leq \int_X |f| \, d\nu \leq \int_X |f| \, d\lambda \leq \|f\| \cdot (\lambda(X))^{1/2}$, where $\|f\| = (\int |f|^2 \, d\lambda)^{1/2}$. Thus, $F \in L^2(\lambda)^*$. Applying Riesz's representation to the Hilbert space $L^2(\lambda)$, there exists $g \in L^2(\lambda)$ such that $F(f) = \int_X f g \, d\lambda$. Thus, $\int_X f \, d\nu = \int_X f g \, d\lambda$. In particular, for $f = 1_B$ we get

$$\nu(B) = \int_B g \, d\lambda \quad \text{for } B \in \mathfrak{M}. \quad (\dagger)$$

But $0 \leq \nu \leq \lambda$, so $0 \leq \frac{1}{\lambda(B)} \int_B g \, d\lambda = \frac{\nu(B)}{\lambda(B)} \leq 1$ for every $B \in \mathfrak{M}_+$. Hence, $0 \leq g \leq 1$ for λ -a.e. x by Lemma 3.40. Modifying g on the λ -null set does not affect (\dagger) , so we may assume $0 \leq g \leq 1$ on X . Let $S_1 = \{x : 0 \leq g(x) < 1\}$ and $S_2 = S_1^c = \{x : g(x) = 1\}$ and

define the measures $\nu_1(B) = \nu(B \cap S_1)$, $\nu_2(B) = \nu(B \cap S_2)$ for $B \in \mathfrak{M}$. Then using (\dagger) and $\lambda = w\mu + \nu$, we get

$$\int_B (1 - g) d\nu = \int_B gw d\mu \quad \text{for } B \in \mathfrak{M}. \quad (\ddagger)$$

For $B = S_2$, the LHS is zero, while the RHS is $\int_{S_2} w d\mu$. Since $w > 0$, it follows by Lemma 3.39(ii) that $\mu(S_2) = 0$. Since $\nu_2(S_2^c) = 0$ by definition, we get $\nu_2 \perp \mu$. Next, since $\int_B (1 - g) = \int_{B \cap S_1} (1 - g)$ and $1 - g > 0$ on S_1 , we have

$$\mu(B) \implies \int_B gw d\mu = 0 \implies \int_{B \cap S_1} (1 - g) d\nu = 0 \implies \nu(B \cap S_1) = 0 \implies \nu_1(B) = 0.$$

Hence, $\nu_1 \ll \mu$. Thus, we have a Lebesgue decomposition $\nu = \nu_a + \nu_s$ with $\nu_a = \nu_1$ and $\nu_s = \nu_2$. Multiplying (\ddagger) by $1, g, g^2, \dots, g^n$ and adding, we get

$$\int_B (1 - g^{n+1}) d\nu = \int_B (g + g^2 + \dots + g^{n+1})w d\mu \quad \text{for } B \in \mathfrak{M}.$$

Since $1 - g^{n+1} \uparrow 1$ on S_1 , denoting by h the increasing limit of $(g + \dots + g^{n+1})w$, we get $\nu_a(B) = \nu_1(B) = \nu(B \cap S_1) = \lim_n \int_{B \cap S_1} (1 - g^{n+1}) d\nu = \lim_n \int_B (1 - g^{n+1}) d\nu = \int_B h d\mu$, which completes the proof of existence for finite positive ν . We now prove uniqueness.

Suppose $\nu = \nu'_a + \nu'_s$ is another Lebesgue decomposition. Then $\nu'_a - \nu_a = \nu_s - \nu'_s$. But $\nu'_a - \nu_a \ll \mu$ and $\nu_s - \nu'_s \perp \mu$. Hence, $\nu'_a - \nu_a = \nu_s - \nu'_s = 0$.⁹

Next, if $\int_B h_1 d\mu = \int_B h_2 d\mu$ for all $B \in \mathfrak{M}$, then $\int_B (h_1 - h_2) d\mu = 0$ for all $B \in \mathfrak{M}$. In particular, $\int_{h_1 > h_2} (h_1 - h_2) d\mu = 0$ and $\int_{h_1 \leq h_2} (h_2 - h_1) d\mu = 0$, so $\int |h_1 - h_2| d\mu = 0$ and thus $h_1 = h_2$. This concludes uniqueness.

Finally, if ν is complex, then $\nu = \nu_1 + i\nu_2$ with ν_1, ν_2 real. Moreover, $\nu_j = \nu_j^+ - \nu_j^-$, where $\nu_j^+ = \frac{1}{2}(|\nu_j| + \nu_j)$ and $\nu_j^- = \frac{1}{2}(|\nu_j| - \nu_j)$ are positive. Hence, we can apply the preceding case to each ν_j^\pm . \square

Theorem 3.43. *Let (X, \mathfrak{M}, μ) be a measure space with μ σ -finite. Let $1 \leq p < \infty$ and q its conjugate exponent (i.e. $\frac{1}{p} + \frac{1}{q} = 1$). Define $\Phi : L^q(\mu) \rightarrow L^p(\mu)^*$ by $g \mapsto \lambda_g$, where $\lambda_g(f) = \int fg$. Then Φ is an isometric isomorphism.*

Proof. We first assume $\mu(X) < \infty$. Clearly, Φ is linear. To see surjectivity, given $\lambda \in L^p(\mu)^*$, let $\nu(E) = \lambda(\chi_E)$ for $E \subseteq X$ measurable. If $E = \cup_1^\infty E_i$ with $\{E_i\}$ disjoint, then $\chi_E = \sum_1^\infty \chi_{E_i}$ in L^p , since $\|\chi_E - \sum_1^n \chi_{E_i}\|_p = \|\sum_{n+1}^\infty \chi_{E_i}\|_p = \mu(\cup_{n+1}^\infty E_i)^{1/p} = (\sum_{n+1}^\infty \mu(E_i))^{1/p} \rightarrow 0$ (we used that $p < \infty$). Since λ is linear and continuous, we get $\nu(E) = \sum_1^\infty \lambda(\chi_{E_i}) = \sum_1^\infty \nu(E_i)$, so ν is a complex measure. Also, if $\mu(E) = 0$, then $\|\chi_E\|_p = 0$ and thus $\nu(E) = 0$. Hence, $\nu \ll \mu$, so by Radon-Nikodym, there exists $g \in L^1(\mu)$ such that $\lambda(\chi_E) = \nu(E) = \int_E g d\mu$ for all measurable E . By linearity, we get

$$\lambda(f) = \int fg d\mu \quad (\star)$$

for all simple f , and thus for all $f \in L^\infty(\mu)$, since every $f \in L^\infty(\mu)$ is a uniform limit of simple functions (f_i) , and $\|f_i - f\|_p \leq \|f_i - f\|_\infty \mu(X)^{1/p} \rightarrow 0$, so that $\lambda(f_i) \rightarrow \lambda(f)$.

9. We have used the following facts: (i) if $\nu_1 \ll \mu$ and $\nu_2 \ll \mu$, then $\nu_1 + \nu_2 \ll \mu$, (ii) if $\nu_1 \perp \mu$ and $\nu_2 \perp \mu$, then $\nu_1 + \nu_2 \perp \mu$, (iii) if $\nu \ll \mu$ and $\nu \perp \mu$, then $\nu = 0$. (i) is obvious. For (ii), if there are $A_1 \cap B_1 = \emptyset$ such that ν_1 is concentrated on A_1 , μ on B_1 , and $A_2 \cap B_2 = \emptyset$ such that ν_2 is concentrated on A_2 , μ on B_2 , then $\nu_1 + \nu_2$ is concentrated on $A = A_1 \cup A_2$, μ on $B = B_1 \cap B_2$, and $A \cap B = \emptyset$. Thus, $\nu_1 + \nu_2 \perp \mu$. Finally, for (iii), if $\nu \perp \mu$, then ν is concentrated on a set A such that $\mu(A) = 0$. But $\nu \ll \mu$, so $\nu(E) = 0$ for every $E \subseteq A$. Hence, ν is also concentrated on A^c . Thus, $\nu = 0$.

We now show that $g \in L^q(\mu)$. If $p = 1$, then by (\star) we have $|\int_E g \, d\mu| \leq \|\lambda\| \|\chi_E\|_1 = \|\lambda\| \mu(E)$ for all $E \in \mathfrak{M}$. Hence, $|g(x)| \leq \|\lambda\|$ a.e. by Lemma 3.40, so $\|g\|_\infty \leq \|\lambda\|$.

If $1 < p < \infty$, let $E_n = \{x : |g(x)| \leq n\}$ and write $|g| = \alpha g$, where α is measurable and $|\alpha| = 1$.¹⁰ Put $f = \chi_{E_n} |g|^{q-1} \alpha$. Then $f \in L^\infty(\mu)$, and $|f|^p = |g|^q = fg$ on E_n , since $pq - p = q$. So by (\star) , $\int_{E_n} |g|^q \, d\mu = \int_X fg \, d\mu = \lambda(f) \leq \|\lambda\| \|f\|_p = \|\lambda\| (\int_{E_n} |g|^q)^{1/p}$. Thus, $(\int_X \chi_{E_n} |g|^q \, d\mu)^{1/q} \leq \|\lambda\|$. Applying monotone convergence, we get $\|g\|_q \leq \|\lambda\|$.

We thus showed in both cases that $g \in L^q(\mu)$, with $\|g\|_q \leq \|\lambda\|$. By (\star) , λ and λ_g coincide on $L^\infty(\mu)$, which is dense in $L^p(\mu)$, hence $\lambda = \lambda_g$. But by Hölder, $|\lambda_g(f)| = |\int fg \, d\mu| \leq \|f\|_p \|g\|_q$, so $\|\lambda\| = \|\lambda_g\| \leq \|g\|_q$. Thus, Φ is an isometry. In particular, Φ is injective, and we just proved it is surjective. This completes the proof if $\mu(X) < \infty$.

If μ is σ -finite, choose $w \in L^1(\mu)$ as in Lemma 3.41 and let $\tilde{\mu} = w\mu$. Define $\Theta : L^p(\tilde{\mu}) \rightarrow L^p(\mu)$ by $F \mapsto w^{1/p} F$. Then Θ is a bijective isometry since $0 < w \leq 1$. Hence, if $\lambda \in L^p(\mu)^*$, then $\Psi := \lambda \circ \Theta \in L^p(\tilde{\mu})^*$ and $\|\Psi\| = \|\lambda\|$. By the finite case, $\exists G \in L^q(\tilde{\mu})$ such that $\Psi(F) = \int FG \, d\tilde{\mu}$ and $\|\Psi\| = \|G\|_q$. Put $g = w^{1/q} G$ if $p > 1$ and $g = G$ if $p = 1$. Then $\int |g|^q \, d\mu = \int |G|^q \, d\tilde{\mu} = \|\Psi\|^q = \|\lambda\|^q$ if $p > 1$, while $\|g\|_\infty = \|G\|_\infty = \|\Psi\| = \|\lambda\|$ if $p = 1$. In any case, $\|g\|_q = \|\lambda\|$. Moreover, since $G\tilde{\mu} = w^{1/p} g\mu$, we finally get $\lambda(f) = \Psi(w^{-1/p} f) = \int w^{-1/p} f G \, d\tilde{\mu} = \int fg \, d\mu$ for every $f \in L^p(\mu)$. \square

Theorem 3.44 (Riesz-Markov). *Let X be a locally compact Hausdorff space and define $\Phi : \mathcal{M}(X) \rightarrow C_0(X)^*$ by $\mu \mapsto \lambda_\mu$, where $\lambda_\mu(f) = \int f \, d\mu$. Then Φ is an isometric isomorphism.*

If X is compact, the map $\mu \mapsto \lambda_\mu$ gives an isometric isomorphism of $\mathcal{M}_{+,1}(X)$ with $C(X)_{+,1}^ = \{\lambda \in C(X)^* : \|\lambda\| = 1 \text{ and } \lambda(f) \geq 0 \text{ whenever } f \geq 0\}$.*

Proof. The proof is really long, so we will only outline the main steps. We encourage the serious student to read the whole proof in [31, Theorem 6.19]. As in Theorem 3.6, the difficulty is to prove that Φ is surjective. Before we begin, let us mention that the proof is much easier in the (important) special case of $C[0,1]$. In this case, one simply argues as follows: given $\lambda \in C[0,1]^*$, use Hahn-Banach to find an extension L of λ to the space of all bounded functions on $[0,1]$. The advantage of L over λ is that it is well-defined on characteristic functions. So imitating the proof of Theorem 3.6, we may put $w(x) = L(\chi_{[0,x]})$. It can be shown that w has a bounded variation. Now given $f \in C[0,1]$, let $f_n = \sum_{i=1}^n f(\frac{i}{n}) \chi_{[\frac{i-1}{n}, \frac{i}{n}]}$. Then f_n converges uniformly to f , so $L(f_n)$ converges to $L(f)$. But $L(\chi_{[\frac{i-1}{n}, \frac{i}{n}]}) = L(\chi_{[0, \frac{i}{n}]}) - L(\chi_{[0, \frac{i-1}{n}]}) = w(\frac{i}{n}) - w(\frac{i-1}{n})$, so $L(f_n) = \int f_n \, dw$ by definition of the Riemann-Stieltjes integral. Hence, $L(f) = \int f \, dw$. For details, see [28, page 354] or [22, Section 4.4]. We now proceed to the general case:

1. Let $\lambda \in C_0(X)^*$. We first assume λ is *positive*, i.e. $\lambda(f) \geq 0$ whenever $f \geq 0$. Given V open in X , define

$$\mu(V) = \sup\{\lambda(\phi) : \phi \in C_c(X), 0 \leq \phi \leq 1 \text{ and } \text{supp } \phi \subseteq V\}.$$

Clearly, $V_1 \subseteq V_2$ implies $\mu(V_1) \leq \mu(V_2)$, so

$$\mu(E) = \inf_{V \supseteq E, V \text{ open}} \mu(V) \quad (*)$$

for any open set E . We now define $\mu(E)$ for any set $E \subseteq X$ by $(*)$. Let \mathfrak{M}_F be the class of all $E \subseteq X$ such that $\mu(E) < \infty$ and $\mu(E) = \sup_{K \subseteq E, K \text{ compact}} \mu(K)$. And let

10. Let $\alpha(x) = \frac{g(x) + \chi_E(x)}{|g(x) + \chi_E(x)|}$, where $E = \{x : g(x) = 0\}$. Then $\alpha(x) = 1$ on E , $\alpha(x) = g(x)/|g(x)|$ on E^c , and α is measurable since E is measurable and $\beta(z) = z/|z|$ is continuous.

\mathfrak{M} be the class of all $E \subseteq X$ such that $E \cap K \in \mathfrak{M}_F$ for every compact K . Then one proves that \mathfrak{M} is a σ -algebra, and μ is a measure on \mathfrak{M} (much effort here).

2. Show that $\lambda(f) = \int_X f \, d\mu$ for every $f \in C_c(X)$. The idea is as follows: let $K = \text{supp } f$ and divide K into pieces E_i on which f is almost constant. Now use a partition of unity (h_i) subordinate to $\{E_i\}$ to write $f = \sum h_i f$. This writes f as something close to a step function, so it is more or less sufficient to describe $\mu(E_i)$ and $\mu(K)$ in terms of $\lambda(h_i)$, and one does this by working from the definition of $\mu(V)$ with V open.
3. Now generalize the previous construction to all $\lambda \in C_0(X)^*$, not just the positive ones. This is done by associating a positive functional to λ , namely $\tilde{\lambda}(f) = \sup\{|\lambda(g)| : g \in C_0(X) \text{ and } |g| \leq f\}$. One can prove that $\tilde{\lambda}$ is indeed linear, and $\|\tilde{\lambda}\| = \|\lambda\|$.

These are the major steps, one then ties everything together and concludes with some additional arguments. □

Chapter 4

Spectral theory

This chapter is taken from [44], [28] and [38], with added details. The letters X, Y, Z, \mathcal{H} denote vector spaces over \mathbb{C} which are never $\{0\}$.

4.1 Compact and Hilbert-Schmidt operators

Definition 4.1. Let X, Y be Banach spaces. We say that $T : X \rightarrow Y$ is *compact* if $\overline{T(B)}$ is compact in Y whenever $B \subset X$ is bounded.

Lemma 4.2. *The following assertions are equivalent:*

- (i) T is compact.
- (ii) $\overline{T(X_1)}$ is compact. Here X_1 is the unit ball of X .
- (iii) Any bounded sequence (x_n) in X has a subsequence (x_{n_k}) such that (Tx_{n_k}) converges.

Proof. Exercise 1. □

Example 4.3. If $T : X \rightarrow Y$ has a finite rank, i.e. $\dim \text{Ran}(T) < \infty$, then T is compact. Indeed, if $B \subset X$ is bounded, then $\overline{T(B)}$ is a closed bounded subset in a finite dimensional normed space, hence compact by Heine-Borel.

The set of compact operators from X to Y is denoted by $K(X, Y)$. The set of bounded operators is denoted by $B(X, Y)$. We also denote $K(X) = K(X, X)$ and $B(X) = B(X, X)$.

Lemma 4.4. *Let X, Y be Banach spaces.*

- (a) $K(X, Y)$ is a closed subspace of $B(X, Y)$. In other words, $\alpha T + \beta S$ is compact whenever T, S are compact, $\alpha, \beta \in \mathbb{K}$, and if $\{T_n\}$ is a sequence of compact operators converging in norm to T , then T is compact.
- (b) If Z is a Banach space, $T \in K(X, Y)$, $S_1 \in B(Y, Z)$ and $S_2 \in B(Z, X)$, then $TS_2 \in K(Z, Y)$ and $S_1T \in K(X, Z)$. In particular, $K(X)$ is an ideal in $B(X)$.
- (c) If $T \in K(X, Y)$, then the adjoint $T^* \in K(Y^*, X^*)$.
If X, Y are Hilbert spaces, then the Hilbert adjoint $T^* \in K(Y, X)$.

Proof. (a) Let (x_n) be bounded in X . There is a subsequence $(x_{\varphi(n)})$ such that $(Tx_{\varphi(n)})$ converges. But $(x_{\varphi(n)})$ is bounded, so it has a subsequence $(x_{\varphi \circ \psi(n)})$ such that $(Sx_{\varphi \circ \psi(n)})$ converges. Hence, $((\alpha T + \beta S)x_{\varphi \circ \psi(n)})$ converges, so $\alpha T + \beta S$ is compact.

Suppose $T_n \in K(X, Y)$ and $\|T_n - T\| \rightarrow 0$. To see that T is compact, it suffices to see that $T(X_1)$ is totally bounded.¹ Let $\varepsilon > 0$. Given $x, y \in X_1$, we have

$$\|Tx - Ty\| \leq \|Tx - T_nx\| + \|T_nx - T_ny\| + \|T_ny - Ty\| \leq 2\|T - T_n\| + \|T_nx - T_ny\|.$$

Fix N large enough such that $\|T_N - T\| \leq \varepsilon/4$. Choose a finite $\varepsilon/2$ -net for $T_N(X_1)$, say $\{T_N(x_1), \dots, T_N(x_r)\}$. Then $\{T(x_1), \dots, T(x_r)\}$ is an ε -net for $T(X_1)$.²

- (b) Let (z_n) be bounded in Z . Then (S_2z_n) is bounded in X (since S_2 is bounded), so there is a subsequence $(S_2z_{\varphi(n)})$ such that $(TS_2z_{\varphi(n)})$ converges, so TS_2 is compact. Next, if (x_n) is bounded in X , there is a subsequence $(x_{\varphi(n)})$ such that $(Tx_{\varphi(n)})$ converges, so $(S_1Tx_{\varphi(n)})$ converges (since S_1 is continuous), so S_1T is compact.
- (c) We only prove (c) for Hilbert spaces; the reader can find the Banach case in [22, Theorem 8.2.5]. So suppose T is compact and let (y_n) be bounded in Y , say $\|y_n\| \leq M$. Since TT^* is compact by (b), then (y_n) has a subsequence $(y_{\varphi(n)})$ such that $(TT^*y_{\varphi(n)})$ converges. Hence,

$$\begin{aligned} \|T^*y_{\varphi(n)} - T^*y_{\varphi(m)}\|^2 &= \langle TT^*(y_{\varphi(n)} - y_{\varphi(m)}), y_{\varphi(n)} - y_{\varphi(m)} \rangle \\ &\leq \|TT^*(y_{\varphi(n)} - y_{\varphi(m)})\| \cdot \|y_{\varphi(n)} - y_{\varphi(m)}\| \\ &\leq 2M\|TT^*y_{\varphi(n)} - TT^*y_{\varphi(m)}\| \rightarrow 0 \end{aligned}$$

as $n, m \rightarrow \infty$. Hence, $(T^*y_{\varphi(n)})$ is Cauchy, hence convergent, so T^* is compact. \square

Example 4.5. Suppose \mathcal{H} is a Hilbert space with orthonormal basis $\{e_i\}_{i=1}^\infty$ and let $T \in B(\mathcal{H})$ be given by the diagonal matrix $T_{i,j} = \lambda_i \delta_{i,j}$, i.e. $Te_i = \lambda_i e_i$. Then T is compact iff $\lambda_i \rightarrow 0$ as $i \rightarrow \infty$.

To see this, first suppose $\lambda_i \rightarrow 0$. By definition, if $x = \sum \alpha_i e_i \in \mathcal{H}$, then $Tx = \sum \alpha_i \lambda_i e_i$. For each n , let $T_n e_i = Te_i$ if $i \leq n$ and $T_n e_i = 0$ if $i > n$. Then T_n has a finite rank ($\text{Ran } T_n \subset \text{sp}\{e_1, \dots, e_n\}$), so T_n is compact. Moreover, $\|(T - T_n)x\|^2 = \|\sum_{i>n} \alpha_i \lambda_i e_i\|^2 = \sum_{i>n} |\alpha_i \lambda_i|^2 \leq \sup_{i>n} |\lambda_i|^2 \sum_{i=n+1}^\infty |\alpha_i \lambda_i|^2 \leq \sup_{i>n} |\lambda_i|^2 \|x\|^2$. Hence, $\|T_n - T\| \leq \sup_{i>n} |\lambda_i| \rightarrow 0$ as $i \rightarrow \infty$, so T is compact by Lemma 4.4(a).

Conversely, suppose $\lambda_i \not\rightarrow 0$. Then there exists ε_0 such that $J = \{i : |\lambda_i| > \varepsilon_0\}$ is infinite.³ For each $i, j \in J$, $i \neq j$, we have $\|Te_i - Te_j\|^2 = |\lambda_i|^2 + |\lambda_j|^2 \geq 2\varepsilon_0^2$. Hence, (Te_i) has no convergent subsequence, so T is not compact.

Compact operators enjoy many interesting properties:

- Lemma 4.6.** (1) *If X, Y are Banach spaces and $T \in K(X, Y)$, then T transforms weak convergence into strong convergence. That is, if $\ell(x_n - x) \rightarrow 0$ for every $\ell \in X^*$, then $\|Tx_n - Tx\| \rightarrow 0$.*
- (2) *If \mathcal{H} is a separable Hilbert space and $T \in K(\mathcal{H})$, then T transforms strong convergence into uniform convergence. That is, if $\|S_n x - Sx\| \rightarrow 0$ for each $x \in \mathcal{H}$, then $\|S_n T - ST\| \rightarrow 0$ and $\|TS_n - TS\| \rightarrow 0$.*

1. Recall that if K is a metric space, then $E \subseteq K$ is an ε -net for K if $\forall p \in K \exists q \in E$ such that $d(p, q) < \varepsilon$. K is *totally bounded* if for each $\varepsilon > 0$ there exists a finite ε -net for K . A classic result is that K is compact iff it is totally bounded and complete. Since Y is Banach, $\overline{T(X_1)}$ is always complete.

2. Longer proof: suppose (x_n) is bounded in X . There is a subsequence $(x_{1,n})$ such that $(T_1 x_{1,n})$ converges. Also, $(x_{1,n})$ has a subsequence $(x_{2,n})$ such that $(T_2 x_{2,n})$ converges, and so on. Let $y_n = x_{n,n}$. Then $(T_n y_n)_m$ is Cauchy. Using an $\varepsilon/3$ argument, it follows that $(T y_m)$ is Cauchy, hence convergent.

3. Details: if for all $\varepsilon > 0$ the set $J = \{i : |\lambda_i| > \varepsilon\}$ is finite, then $\forall \varepsilon > 0 \exists N \in \mathbb{N}$ (namely $N = \max J$) such that $|\lambda_i| \leq \varepsilon$ if $i > N$, so $\lambda_i \rightarrow 0$.

(3) If \mathcal{H} is a separable Hilbert space, then every $T \in K(\mathcal{H})$ is the norm limit of a sequence of operators of finite rank.

We will not need this lemma, but we prove (1) and (3) for the curious reader. Let us mention that (3) is not true for general Banach spaces. We say that a Banach space has the *approximation property* if (3) is true. The first separable Banach space without this property was constructed by Enflo in 1972. He was awarded a live goose for his achievement, which was promised by Mazur in 1936. Enflo was 28 years old.

Proof. For (1), suppose x_n converges weakly to x . By the uniform boundedness principle, the sequence $(\|x_n\|)$ is bounded (see [22]). Let $y_n = Tx_n$ and $\ell \in Y^*$. Then $|\ell^*(y_n) - \ell^*(y)| = |(T^*\ell)(x_n - x)| \rightarrow 0$ since $T^*\ell \in X^*$. Hence, y_n converges weakly to y . Suppose y_n does not converge to y in norm. Then there exists ε_0 and a subsequence $(y_{\varphi(n)})$ of (y_n) such that $\|y_{\varphi(n)} - y\| \geq \varepsilon_0$. But $(x_{\varphi(n)})$ is bounded and T is compact, so $y_{\varphi(n)} = Tx_{\varphi(n)}$ has a subsequence $(y_{\varphi \circ \psi(n)})$ that converges (in norm) to some \tilde{y} . Now $\tilde{y} \neq y$ since $\|\tilde{y} - y\| = \lim \|y_{\varphi \circ \psi(n)} - y\| \geq \varepsilon_0$. But this is impossible since $(y_{\varphi \circ \psi(n)})$ converges weakly to \tilde{y} and (y_n) converges weakly to y . This contradiction shows that $\|y_n - y\| \rightarrow 0$.

We omit the proof of (2). *Sketch:* we may assume $S = 0$. First assume T has a finite rank, then generalize using (3).

For (3), let $\{e_i\}$ be an orthonormal basis of \mathcal{H} and let $T_n x = \sum_1^n \langle x, e_j \rangle T e_j$. Then $Tx - T_n x = T y_n$, where $y_n = \sum_{n+1}^\infty \langle x, e_j \rangle e_j \in \{e_1, \dots, e_n\}^\perp$ and $\|y_n\| \leq \|x\|$. Hence, $\|T - T_n\| = \sup_{\|x\| \leq 1} \|Tx - T_n x\| \leq \sup_{y \in \{e_1, \dots, e_n\}^\perp, \|y\| \leq 1} \|Ty\| =: \alpha_n$. Clearly, α_n is monotone decreasing, so it converges to a limit $\alpha \geq 0$. Choose $y_n \in \{e_1, \dots, e_n\}^\perp$ with $\|y_n\| \leq 1$ such that $\|Ty_n\| \geq \frac{3}{4}\alpha_n$. Then $\|Ty_n\| \geq \frac{1}{2}\alpha$ if n is large enough. Since y_n converges weakly to 0, then $\|Ty_n\| \rightarrow 0$ by (1). Thus, $\alpha = 0$. Hence, $\|T - T_n\| \rightarrow 0$. \square

Definition 4.7. Let \mathcal{H} be a Hilbert space with orthonormal basis $\{e_j\}$ and let $T \in B(\mathcal{H})$. We say that T is a *Hilbert-Schmidt operator* if $\sum_j \|Te_j\|^2 < \infty$. Equivalently, $\sum_{i,j} |T_{ij}|^2 = \sum_{i,j} |\langle Te_j, e_i \rangle|^2 < \infty$.

This seems to depend on the basis, but it does not:

Lemma 4.8. $\sum_{i,j} |\langle Te_j, e_i \rangle|^2 = \sum_{i,j} |\langle T f_j, f_i \rangle|^2$ for any orthonormal bases $\{e_i\}$ and $\{f_j\}$.

Proof. We have

$$\sum_{i,j} |\langle Te_j, e_i \rangle|^2 = \sum_j \|Te_j\|^2 = \sum_{i,j} |\langle Te_j, f_i \rangle|^2 = \sum_{i,j} |\langle e_j, T^* f_i \rangle|^2 = \sum_j \|T^* f_i\|^2.$$

The RHS is independent of $\{e_i\}$, and hence so is the LHS. \square

Definition 4.9. If T is Hilbert-Schmidt, we define the *Hilbert-Schmidt norm* of T by $\|T\|_2 = (\sum_j \|Te_j\|^2)^{1/2}$, where $\{e_j\}$ is any orthonormal basis of \mathcal{H} .

Lemma 4.10. Let T be a Hilbert-Schmidt operator. Then

- (i) T^* is Hilbert-Schmidt with $\|T^*\|_2 = \|T\|_2$,
- (ii) $\|T\| \leq \|T\|_2$,
- (iii) T is compact.

Proof. For (i), see the proof of Lemma 4.8. For (ii), let $\{e_j\}$ be an orthonormal basis of \mathcal{H} and $x = \sum \alpha_j e_j \in \mathcal{H}$. Then by Cauchy-Schwarz,

$$\begin{aligned} \|Tx\|^2 &= \sum_i |\langle Tx, e_i \rangle|^2 = \sum_i \left| \sum_j \alpha_j \langle Te_j, e_i \rangle \right|^2 \\ &\leq \sum_i \left(\sum_j |\alpha_j|^2 \right) \left(\sum_j |\langle Te_j, e_i \rangle|^2 \right) = \|x\|^2 \|T\|_2^2. \end{aligned}$$

For (iii), for each n , define $T_n \in B(\mathcal{H})$ by $T_n e_i = T e_i$ if $i \leq n$ and $T_n e_i = 0$ if $i > n$. Then T_n has finite rank. Furthermore, $T - T_n$ is Hilbert-Schmidt, with $\|T - T_n\|_2^2 = \sum_{j>n} \|T e_j\|^2$. Since $\sum_j \|T e_j\|^2 < \infty$, we have $\|T - T_n\|_2 \rightarrow 0$, so by (ii), $\|T - T_n\| \rightarrow 0$. Hence, T is compact by Lemma 4.4(a). \square

Example 4.11. Let (X, μ) be a σ -finite measure space and let $K \in L^2(X \times X)$. Define the integral operator $T_K : L^2(\mu) \rightarrow L^2(\mu)$ by $(T_K f)(x) = \int_X K(x, y) f(y) d\mu(y)$. Then T_K is Hilbert-Schmidt, and $\|T_K\|_2 = \|K\|_{L^2(X \times X)}$.

To see this, fix x and let $K_x(y) = K(x, y)$. Then $K_x \in L^2(\mu)$ for μ -a.e. x by Fubini's theorem. Let $\{e_i\}$ be an orthonormal basis of $L^2(\mu)$. Then by monotone convergence,

$$\begin{aligned} \sum \|T_K e_i\|^2 &= \sum \int |(T_K e_i)(x)|^2 = \sum \int |\langle K_x, \bar{e}_i \rangle|^2 \\ &= \int \sum |\langle \bar{K}_x, e_i \rangle|^2 = \int \|\bar{K}_x\|_{L^2(X)}^2 = \|K\|_{L^2(X \times X)}^2. \end{aligned}$$

The space of Hilbert-Schmidt operators \mathcal{S}_2 is a Hilbert space when endowed with the inner product $(T, S) = \sum_{i,j} T_{ij} \bar{S}_{i,j}$.

We showed above that $\mathcal{S}_2 \subset \mathcal{S}_\infty \subset B(\mathcal{H})$, where \mathcal{S}_∞ are the compact operators, and we showed that \mathcal{S}_∞ is a closed ideal. There are analogous classes of operators \mathcal{S}_p for every $1 \leq p < \infty$. The most important one is $p = 1$, the set of *trace class operators*. They are defined as follows: call an operator T *positive* if $\langle Tx, x \rangle \geq 0$ for all $x \in \mathcal{H}$. If T is positive, define the *trace of T* by $\text{tr}(T) = \sum_1^\infty \langle T e_i, e_i \rangle$. If T is a matrix, the trace is just the sum of the diagonal elements, and also the sum of the eigenvalues. We will give a meaning later on to the operator $|T| = \sqrt{T^* T}$ (it is the unique operator B satisfying $B \geq 0$ and $B^2 = T^* T$). We say that a bounded operator A is trace class if $\text{tr}|A| < \infty$. It can be shown that any trace class operator A takes the form $A = BC$, where $B, C \in \mathcal{S}_2$. The classes \mathcal{S}_1 and \mathcal{S}_2 also form ideals.

4.2 Spectral theorem for compact operators

It is well-known that a Hermitian matrix on \mathbb{C}^n (i.e. $a_{i,j} = \bar{a}_{j,i}$) has an orthonormal basis of eigenvectors. In this section, we generalize this result to compact self-adjoint (more generally, normal) operators on a separable Hilbert space.

Definition 4.12. Let \mathcal{H} be a Hilbert space and $T \in B(\mathcal{H})$. Given $\lambda \in \mathbb{C}$, we define $\mathcal{H}_\lambda = \{x \in \mathcal{H} : Tx = \lambda x\}$. Then \mathcal{H}_λ is a closed subspace of \mathcal{H} . λ is called an *eigenvalue* if $\mathcal{H}_\lambda \neq \{0\}$, and $x \in \mathcal{H}_\lambda$ is called an *eigenvector*.

Example 4.13. In contrast to finite dimensions, a self-adjoint operator need not have any eigenvalue. For example, define $T : L^2[a, b] \rightarrow L^2[a, b]$ by $(Tf)(x) = xf(x)$. Then T is self-adjoint, T is bounded since $\|Tf\|_{L^2[a,b]} \leq \|x\|_{L^\infty[a,b]} \|f\|_{L^2[a,b]}$, but T has no eigenvalue, since $Tf = \lambda f$ implies $(x - \lambda)f(x) = 0$ a.e. and hence $f = 0$ in $L^2[a, b]$.

Lemma 4.14. *Let \mathcal{H} be a Hilbert space and $T \in B(\mathcal{H})$ be self-adjoint (i.e. $T = T^*$).*

- (i) *If W is T -invariant (i.e. $T(W) \subseteq W$), then W^\perp is also T -invariant.*
- (ii) *For every $x \in \mathcal{H}$, $\langle Tx, x \rangle \in \mathbb{R}$. In particular, all eigenvalues are real.*
- (iii) $\|T\| = \sup_{\|x\| \leq 1} |\langle Tx, x \rangle|$.
- (iv) *If $\lambda \neq \beta$, then $\mathcal{H}_\lambda \perp \mathcal{H}_\beta$.*

Proof. (i) If $y \in W^\perp$ and $x \in W$, $\langle x, Ty \rangle = \langle Tx, y \rangle = 0$ since $Tx \in W$. Hence, $Ty \in W^\perp$.

(ii) Note that $\overline{\langle Tx, x \rangle} = \langle x, Tx \rangle = \langle T^*x, x \rangle = \langle Tx, x \rangle$, so $\langle Tx, x \rangle \in \mathbb{R}$.

If $Tx = \lambda x$, $x \neq 0$, then $\lambda \|x\|^2 = \langle Tx, x \rangle \in \mathbb{R}$, so $\lambda \in \mathbb{R}$.

(iii) Let $\alpha = \sup_{\|x\| \leq 1} |\langle Tx, x \rangle|$. By Cauchy-Schwarz, $\alpha \leq \|T\|$. To prove that $\|T\| \leq \alpha$, we first show that $|\langle Tx, y \rangle| \leq \alpha \|x\| \|y\|$. Since this inequality is unchanged if we multiply y by a complex number of modulus 1, we may assume $\langle Tx, y \rangle \in \mathbb{R}$. Then

$$4\langle Tx, y \rangle = \langle T(x+y), x+y \rangle - \langle T(x-y), x-y \rangle,$$

so using the definition of α and the parallelogram identity,

$$|\langle Tx, y \rangle| \leq \frac{\alpha}{4} (\|x+y\|^2 + \|x-y\|^2) = \frac{\alpha}{2} (\|x\|^2 + \|y\|^2).$$

Apply this inequality to $\sqrt{a}x$ and $\frac{1}{\sqrt{a}}y$ for any $a > 0$. We get

$$|\langle Tx, y \rangle| \leq \frac{\alpha}{2} (a\|x\|^2 + \frac{1}{a}\|y\|^2).$$

If $x = 0$, then trivially $|\langle Tx, y \rangle| \leq \alpha \|x\| \|y\|$. Otherwise, choose $a = \|y\|/\|x\|$ to get $|\langle Tx, y \rangle| \leq \alpha \|x\| \|y\|$. To finally complete the proof of (iii), choose x such that $Tx \neq 0$, and apply this inequality to $y = Tx/\|Tx\|$ (the case $T = 0$ is trivial). Then we get $\|Tx\| = |\langle Tx, y \rangle| \leq \alpha \|x\| \cdot 1$, i.e. $\|T\| \leq \alpha$.

(iv) If $Tx = \lambda x$ and $Ty = \beta y$, then $\langle Tx, y \rangle = \lambda \langle x, y \rangle$, and $\langle Tx, y \rangle = \langle x, Ty \rangle = \beta \langle x, y \rangle$ ($\beta \in \mathbb{R}$ by (ii)). Hence, $(\lambda - \beta)\langle x, y \rangle = 0$. If $\lambda \neq \beta$, we must have $\langle x, y \rangle = 0$. \square

Theorem 4.15 (Spectral theorem I). *Let T be a compact self-adjoint operator on a separable Hilbert space \mathcal{H} . Then \mathcal{H} has an orthonormal basis $\{e_n\}$ consisting of eigenvectors of T , i.e. $Te_n = \lambda_n e_n$. If \mathcal{H} is infinite-dimensional, then $\lambda_n \rightarrow 0$.*

This is known as the *Hilbert-Schmidt theorem*. Note that since $\lambda_n \rightarrow 0$, then each nonzero eigenvalue has a finite multiplicity. So here, the *spectrum* we will soon define is a discrete set, with no accumulation point except perhaps $\lambda = 0$.

Proof. We first claim that we can find at least one non-zero eigenvector. If $T = 0$, any vector is an eigenvector, so we may assume $T \neq 0$. By Lemma 4.14(iii), we can choose $x_n \in \mathcal{H}$ with $\|x_n\| \leq 1$ such that $|\langle Tx_n, x_n \rangle| \rightarrow \|T\|$. By passing to a subsequence, we may assume $\langle Tx_n, x_n \rangle \rightarrow \lambda$, where $|\lambda| = \|T\|$. Since (x_n) is bounded and T is compact, by passing to a subsequence we may assume $Tx_n \rightarrow y$ for some $y \in \mathcal{H}$. Since $\|T\| \neq 0$ and $\|T\| = \lim |\langle Tx_n, x_n \rangle|$, then $y \neq 0$. Now using Lemma 4.14(ii),

$$\|Tx_n - \lambda x_n\|^2 = \|Tx_n\|^2 - 2\lambda \langle Tx_n, x_n \rangle + \lambda^2 \|x_n\|^2 \leq 2\|T\|^2 - 2\lambda \langle Tx_n, x_n \rangle.$$

As $n \rightarrow \infty$, we get $\|Tx_n - \lambda x_n\| \rightarrow 0$. Since $Tx_n \rightarrow y$, we thus have $\lambda x_n \rightarrow y$. As T is continuous, we get $\lambda Tx_n \rightarrow Ty$. But since $Tx_n \rightarrow y$, we also have $\lambda Tx_n \rightarrow \lambda y$. Hence, $Ty = \lambda y$, so y is an eigenvector.

Using Zorn's lemma, we can now choose an orthonormal set of eigenvectors E for T which is maximal among all orthonormal sets of eigenvectors.⁴ Let $W = \text{sp } E$ be the span of these vectors. It suffices to see that $\overline{W} = \mathcal{H}$, i.e. $W^\perp = \{0\}$. We clearly have $TW \subseteq W$, so by Lemma 4.14(i), $TW^\perp \subseteq W^\perp$. Hence, $T|_{W^\perp} \in B(W^\perp)$ is self-adjoint, and if $W^\perp \neq \{0\}$, then by the preceding paragraph, there is an eigenvector for T in W^\perp . This contradicts the maximality of E and proves the first assertion of the theorem. The second assertion follows from Example 4.5. \square

We now generalize this result to normal operators. We first note the following.

Corollary 4.16. *Let S, T be compact self-adjoint operators on a separable Hilbert space \mathcal{H} such that $ST = TS$. Then \mathcal{H} has an orthonormal basis $\{e_n\}$ such that e_n is an eigenvector for both S and T .*

Proof. Let \mathcal{H}_λ be the eigenspace for T as in Definition 4.12. If $x \in \mathcal{H}_\lambda$, then $T(Sx) = STx = S(\lambda x) = \lambda(Sx)$. Hence, $Sx \in \mathcal{H}_\lambda$, so $S : \mathcal{H}_\lambda \rightarrow \mathcal{H}_\lambda$. Since $S|_{\mathcal{H}_\lambda}$ is compact and self-adjoint we may choose an orthonormal basis $\{e_{n,\lambda}\}_n$ of \mathcal{H}_λ consisting of eigenvectors of S . Of course, $\{e_{n,\lambda}\}_n$ are also eigenvectors of T since $e_{n,\lambda} \in \mathcal{H}_\lambda$. Taking $\{e_n\} = \cup_\lambda \{e_{n,\lambda}\}$ we get the basis. \square

Definition 4.17. Let \mathcal{H} be a Hilbert space. We say that $T \in B(\mathcal{H})$ is *normal* if it commutes with its adjoint, i.e. $T^*T = TT^*$.

For example, self-adjoint operators are normal. Unitary operators are also normal (recall that U is unitary if U is bijective and $U^* = U^{-1}$). On the other hand, $T = 2iI$ is normal, but neither self-adjoint nor unitary.

Theorem 4.18 (Spectral Theorem II). *Let T be a compact normal operator on a separable Hilbert space \mathcal{H} . Then \mathcal{H} has an orthonormal basis $\{e_n\}$ consisting of eigenvectors of T , i.e. $Te_n = \gamma_n e_n$. If \mathcal{H} is infinite-dimensional, then $\gamma_n \rightarrow 0$.*

Proof. Let $T_1 = \frac{T+T^*}{2}$ and $T_2 = \frac{T-T^*}{2i}$. Then $T = T_1 + iT_2$ and T_j are self-adjoint. Moreover, T_1 and T_2 commute since T is normal. So by Corollary 4.16, there is an orthonormal basis $\{e_n\}$ for \mathcal{H} such that $Te_n = T_1 e_n + iT_2 e_n = \lambda_n e_n + i\mu_n e_n = (\lambda_n + i\mu_n)e_n$. So we may take $\gamma_n = \lambda_n + i\mu_n$. Finally, $\gamma_n \rightarrow 0$ by Example 4.5. \square

4.3 Application: Peter-Weyl Theorem

The Peter-Weyl theorem is an important result in harmonic analysis which can be regarded as a generalization of the completeness of the Fourier series. Let us start with two examples to better understand the theorem.

Example 4.19. Consider the unit circle $\mathbb{T} \subset \mathbb{C}$ and let $e_k(z) = z^k$ on \mathbb{T} . Then the Fourier basis $\{e_k\}_{k \in \mathbb{Z}}$ is an orthogonal basis of $L^2(\mathbb{T})$. The fact that they are orthogonal is a simple calculation (here $\int_{\mathbb{T}} z^k \overline{z^j} dz = \int_0^{2\pi} e^{ikx} e^{-ijx} dx$). Their span is dense in $L^2(\mathbb{T})$ because by Stone-Weierstrass, their span is (uniformly) dense in $C(\mathbb{T})$, which is dense in $L^2(\mathbb{T})$.

Hence, if we let $\mathcal{H}_k = \mathbb{C} \cdot e_k = \{\alpha e_k : \alpha \in \mathbb{C}\}$, then $L^2(\mathbb{T}) = \widehat{\bigoplus_{k \in \mathbb{Z}} \mathcal{H}_k}$. Here $\widehat{\bigoplus}$ denotes the closure of the direct sum. Moreover, the \mathcal{H}_k are mutually orthogonal. Also, if we define $(\pi_a f)(z) = f(az)$ for $a \in \mathbb{T}$ and $f \in L^2(\mathbb{T})$, then $(\pi_a e_k)(z) = e_k(az) = a^k z^k$, i.e. $\pi_a e_k = a^k e_k$. Hence, \mathcal{H}_k is invariant under each π_a .

4. As usual, consider the class of all orthonormal sets of eigenvectors for T , ordered by set inclusion. This class is nonempty since it contains $\{v\}$, where $v = y/\|y\|$ is the eigenvector we constructed. Any chain $\{E_\alpha\}$ has an upper bound, namely $\cup E_\alpha$, which is clearly orthonormal. So we can apply Zorn's lemma.

Example 4.20. Let $G \subset GL(n, \mathbb{R})$ be a compact subgroup. For each r , let P_r be the space of polynomial functions on $\mathbb{R}^{n \times n}$ of degree at most r , i.e. if $p \in P_r$, then $p : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ and $p : (a_{i,j})_{i,j=1}^n \mapsto \sum_{k=0}^r \sum_{i,j=1}^n c_{i,j,k} a_{i,j}^k$. Each $g \in G$ acts on $\mathbb{R}^{n \times n}$ by matrix multiplication. Moreover, for a fixed g , this map $\mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ is linear. Hence, if $f \in P_r$ and $\pi_g f(x) = f(g^{-1}x)$, then $\pi_g f \in P_r$. Thus, if $V_r = \{f|_G : f \in P_r\}$, then V_r is π_g -invariant. Moreover, $V_r \subset V_{r+1}$, $\dim V_r < \infty$, and $\cup_i V_i$ is (uniformly) dense in $C(G)$ by Stone-Weierstrass. Hence, $\cup_i V_i$ is dense in $L^2(G)$. Let $\mathcal{H}_r = \{x \in V_r : x \perp V_{r-1}\}$. Then $\dim \mathcal{H}_r < \infty$, $\{\mathcal{H}_r\}$ are mutually orthogonal, $L^2(G) = \widehat{\bigoplus_r \mathcal{H}_r}$, and \mathcal{H}_r is invariant under each π_g , $g \in G$, assuming G is equipped with a left invariant (Haar) measure.

Definition 4.21. Let G be a group, V a vector space and $GL(V)$ the group of invertible linear maps on V . A *representation of G on V* is a homomorphism $\pi : G \rightarrow GL(V)$, $\pi : g \mapsto \pi_g$. We say that $W \subseteq V$ is *invariant under π* if $\pi_g(W) \subseteq W$ for all $g \in G$. We say that the representation π is *irreducible* if it has no closed invariant subspaces other than $\{0\}$ and V . If G is a topological group and \mathcal{H} is a Hilbert space, we say that a representation $\pi : G \rightarrow GL(\mathcal{H})$ is *unitary* if each π_g is a unitary operator on \mathcal{H} , and if the map $g \mapsto \pi_g f$ is continuous for all $f \in \mathcal{H}$.

Definition 4.22. Let G be a compact group with a left invariant (Haar) measure μ (see Chapter 3). We define the *(left) regular representation of G* by $\pi : G \rightarrow L^2(G, \mu)$, $(\pi_g f)(x) = f(g^{-1}x)$.

It can be shown that the regular representation is unitary.

Theorem 4.23 (Peter-Weyl). *Let G be a compact group and π the regular representation on $L^2(G)$. Then $L^2(G) = \widehat{\bigoplus_r \mathcal{H}_r}$, where \mathcal{H}_r are mutually orthogonal, $\dim \mathcal{H}_r < \infty$ and \mathcal{H}_r is invariant and irreducible for π .*

Proof. Given $\varphi \in C(G)$, define $K(x, y) = \varphi(x^{-1}y)$. Then the integral operator $T = T_K$ is compact by Example 4.11. Hence, T^*T is compact by Lemma 4.4(b) and clearly self-adjoint. By the Spectral Theorem 4.15, it follows that

$$L^2(G) = \ker(T^*T) \oplus \widehat{\bigoplus_{\lambda \neq 0} \mathcal{H}_\lambda}, \quad (\dagger)$$

where \mathcal{H}_λ is the corresponding eigenspace for T^*T , and for $\lambda \neq 0$, $\dim \mathcal{H}_\lambda < \infty$ since each nonzero eigenvalue has a finite multiplicity.

We claim that each \mathcal{H}_λ is invariant for π . First note that if $g \in G$, then $[T(\pi_g f)](x) = \int K(x, y) f(g^{-1}y) d\mu(y)$. As μ is invariant, we may replace y by gy to get $[T(\pi_g f)](x) = \int K(x, gy) f(y) d\mu(y) = \int K(g^{-1}x, y) f(y) d\mu(y) = [\pi_g(Tf)](x)$. Hence, T commutes with π_g . Also, T^*T commutes with π_g . Indeed, first note that $(\pi_g f, h) = \int f(g^{-1}x) \bar{h}(x) = \int f(x) \bar{h}(gx) = (f, \pi_{g^{-1}} h)$, so $\pi_g^* = \pi_{g^{-1}}$. Hence, $T^* \pi_g = T^* \pi_{g^{-1}} = (\pi_{g^{-1}} T)^* = (T \pi_{g^{-1}})^* = \pi_g T^*$ and thus $T^* T \pi_g = T^* \pi_g T = \pi_g T^* T$. It follows that \mathcal{H}_λ is invariant for π . Indeed, if $v \in \mathcal{H}_\lambda$, then $T^* T(\pi_g v) = \pi_g(T^* T v) = \pi_g(\lambda v) = \lambda(\pi_g v)$, so $\pi_g v \in \mathcal{H}_\lambda$ as asserted.

We can now use Zorn's lemma to choose a maximal collection $\{V_j\}$ of mutually orthogonal, finite dimensional subspaces $V_j \subset L^2(G)$ with V_j invariant and irreducible for π (such V_j exist by the preceding paragraph). Let $W = \bigoplus V_j$. We claim that $\bar{W} = L^2(G)$. If not, then we can choose $f \in W^\perp$, $f \neq 0$. Since $C(G)$ is dense in $L^2(G)$, we can find $\varphi \in C(G)$ such that $(\varphi, \bar{f}) \neq 0$. Again define $K(x, y) = \varphi(x^{-1}y)$ and $T = T_K$ to obtain the decomposition (\dagger) . Now let $P : L^2(G) \rightarrow W^\perp$ be the orthogonal projection. By construction, W is invariant for π . So W^\perp is also invariant for π . Indeed, if $v \in W^\perp$ and

$w \in W$, then $(\pi_g v, w) = (v, \pi_{g^{-1}} w) = 0$ because $\pi_{g^{-1}} w \in W$, hence $\pi_g v \in W^\perp$ as asserted. It follows that P commutes with all π_g . Indeed, if $u \in L^2(G)$ and $v \in W^\perp$, we have

$$(\pi_g u - \pi_g P u, v) = (u, \pi_{g^{-1}} v) - (P u, \pi_{g^{-1}} v) = (u, \pi_{g^{-1}} v) - (u, P \pi_{g^{-1}} v).$$

But $P \pi_{g^{-1}} v = \pi_{g^{-1}} v$ since $\pi_{g^{-1}} v \in W^\perp$. Hence, $(\pi_g u - \pi_g P u, v) = 0$. Thus, $\pi_g u = \pi_g P u + (\pi_g u - \pi_g P u)$ with $\pi_g P u \in W^\perp$ and $(\pi_g u - \pi_g P u) \perp W^\perp$. So by definition of the orthogonal projection, $P \pi_g u = \pi_g P u$, so P commutes with π_g . Hence, $\pi_g(P \mathcal{H}_\lambda) = P(\pi_g \mathcal{H}_\lambda) \subseteq P(\mathcal{H}_\lambda)$, so $P(\mathcal{H}_\lambda) \subseteq W^\perp$ is also invariant for π . Since $\dim P(\mathcal{H}_\lambda) < \infty$, this will contradict the maximality of $\{V_j\}$ if we can show that at least one $P(\mathcal{H}_\lambda) \neq \{0\}$.

However, if $P(\mathcal{H}_\lambda) = \{0\}$ for all $\lambda \neq 0$, then $\widehat{\bigoplus} \mathcal{H}_\lambda \subseteq \overline{W}$ and thus $W^\perp = \overline{W}^\perp \subseteq (\widehat{\bigoplus} \mathcal{H}_\lambda)^\perp$, i.e. $W^\perp \subseteq \ker T^* T$. But if $T^* T v = 0$, then $0 = (T^* T v, v) = (T v, T v)$, so $T v = 0$. Thus, we would have $W^\perp \subseteq \ker T$. However, $(T f)(e) = \int K(e, y) f(y) d\mu(y) = \int \varphi(y) f(y) d\mu(y) = (\varphi, f) \neq 0$. Moreover, $T f$ is a continuous function on G . Hence, $T f \neq 0$ in $L^2(G)$. But $f \in W^\perp$, so this contradicts $W^\perp \subseteq \ker T$. Hence, at least one $P(\mathcal{H}_\lambda) \neq \{0\}$, which completes the proof. \square

The previous theorem can be used to prove other versions of the Peter-Weyl theorem, for example the following one.

Theorem 4.24 (Peter-Weyl II). *Let G be a compact group and σ a unitary representation of G on a Hilbert space \mathcal{H} . Then $\mathcal{H} = \widehat{\bigoplus}_r \mathcal{H}_r$, where \mathcal{H}_r are mutually orthogonal, $\dim \mathcal{H}_r < \infty$ and \mathcal{H}_r is invariant and irreducible for σ .*

Proof. See [44, Theorem 3.3.9], or your favorite book in harmonic analysis or representation theory. \square

4.4 The spectrum of a bounded operator

We saw in Example 4.13 that a bounded self-adjoint operator may have no eigenvalues. Hence, if we want to obtain a spectral theorem for such operators, we should weaken the notion of eigenvalues. This motivates the following.

Definition 4.25. Let X be a Banach space and $T \in B(X)$.

We say that T is *invertible* if T^{-1} exists and $T^{-1} \in B(X)$.

We define the *spectrum* of T by

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda I \text{ is not invertible}\}.$$

Here $T - \lambda := T - \lambda I$; a standard abuse of notation.

Example 4.26. 1. If λ is an eigenvalue, then $\lambda \in \sigma(T)$ since $T_\lambda := T - \lambda I$ is not injective.

If $\dim X < \infty$, then $\lambda \in \sigma(T)$ iff λ is an eigenvalue. Indeed, if λ is not an eigenvalue, then T_λ is injective. So $\dim X = \dim T_\lambda(X)$, so $T_\lambda(X) = X$, so T_λ is surjective. Hence, T_λ^{-1} exists, and $T_\lambda^{-1} \in B(X)$ by the open mapping theorem. Hence, $\lambda \notin \sigma(T)$.

2. Let \mathcal{H} be a Hilbert space with orthonormal basis $\{e_i\}$. Define $T \in B(\mathcal{H})$ by $T e_i = \lambda_i e_i$ with $\lambda_i \rightarrow 0$ as $i \rightarrow \infty$, but $\lambda_i \neq 0$ for all i . Then 0 is not an eigenvalue, but $0 \in \sigma(T)$. Indeed, if T was invertible, we would have $T^{-1} e_i = \lambda_i^{-1} e_i$. Since $\lambda_i \rightarrow 0$, this cannot be a bounded operator.

Definition 4.27. Let X be a Banach space and $T \in B(X)$. We say that T is *bounded below* if there is some $k > 0$ such that $\|T x\| \geq k \cdot \|x\|$ for all $x \in X$.

Lemma 4.28. *If $T \in B(X)$, then T is invertible iff T is bounded below and $T(X)$ is dense.*

Proof. If T is invertible, then $T(X) = X$ and $\|x\| = \|T^{-1}Tx\| \leq \|T^{-1}\|\|Tx\|$, i.e. T is bounded below by $k = \|T^{-1}\|^{-1}$. Conversely, if T is bounded below by k , then T is injective since $Tx = 0$ implies $\|x\| \leq k^{-1}\|Tx\| = 0$, i.e. $x = 0$. If $T(X)$ is moreover dense, given $y \in X$, choose $x_n \in X$ such that $Tx_n \rightarrow y$. Then $\{Tx_n\}$ is Cauchy and $\|Tx_n - Tx_m\| \geq k\|x_n - x_m\|$. Hence, $\{x_n\}$ is also Cauchy, so $x_n \rightarrow x$ for some x since X is complete. Hence, $Tx_n \rightarrow Tx$, so $Tx = y$. Thus, T is surjective, hence bijective. Finally, $\|T^{-1}x\| \leq k^{-1}\|Tx\| = k^{-1}\|x\|$, so $\|T^{-1}\| \leq k^{-1}$. Conclusion: T is invertible. \square

Example 4.29. Let $\mathcal{H} = L^2(X, \mu)$ and suppose $\varphi \in L^\infty(X)$. Let M_φ be the multiplication operator on \mathcal{H} given by $M_\varphi(f) = \varphi f$. Then for any $\lambda \in \mathbb{C}$, $M_\varphi - \lambda = M_{\varphi - \lambda}$. We claim that $\sigma(M_\varphi) = \text{ess range}(\varphi)$, where⁵

$$\text{ess range}(\varphi) = \{\lambda \in \mathbb{C} \mid \forall \varepsilon > 0 : \mu\{x : |\varphi(x) - \lambda| < \varepsilon\} > 0\}.$$

Indeed, if $\lambda \notin \text{ess range}(\varphi)$, then there exists $\varepsilon > 0$ such that, if $A_\varepsilon = \{x : |\varphi(x) - \lambda| < \varepsilon\}$, then $\mu(A_\varepsilon) = 0$. Hence, given $f \in \mathcal{H}$, $\|M_{1/(\varphi - \lambda)}f\|_2^2 = \int_{X \setminus A_\varepsilon} \left| \frac{f(x)}{\varphi(x) - \lambda} \right|^2 \leq \varepsilon^{-2} \|f\|_2^2$. Hence, $M_{\varphi - \lambda}$ is invertible, with inverse $M_{1/(\varphi - \lambda)}$, so $\lambda \notin \sigma(M_\varphi)$. Conversely, if $\lambda \in \text{ess range}(\varphi)$, define $S_m = \{x : |\varphi(x) - \lambda| < 2^{-m}\}$ and let $\chi_m = \chi_{S_m}$. Then $\mu(S_m) > 0$, so $\chi_m \neq 0$. Moreover, $\|M_{\varphi - \lambda}\chi_m\|_2^2 = \int_{S_m} |\varphi(x) - \lambda|^2 \leq 2^{-2m} \mu(S_m) = 2^{-2m} \|\chi_m\|_2^2$. Hence, $M_{\varphi - \lambda}$ is not bounded below, so not invertible. Thus, $\lambda \in \sigma(M_\varphi)$.

Lemma 4.30. *Let \mathcal{H} be a Hilbert space and $T \in B(\mathcal{H})$. If T and T^* are bounded below, then T is invertible.*

Proof. By Lemma 4.28, it suffices to show that $T(\mathcal{H})$ is dense. But $T(\mathcal{H})^\perp = \ker T^*$. Indeed, $y \in T(\mathcal{H})^\perp \iff \langle y, Tx \rangle = 0$ for all $x \in \mathcal{H} \iff \langle T^*y, x \rangle = 0$ for all $x \in \mathcal{H} \iff T^*y = 0$. Since T^* is bounded below, $\ker T^* = \{0\}$. Hence, $T(\mathcal{H})^\perp = \{0\}$, so $T(\mathcal{H})$ is dense. \square

We saw in Lemma 4.14 that the eigenvalues of a bounded self-adjoint operator are real. We now have the following generalization.

Lemma 4.31. *Let \mathcal{H} be a Hilbert space and $T \in B(\mathcal{H})$.*

- (a) *If T is self-adjoint, then $\sigma(T) \subseteq \mathbb{R}$.*
- (b) *If T is unitary, then $\sigma(T) \subseteq \{\lambda \in \mathbb{C} : |\lambda| = 1\}$.*

Proof. (a) By Lemma 4.30, it suffices to show that if $|\text{Im } \lambda| \neq 0$, then $T - \lambda$ and $(T - \lambda)^*$ are bounded below. Let $\|y\| = 1$. Then $\|(T - \lambda)y\| \geq |\langle (T - \lambda)y, y \rangle| = |\langle Ty, y \rangle - \lambda \langle y, y \rangle|$. Since $\langle Ty, y \rangle \in \mathbb{R}$ by Lemma 4.14 and $\|y\| = 1$, then $\|(T - \lambda)y\| \geq |\text{Im } \lambda|$. Thus, $\|(T - \lambda)x\| \geq |\text{Im } \lambda| \cdot \|x\|$ for all $x \in \mathcal{H}$. Also, $\|(T - \lambda)^*x\| = \|(T - \bar{\lambda})x\| \geq |\text{Im } \bar{\lambda}| \cdot \|x\| = |\text{Im } \lambda| \cdot \|x\|$. Hence, $T - \lambda$ and $(T - \lambda)^*$ are bounded below.

5. Let $\varphi_*\mu$ be the pushforward measure $\varphi_*\mu(B) = \mu(\varphi^{-1}(B))$ for Borel $B \subseteq \mathbb{C}$. Then λ is in the essential range of φ iff every neighborhood of λ has positive $\varphi_*\mu$ -measure. In general, $\text{ess range}(\varphi) = \bigcap_{\psi=\varphi \text{ a.e.}} \overline{\text{Ran } \psi}$.

If X is a compact subset of \mathbb{R}^n , $\varphi \in C(X)$ and μ is the Lebesgue measure, then $\text{ess range}(\varphi) = \text{Ran}(\varphi)$. Indeed, if $\lambda \in \text{ess range}(\varphi)$, then for each $\varepsilon > 0$, we have $\mu\{x : |\varphi(x) - \lambda| < \varepsilon\} > 0$. In particular, $\{x : |\varphi(x) - \lambda| < \varepsilon\} \neq \emptyset$. Thus, $B_{\lambda, \varepsilon} \cap \text{Ran}(\varphi) \neq \emptyset$. Since $\varepsilon > 0$ is arbitrary, we get $\lambda \in \text{Ran}(\varphi)$. But $\text{Ran}(\varphi)$ is compact, hence closed, so $\text{ess range}(\varphi) \subseteq \text{Ran}(\varphi)$. Conversely, if $\lambda \in \text{Ran}(\varphi)$, then $\lambda = \varphi(x_0)$ for some x_0 . Let $\varepsilon > 0$. Since φ is continuous, there exists $\delta > 0$ such that $|\varphi(x_0) - \varphi(x)| < \varepsilon$ whenever $|x - x_0| < \delta$. Hence, $\mu\{x : |\varphi(x_0) - \varphi(x)| < \varepsilon\} \geq \mu\{x : |x - x_0| < \delta\} = c_n \delta^n > 0$. Thus, $\lambda \in \text{ess range}(\varphi)$.

- (b) It suffices to show that if $|\lambda| \neq 1$, then $T - \lambda$ and $(T - \lambda)^*$ are bounded below. We have $\|(T - \lambda)x\| \geq |\|Tx\| - \|\lambda x\|| = |1 - |\lambda|| \cdot \|x\|$, since $\|Tx\| = \|x\|$. Also, $\|(T - \lambda)^*x\| = \|(T^{-1} - \bar{\lambda})x\| \geq |1 - |\bar{\lambda}|| \cdot \|x\|$, since T^{-1} is unitary. Hence, $T - \lambda$ and $(T - \lambda)^*$ are bounded below. \square

The following lemma is a special case of the *spectral mapping theorem*.

Lemma 4.32. *Let X be a Banach space, $T \in B(X)$ and $p(t)$ a polynomial. Then*

$$\sigma(p(T)) = p(\sigma(T)).$$

Proof. If $p(t) = c$ is constant, then $p(T) = cI$ and $\sigma(cI) = c = p(\sigma(T))$. So suppose p is not constant. Given $\lambda \in \mathbb{C}$, let $\{\mu_i\}$ be the roots of the polynomial $p(t) - \lambda$. Then we have $p(T) - \lambda = a_0 \prod_{i=1}^n (T - \mu_i)$, where $a_0 \neq 0$. If $T - \mu_i$ is invertible for each i , then $p(T) - \lambda$ is invertible. Conversely, if $T - \mu_{i_0}$ is not invertible for some i_0 , then $p(T) - \lambda$ is not invertible. Indeed, by the open mapping theorem, $T - \mu_{i_0}$ is either non-injective or non-surjective. If it is not injective, then $p(T) - \lambda = (T - \mu_1) \cdots (T - \mu_n)(T - \mu_{i_0})$ is not injective. If it is not surjective, then $p(T) - \lambda = (T - \mu_{i_0})(T - \mu_1) \cdots (T - \mu_n)$ is not surjective. We thus showed that $p(T) - \lambda$ is invertible iff each $T - \mu_i$ is invertible.

Hence, $\lambda \in \sigma(p(T)) \iff \mu_i \in \sigma(T)$ for some i . But $\mu_i \in \sigma(T)$ implies $\lambda \in p(\sigma(T))$, since $p(\mu_i) - \lambda = 0$. Conversely, $\lambda \in p(\sigma(T))$ implies $\lambda = p(\mu)$ for some $\mu \in \sigma(T)$. So μ is a root of $p(t) - \lambda$, so $\mu = \mu_i$ for some i . Hence, $\mu_i \in \sigma(T)$. Conclusion: $\lambda \in \sigma(p(T)) \iff \lambda \in p(\sigma(T))$. \square

We now turn to the main result on the spectrum of a bounded operator.

Definition 4.33. Let $T \in B(X)$. We define

- The *resolvent set* of T by $\rho(T) = \mathbb{C} \setminus \sigma(T)$.
- The *spectral radius* of T by $\|T\|_\sigma = \sup_{\lambda \in \sigma(T)} |\lambda|$.
- The *resolvent operator* of T by $R(\lambda) = (T - \lambda)^{-1}$ for $\lambda \in \rho(T)$.

Let X be a Banach space and $x_i \in X$. Recall that $\sum x_i$ is *absolutely convergent* if $\sum \|x_i\| < \infty$. Since $\|\sum_{i=n}^m x_i\| \leq \sum_{i=n}^m \|x_i\|$, then absolute convergence implies convergence by the Cauchy criterion.

Theorem 4.34. (i) *If $T \in B(X)$ and $\|T\| < 1$, then $I - T$ is invertible. Moreover, $(I - T)^{-1} = \sum_{n=0}^{\infty} T^n$ and $\|(I - T)^{-1}\| \leq \frac{1}{1 - \|T\|}$.*

(ii) *The set of invertible operators is open in $B(X)$.*

(iii) *If $|\lambda| > \|T\|$, then $\lambda \in \rho(T)$, $R(\lambda) = -\sum_{n=0}^{\infty} \lambda^{-n-1} T^n$ and $\|R(\lambda)\| \leq \frac{1}{|\lambda| - \|T\|}$.*

(iv) (Resolvent identity). *For all $\lambda, \mu \in \rho(T)$ we have $R(\lambda) - R(\mu) = (\lambda - \mu)R(\lambda)R(\mu)$.*

(v) *$\rho(T)$ is open, and the resolvent $R(\lambda)$ is an analytic operator-valued function. That is, any $\lambda \in \rho(T)$ has a small neighborhood in which $R(\mu)$ can be expressed as a convergent power series $R(\mu) = \sum_{n=1}^{\infty} (\mu - \lambda)^{n-1} R(\lambda)^n$.*

(vi) *$\sigma(T)$ is a compact non-empty set with $\|T\|_\sigma \leq \|T\|$.*

Proof. (i) Since $\|T\| < 1$, the series $\sum T^n$ is absolutely convergent, hence convergent, say to S . Let $S_N = \sum_{n=0}^{N-1} T^n$. Then $(I - T)S_N = S_N(I - T) = I - T^N$. Letting $N \rightarrow \infty$ we obtain $S = (I - T)^{-1}$. Finally, $\|(I - T)^{-1}\| \leq \sum_{n=0}^{\infty} \|T\|^n = \frac{1}{1 - \|T\|}$.

(ii) If T is invertible and $S \in B(X)$ satisfies $\|T - S\| < \|T^{-1}\|^{-1}$, then $\|ST^{-1} - I\| \leq \|S - T\| \|T^{-1}\| < 1$, so $ST^{-1} = I - (I - ST^{-1})$ is invertible by (i). Hence, S is invertible, so the set of invertible operators is open.

- (iii) $T - \lambda = -\lambda(I - \frac{T}{\lambda})$. Since $\|\frac{T}{\lambda}\| < 1$, $R(\lambda)$ exists by (i) and equals $-\lambda^{-1} \sum_{n=0}^{\infty} (\frac{T}{\lambda})^n$. The norm is bounded as in (i).
- (iv) $(\lambda - \mu)R(\lambda)R(\mu) = R(\lambda)(\lambda - \mu)R(\mu) = R(\lambda)[(T - \mu) - (T - \lambda)]R(\mu) = R(\lambda) - R(\mu)$.
- (v) Let $\lambda \in \rho(T)$ and $\mu \in \mathbb{C}$. Then $T - \mu = (T - \lambda) - (\mu - \lambda) = (T - \lambda)(I - (\mu - \lambda)R(\lambda))$. So by (i), $(T - \mu)$ is invertible if $\|(\mu - \lambda)R(\lambda)\| < 1$, i.e. $|\mu - \lambda| < \|R(\lambda)\|^{-1}$. This shows that $\rho(T)$ is open. By (i), $R(\mu) = R(\lambda) \sum_{n=0}^{\infty} (\mu - \lambda)^n R(\lambda)^n$ for such μ .
- (vi) By (iii), if $\lambda \in \sigma(T)$, then $|\lambda| \leq \|T\|$. Hence, $\|T\|_{\sigma} \leq \|T\|$. In particular, $\sigma(T)$ is bounded, and it is closed by (v). Suppose that $\sigma(T) = \emptyset$. Then $R(\lambda)$ is analytic on all \mathbb{C} by (v). We claim it is also bounded. Indeed, by continuity, it is bounded on the compact set $\{\lambda : |\lambda| \leq 2\|T\|\}$, and if $|\lambda| > 2\|T\|$, then by (iii), we have $\|R(\lambda)\| \leq \frac{1}{|\lambda| - \|T\|} \leq \frac{1}{\|T\|}$. Fix $f \in B(X)^*$. We thus showed that $f(R(\lambda))$ is a bounded entire function, so by Liouville's theorem, $f(R(\lambda))$ is constant. But $|f(R(\lambda))| \leq \|f\| \cdot \|R(\lambda)\| \leq \frac{\|f\|}{|\lambda| - \|T\|} \rightarrow 0$ as $|\lambda| \rightarrow \infty$. Hence, $f(R(\lambda)) = 0$, since it is constant. Since this holds for all $f \in B(X)^*$, then by Hahn-Banach, $R(\lambda) = 0$. This contradicts the fact that $R(\lambda)$ is invertible, so $\sigma(T) \neq \emptyset$. \square

Theorem 4.35. *Let X be a Banach space and $T \in B(X)$. Then*

$$\|T\|_{\sigma} = \lim_{n \rightarrow \infty} \|T^n\|^{1/n} = \inf_{n \in \mathbb{N}} \|T^n\|^{1/n}.$$

Proof. Upper bound. If $\lambda \in \sigma(T)$, then $\lambda^n \in \sigma(T^n)$ by Lemma 4.32, so $|\lambda^n| \leq \|T^n\|_{\sigma} \leq \|T^n\|$ by Theorem 4.34(v). Hence, $|\lambda| \leq \|T^n\|^{1/n}$. Thus, $\|T\|_{\sigma} \leq \inf_n \|T^n\|^{1/n}$.

Lower bound. Fix $f \in B(X)^*$. By Theorem 4.34(v), $f(R(\lambda))$ is analytic on the annulus $|\lambda| > \|T\|_{\sigma}$. And by Theorem 4.34(iii), $f(R(\lambda)) = -\sum_{n=0}^{\infty} \lambda^{-n-1} f(T^n)$ in the smaller annulus $|\lambda| > \|T\|$. By the theory of Laurent series from complex analysis, we also have $f(R(\lambda)) = -\sum_{n=0}^{\infty} \lambda^{-n-1} f(T^n)$ in the larger annulus $|\lambda| > \|T\|_{\sigma}$. In particular, $\sup_n |\lambda^{-n-1} f(T^n)| < \infty$. By the uniform boundedness principle, it follows that $\sup_n \|\lambda^{-n-1} T^n\| =: K < \infty$. Hence, $\|T^n\|^{1/n} \leq K^{1/n} |\lambda|^{1+1/n}$ for all n . So $\overline{\lim} \|T^n\|^{1/n} \leq |\lambda|$. Since this holds for all λ such that $|\lambda| > \|T\|_{\sigma}$, we have proved that $\overline{\lim} \|T^n\|^{1/n} \leq \|T\|_{\sigma}$. Putting this together with the upper bound, we get

$$\|T\|_{\sigma} \leq \inf_n \|T^n\|^{1/n} \leq \underline{\lim} \|T^n\|^{1/n} \leq \overline{\lim} \|T^n\|^{1/n} \leq \|T\|_{\sigma}.$$

This completes the proof. \square

Corollary 4.36. *If \mathcal{H} is a Hilbert space and $T \in B(\mathcal{H})$ is normal, then $\|T\|_{\sigma} = \|T\|$. More generally, $\|p(T)\| = \max_{\lambda \in \sigma(T)} |p(\lambda)|$ for any polynomial $p(t)$.*

Proof. As T^*T is self-adjoint, $\|T\|^2 = \sup_{\|x\| \leq 1} |\langle Tx, Tx \rangle| = \sup_{\|x\| \leq 1} |\langle T^*Tx, x \rangle| = \|T^*T\|$ by Lemma 4.14. But $\|T^*T\|^2 = \sup_{\|x\| \leq 1} |\langle T^*Tx, T^*Tx \rangle| = \sup_{\|x\| \leq 1} |\langle T^2x, T^2x \rangle| = \|T^2\|^2$, since $TT^* = T^*T$. Hence, $\|T\|^2 = \|T^2\|$. By induction, $\|T\|^n = \|T^n\|$ for $n = 2, 2^2, 2^3, \dots$. Thus, using Theorem 4.35, $\|T\|_{\sigma} = \lim_n \|T^n\|^{1/n} = \lim_k \|T^{2^k}\|^{1/2^k} = \lim_k \|T\| = \|T\|$.

Next, $p(T)$ is normal by induction, so using Lemma 4.32, we get $\|p(T)\| = \|p(T)\|_{\sigma} = \max_{\mu \in \sigma(p(T))} |\mu| = \max_{\mu \in p(\sigma(T))} |\mu| = \max_{\lambda \in \sigma(T)} |p(\lambda)|$. \square

Example 4.37. Let \mathcal{H} be the space of 2×2 complex matrices.

- (a) If $T = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ for $\lambda_i \in \mathbb{C}$, then T is normal. We have $\sigma(T) = \{\lambda_1, \lambda_2\}$, so

$$\|T\|_{\sigma} = \max\{|\lambda_1|, |\lambda_2|\} = \|T\|. \text{ Note that } T^n = \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix}, \text{ so } \|T^n\|^{1/n} = \|T\|.$$

- (b) Let $T = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ with $\lambda \neq 0$. Then T is not normal. We have $\sigma(T) = \{1\}$, $\|T\|_\sigma = 1$ but $\|T\| = (1 + |\lambda|^2)^{1/2}$. Here $T^n = \begin{pmatrix} 1 & n\lambda \\ 0 & 1 \end{pmatrix}$, so $\|T^n\| = (1 + |n\lambda|^2)^{1/2}$. Hence, $\|T^n\|^{1/n} \rightarrow 1$, as asserted in Theorem 4.35.

4.5 Continuous functional calculus

Definition 4.38. Let \mathcal{H} be a Hilbert space, suppose $T \in B(\mathcal{H})$ is self-adjoint, and let $f \in C(\sigma(T))$. Since $\sigma(T)$ is a compact subset of \mathbb{R} , by the Weierstrass approximation, we may find polynomials $p_n(t)$ such that $p_n(t) \rightarrow f(t)$ uniformly on $\sigma(T)$. We define the operator $f(T)$ to be the limit in $B(\mathcal{H})$ of the operators $p_n(T)$.

Lemma 4.39. $f(T)$ is well-defined. That is, the sequence $p_n(T)$ converges, and its limit is independent of the choice of the approximating polynomials.

Proof. Using Corollary 4.36, $\|p_n(T) - p_m(T)\| = \|(p_n - p_m)(T)\| = \max_{\lambda \in \sigma(T)} |(p_n - p_m)(\lambda)| \rightarrow 0$ as $n, m \rightarrow \infty$, since $\{p_n(t)\}$ converges uniformly. Hence, $\{p_n(T)\}$ is Cauchy, and converges by the completeness of $B(\mathcal{H})$.

If $q_n(t)$ is another approximating sequence of polynomials, then repeating the previous argument we get $\|q_n(T) - p_n(T)\| \rightarrow 0$, hence $\lim q_n(T) = \lim p_n(T) = f(T)$. \square

Definition 4.40. Let \mathcal{H} be a Hilbert space and $T \in B(\mathcal{H})$. We say that T is *positive*, and denote $T \geq 0$, if $\langle Tx, x \rangle \geq 0$ for all $x \in \mathcal{H}$. We say that $T \geq S$ if $T - S \geq 0$.

Theorem 4.41. Let \mathcal{H} be a Hilbert space, $T \in B(\mathcal{H})$ self-adjoint and $f \in C(\sigma(T))$.

- (a) The map $f \mapsto f(T)$ is an algebraic $*$ -homomorphism from $C(\sigma(T)) \rightarrow B(\mathcal{H})$. That is, $(fg)(T) = f(T)g(T)$, $(\alpha f)(T) = \alpha f(T)$, $1(T) = \text{id}$ and $\bar{f}(T) = f(T)^*$.
- (b) $f(T)$ is invertible iff $f(\lambda) \neq 0$ for all $\lambda \in \sigma(T)$.
- (c) (Spectral mapping theorem). $\sigma(f(T)) = f(\sigma(T))$.
- (d) $\|f(T)\| = \|f\|_\infty$.
- (e) If $Tx = \lambda x$, then $f(T)x = f(\lambda)x$.
- (f) If $f \geq 0$, then $f(T) \geq 0$.
- (g) If $TS = ST$, then $f(T)S = Sf(T)$.

Proof. (a), (e) and (g) are clear for f, g polynomials, and follow by passing to the limit. For (b), suppose first that $f(\lambda) \neq 0$ on $\sigma(T)$. Then $g = 1/f \in C(\sigma(T))$. Moreover, $f(t)g(t) = g(t)f(t) = 1$, so by (a), $f(T)g(T) = g(T)f(T) = \text{id}$, hence $f(T)$ is invertible. Conversely, suppose $f(\mu) = 0$ for some $\mu \in \sigma(T)$. Let $\{p_n\}$ be a sequence of polynomials converging uniformly to f . We may assume $p_n(\mu) = f(\mu) = 0$ by considering the sequence $q_n(t) = p_n(t) - p_n(\mu)$. Since $\mu \in \sigma(T)$, then $0 = p_n(\mu) \in \sigma(p_n(T))$ by Lemma 4.32. Hence, $p_n(T)$ is not invertible. But the set of non-invertible operators is closed by Theorem 4.34. Hence, $f(T)$ is not invertible.

For (c), we have $\mu \in \sigma(f(T))$ iff $f(T) - \mu$ is not invertible. Applying (b) for $f - \mu$, this happens iff $f - \mu$ has a zero in $\sigma(T)$, i.e. iff $\mu \in f(\sigma(T))$.

By (a), $f(T)$ is normal, so using Corollary 4.36 and (c), we have $\|f(T)\| = \|f(T)\|_\sigma = \max_{\mu \in \sigma(f(T))} |\mu| = \max_{\mu \in f(\sigma(T))} |\mu| = \max_{\lambda \in \sigma(T)} |f(\lambda)| = \|f\|_\infty$.

For (f), note that $f \geq 0$ implies $f = g^2$ for some real-valued $g \in C(\sigma(T))$, so $f(T) = g(T)^2$ by (a). Moreover, $g(T)$ is self-adjoint by (a), so $\langle f(T)x, x \rangle = \|g(T)x\|^2 \geq 0$. \square

Remark 4.42. The previous functional calculus can be generalized to normal operators. For this, one first shows that for polynomials p in z and $\bar{z} \in \mathbb{C}$, we have $\sigma(p(T, T^*)) = \{p(z, \bar{z}) : z \in \sigma(T)\}$, deduce that $\|p(T, T^*)\| = \max_{z \in \sigma(T)} |p(z, \bar{z})|$, then define $f(T)$ and find its properties as we did. The reason for this detour is that now $\sigma(T)$ can be complex, so the polynomials $p(z)$ are no longer dense in $C(K)$. To apply Stone-Weierstrass, we must consider polynomials $p(z, \bar{z})$.

The functional calculus can be used to define the operator \sqrt{T} for any positive self-adjoint operator $T \in B(\mathcal{H})$. It turns out to be the unique operator satisfying $(\sqrt{T})^2 = T$. One can then define the modulus $|T| = \sqrt{T^*T}$ for any $T \in B(\mathcal{H})$, and it can be shown that T has a *polar decomposition* $T = U|T|$. If T is invertible, then U is unitary.

4.6 Borel functional calculus

The functional calculus admits an important generalization to the set $\mathcal{B}(\sigma(T))$ of bounded Borel functions. The great advantage is that one can now consider characteristic functions χ_A , and it turns out that $\chi_A(T)$ is an orthogonal projection. The proof is a bit more difficult because Borel functions are not uniform limits of continuous functions.

Theorem 4.43. *Let \mathcal{H} be a Hilbert space, $T \in B(\mathcal{H})$ self-adjoint and $f \in \mathcal{B}(\sigma(T))$. One can define an operator $f(T) \in B(\mathcal{H})$ such that*

- (a) *The map $f \mapsto f(T)$ is an algebraic $*$ -homomorphism from $\mathcal{B}(\sigma(T)) \rightarrow B(\mathcal{H})$.*
- (b) $\|f(T)\| \leq \|f\|_\infty$.
- (c) *If $f(t) = t$, then $f(T) = T$.*
- (d) *If $f_n, f \in \mathcal{B}(\sigma(T))$ satisfy $\sup_n \|f_n\|_\infty < \infty$, and $f_n \rightarrow f$ pointwise, then $f_n(T) \rightarrow f(T)$ strongly in $B(\mathcal{H})$.*
- (e) *If $Tx = \lambda x$, then $f(T)x = f(\lambda)x$.*
- (f) *If $f \geq 0$, then $f(T) \geq 0$.*
- (g) *If $TS = ST$, then $f(T)S = Sf(T)$.*

Proof. Fix $x, y \in \mathcal{H}$. Define a functional $F_{x,y}$ on $C(\sigma(T))$ by $F_{x,y}(f) = \langle f(T)x, y \rangle$. Then $F_{x,y}$ is bounded: $|F_{x,y}(f)| \leq \|f(T)\| \|x\| \|y\| = \|f\|_\infty \|x\| \|y\|$. Hence, $F_{x,y} \in C(\sigma(T))^*$ and $\|F_{x,y}\| \leq \|x\| \|y\|$. By the Riesz-Markov representation, we may find a unique Borel regular measure $\mu_{x,y}$ on $\sigma(T)$ with total variation $|\mu_{x,y}|(\sigma(T)) \leq \|x\| \|y\|$, such that $F_{x,y}(f) = \int f d\mu_{x,y}$, i.e. $\langle f(T)x, y \rangle = \int_{\sigma(T)} f(\lambda) d\mu_{x,y}(\lambda)$ for any $f \in C(\sigma(T))$. We extend this to Borel functions by defining $(Bf)(x, y) = \int_{\sigma(T)} f(\lambda) d\mu_{x,y}(\lambda)$ for $f \in \mathcal{B}(\sigma(T))$. Then B is linear in x , conjugate linear in y , and $|(Bf)(x, y)| \leq \|f\|_\infty |\mu_{x,y}|(\sigma(T)) \leq \|f\|_\infty \|x\| \|y\|$. Thus, B is a bounded sesquilinear form, so by the Riesz representation theorem of Hilbert spaces, there exists a unique operator, call it $f(T)$, such that $(Bf)(x, y) = \langle f(T)x, y \rangle$, and $\|f(T)\| = \|B\| \leq \|f\|_\infty$. We thus defined $f(T)$ for all $f \in \mathcal{B}(\sigma(T))$ and verified (b).

If $f(t) = t$, then $\langle f(T)x, y \rangle = (Bf)(x, y) = \int_{\sigma(T)} f(\lambda) d\mu_{x,y}(\lambda) = \int_{\sigma(T)} \lambda d\mu_{x,y}(\lambda) = \langle Tx, y \rangle$. This holds for all x, y , so (c) is proved. Similarly, $(\alpha f)(T) = \alpha f(T)$ and $1(T) = \text{id}$.

Let f_n, f as in (d). By dominated convergence, $\langle f_n(T)x, y \rangle = \int_{\sigma(T)} f_n(\lambda) d\mu_{x,y}(\lambda) \rightarrow \int_{\sigma(T)} f(\lambda) d\mu_{x,y}(\lambda) = \langle f(T)x, y \rangle$ for $x, y \in \mathcal{H}$. Thus, $f_n(T)$ converges weakly to $f(T)$. We prove strong convergence later.

We now prove $(fg)(T) = f(T)g(T)$. Fix $g \in C(\sigma(T))$ and let $\mathcal{V} = \{f \in \mathcal{B}(\sigma(T)) : (fg)(T) = f(T)g(T)\}$. Then $\mathcal{V} \supset C(\sigma(T))$ by Theorem 4.41. Moreover, \mathcal{V} is a vector space since $[(\alpha f_1 + \beta f_2)g](T) = \alpha(f_1g)(T) + \beta(f_2g)(T)$. To see this equality, note that

$\langle [\alpha h_1 + \beta h_2](T)x, y \rangle = \int [\alpha h_1 + \beta h_2] d\mu_{x,y} = \alpha \int h_1 d\mu_{x,y} + \beta \int h_2 d\mu_{x,y} = \langle [\alpha h_1(T) + \beta h_2(T)]x, y \rangle$ for any x, y . Clearly, the constant function $1 \in \mathcal{V}$. Finally, suppose $f_n \in \mathcal{V}$ satisfy $\sup_n \|f_n\|_\infty < \infty$ and $f_n \rightarrow f$ pointwise. Then $\langle (fg)(T)x, y \rangle = \lim \langle (f_n g)(T)x, y \rangle = \lim \langle f_n(T)g(T)x, y \rangle = \langle f(T)g(T)x, y \rangle$ by weak convergence. This holds for any x, y , so $(fg)(T) = f(T)g(T)$. Hence, $f \in \mathcal{V}$. By the monotone class theorem, we thus have $\mathcal{V} = \mathcal{B}(\sigma(T))$. Now fix $f \in \mathcal{B}(\sigma(T))$ and let $\mathcal{V}' = \{g \in \mathcal{B}(\sigma(T)) : (fg)(T) = f(T)g(T)\}$. Then \mathcal{V}' is again a vector space, $1 \in \mathcal{V}'$, and $\mathcal{V}' \supset C(\sigma(T))$ by what we just proved. If $g_n \in \mathcal{V}'$, $\sup \|g_n\|_\infty < \infty$ and $g_n \rightarrow g$ pointwise, then $\langle (fg)(T)x, y \rangle = \lim \langle (fg_n)(T)x, y \rangle = \lim \langle f(T)g_n(T)x, y \rangle = \lim \langle g_n(T)x, f(T)^*y \rangle = \langle g(T)x, f(T)^*y \rangle = \langle f(T)g(T)x, y \rangle$. Hence, $(fg)(T) = f(T)g(T)$. Using monotone class again, this completes the proof. Similarly we prove $\bar{f}(T) = f(T)^*$ and items (e) and (g). Moreover, (a) implies (f) as before.

We finally prove strong convergence for f_n, f as in (d). We know $|f_n - f|^2(T)$ converges weakly to 0, so using (a) we get $\|(f_n - f)(T)x\|^2 = \langle (f_n - f)(T)x, (f_n - f)(T)x \rangle = \langle |f_n - f|^2(T)x, x \rangle \rightarrow 0$, as asserted. \square

Definition 4.44. Let $T \in B(\mathcal{H})$ be self-adjoint and let $x, y \in \mathcal{H}$. We define the *spectral measure* $\mu_{x,y}$ of T to be the unique regular Borel measure satisfying

$$\langle f(T)x, y \rangle = \int_{\sigma(T)} f(\lambda) d\mu_{x,y}(\lambda)$$

for every $f \in \mathcal{B}(\sigma(T))$. As shown in the previous proof, such a measure exists, and has a total variation $|\mu_{x,y}|(\sigma(T)) \leq \|x\|\|y\|$. We denote $\mu_x = \mu_{x,x}$.

Example 4.45. Let M_φ be the multiplication operator of Example 4.29 and let $f \in \mathcal{B}(\sigma(M_\varphi))$. Then $f(M_\varphi) = M_{f \circ \varphi}$. To see this, let $\mathcal{V} = \{f \in \mathcal{B}(\sigma(M_\varphi)) : f(M_\varphi) = M_{f \circ \varphi}\}$. Clearly, \mathcal{V} contains polynomials. If $f \in C(\sigma(M_\varphi))$, let p_n be polynomials converging uniformly to f . Then $\|M_{p_n \circ \varphi} \psi - M_{f \circ \varphi} \psi\|_{L^2} = \|(p_n \circ \varphi)\psi - (f \circ \varphi)\psi\| = \|[(p_n - f) \circ \varphi]\psi\| \leq \|p_n - f\|_\infty \|\psi\|_{L^2}$, so $\|M_{p_n \circ \varphi} - M_{f \circ \varphi}\| \leq \|p_n - f\|_\infty \rightarrow 0$, hence $M_{f \circ \varphi} = \lim M_{p_n \circ \varphi} = \lim p_n(M_\varphi) = f(M_\varphi)$, where the last equality holds by definition of $f(M_\varphi)$. We thus showed that $\mathcal{V} \supset C(\sigma(M_\varphi))$. Clearly, \mathcal{V} is a vector space containing the constant function 1. If $f_n \in \mathcal{V}$, $f_n \geq 0$, $\sup_n \|f_n\|_\infty < \infty$ and $f_n \uparrow f$ pointwise, then by Theorem 4.43, $f(M_\varphi)\psi = \lim f_n(M_\varphi)\psi = \lim M_{f_n \circ \varphi} \psi = \lim [f_n \circ \varphi]\psi = [f \circ \varphi]\psi = M_{f \circ \varphi} \psi$. Here $\|[f_n \circ \varphi]\psi - [f \circ \varphi]\psi\|_{L^2} \rightarrow 0$ by monotone convergence. As $\psi \in L^2$ is arbitrary, we get $f(M_\varphi) = M_{f \circ \varphi}$, i.e. $f \in \mathcal{V}$. Using monotone class, we thus get $\mathcal{V} = \mathcal{B}(\sigma(M_\varphi))$.

Definition 4.46. Let T be self-adjoint. We define the *spectral projection* $P_T : \mathcal{B}(\sigma(T)) \rightarrow B(\mathcal{H})$ by $P_T(E) = \chi_E(T)$, where χ_E is the characteristic function of E .

Lemma 4.47 (Projection-valued measure). $P_T(E)$ is an orthogonal projection for any $E \in \mathcal{B}(\sigma(T))$. Moreover,

- (i) $P_T(\emptyset) = 0$, $P_T(\sigma(T)) = \text{id}$,
- (ii) For any Borel partition $\sigma(T) = \cup_1^\infty E_k$, we have $\text{id} = \sum_1^\infty P_T(E_k)$. Here convergence is in the strong sense.

Proof. As $\chi_E(t)^2 = \chi_E(t)$ and $\chi_E(t)$ is real-valued, then by Theorem 4.43(a), $P_T(E)^2 = P_T(E)$ and $P_T(E)$ is self-adjoint. Hence, $P_T(E)$ is a projection.

Let $x, y \in \mathcal{H}$. Then $\langle \chi_\emptyset(T)x, y \rangle = \int \chi_\emptyset(\lambda) d\mu_{x,y}(\lambda) = 0$, so $P_T(\emptyset) = 0$. Similarly, $\chi_{\sigma(T)}(T) = 1(T) = \text{id}$. If $f_n(t) = \sum_1^n \chi_{E_k}(t)$, then $\|f_n\|_\infty = 1$ since the E_k are disjoint, and $f_n(t) \rightarrow \chi_{\sigma(T)}$ pointwise since $\cup E_k = \sigma(T)$. Hence, $\sum_1^n P_T(E_k) \rightarrow \text{id}$ strongly. \square

Theorem 4.48 (Spectral Theorem III). *Let \mathcal{H} be a Hilbert space and let $T \in B(\mathcal{H})$ be self-adjoint. Then for any $f \in \mathcal{B}(\mathbb{R})$, we have*

$$f(T) = \int_{\sigma(T)} f(\lambda) dP_T(\lambda),$$

in the sense that $\langle f(T)x, x \rangle = \int_{\sigma(T)} f(\lambda) d\langle P_T(\lambda)x, x \rangle$ for any $x \in \mathcal{H}$.

If T is a compact self-adjoint operator, then it has an orthonormal basis of eigenvectors. If λ_k are its eigenvalues and if P_k is the orthogonal projection onto $\mathcal{H}_k = \{x : Tx = \lambda_k x\}$, then $T = \sum_k \lambda_k P_k$, as the reader can check. If T is only bounded, the spectrum is not discrete in general, and this is why the sum gets replaced by an integral.

Proof. We have $\langle P_T(E)x, x \rangle = \int_{\sigma(T)} \chi_E(\lambda) d\mu_x(\lambda) = \mu_x(E)$ for any Borel $E \subset \sigma(T)$. So the theorem follows from the functional calculus. \square

4.7 Further results

Each section of this chapter could be considerably expanded. Compact operators have a rich theory that extends beyond spectral theory, especially in connection with the operator ideals \mathcal{S}_p we briefly mentioned. See [29] for an elementary treatment and [35] for the advanced theory. There are also many results in the spectral theory of compact operators which we did not cover. A famous one is the *Fredholm alternative*, which says that if X is a Banach space, $T \in K(X)$ and $\lambda \in \mathbb{C}^*$, then exactly one of the following statements is true:

- (i) either the homogeneous equation $\lambda x - Tx = 0$ has a nontrivial solution,
- (ii) or the inhomogeneous equation $\lambda x - Tx = b$ has a solution for every $b \in X$, and this solution is automatically unique.

This alternative is very useful for studying integral equations. In this case, the homogeneous equation is $\lambda f(t) - \int_0^1 k(t, s)f(s) ds = 0$, while the inhomogeneous equation is $\lambda f(t) - \int_0^1 k(t, s)f(s) ds = b(t)$. Quite interestingly, this alternative also implies that if T is a compact operator on an infinite-dimensional Banach space, then $\sigma(T) = \sigma_p(T) \cup \{0\}$, where $\sigma_p(T)$ are the eigenvalues of T . This in turn can be used to prove the spectral theorem for compact self-adjoint operators. For details, see [38] and [28].

The spectrum can be defined in any Banach algebra, not just the algebra of operators. Many results generalize without difficulty to this setting. To speak about adjoints, one needs to add some structure to the Banach algebra (namely, an involution) and specify its relation with the norm. This gives rise to C^* -algebras. An important result in this theory is the Gelfand-Naimark theorem for commutative C^* -algebras. We shall not state it here to avoid too many definitions, but we mention that it can be used to prove a generalized spectral theorem, which yields Theorem 4.48 for normal operators as a special case. Another famous theorem in this theory is the Gelfand-Naimark-Segal (GNS) construction, which is perhaps more algebraic in nature, and states that every C^* -algebra can be regarded as a C^* -subalgebra of $B(\mathcal{H})$ for some Hilbert space \mathcal{H} . For these results, see [32] and [26].

The spectral theorem for bounded self-adjoint operators is hailed as “one of the most important mathematical achievements of all times” in [20]. It has another version which goes as follows: any self-adjoint $T \in B(\mathcal{H})$ is unitarily equivalent to a multiplication operator. More precisely, there is a measure space (X, μ) and $\varphi \in L^\infty(X)$ such that T and M_φ are unitarily equivalent. This theorem can be refined to give a complete classification

of self-adjoint operators up to unitary equivalence. The reader can find these results in [28] and [20].

The Borel functional calculus is a special feature of normal, in particular self-adjoint operators. However, we can define a functional calculus for any bounded operator, if we consider the smaller set of holomorphic functions. For this, see [32]. Finally, the Borel functional calculus and the spectral theorem can be extended to unbounded self-adjoint operators. Such an extension is really important in mathematical physics, since the Hamiltonian is an unbounded operator. This generalization is not exactly “proved similarly” for many reasons. One of them is that we used the compactness of $\sigma(T)$ to define the continuous functional calculus and then the Borel calculus. For unbounded operators, $\sigma(T)$ is no longer compact, so one proceeds differently. There are two ways to do this: either prove the spectral theorem for unbounded operators from scratch, or prove it using the spectral theorem for bounded operators. For the first approach, see [39], for the second one see [28] and [34]. Here is the result:

Theorem 4.49. *To every self-adjoint operator T on a Hilbert space \mathcal{H} , we can find a unique projection-valued measure P_T such that $T = \int_{\sigma(T)} \lambda dP_T(\lambda)$.*

If $\mu_x(E) := \langle P_T(E)x, x \rangle$, then we can define an operator $f(T)$ for any Borel function $f : \mathbb{R} \rightarrow \mathbb{C}$ by

$$D(f(T)) = \{x \in \mathcal{H} : f \in L^2(\mathbb{R}, \mu_x)\}, \quad f(T) = \int_{\sigma(T)} f(\lambda) dP_T(\lambda).$$

This operator has the following properties:

- (a) $\text{id}_{\mathbb{R}}(T) = T$.
- (b) For $x \in D(f(T))$, we have $\|f(T)x\|^2 = \int |f(\lambda)|^2 d\mu_x(\lambda)$.
- (c) $f(T)^* = \bar{f}(T)$. In particular, $f(T)$ is a normal operator, it is self-adjoint if f is real-valued, it is positive if $f \geq 0$, and it is unitary if $|f| = 1$.
- (d) If f and fg are bounded, then $(fg)(T) = f(T)g(T)$.
- (e) If $\lambda \in \mathbb{C} \setminus \mathbb{R}$, then $\frac{1}{T-\lambda} = R(\lambda)$.
- (f) If $Tx = \lambda x$, then $f(T)x = f(\lambda)x$.

4.8 Exercises

1. Let X, Y be Banach spaces, $T : X \rightarrow Y$. Show that the following assertions are equivalent:
 - (i) T is compact.
 - (ii) $\overline{T(X_1)}$ is compact. Here X_1 is the closed unit ball of X .
 - (iii) Any bounded sequence (x_n) in X has a subsequence (x_{n_k}) such that (Tx_{n_k}) converges.
2. Define $T : \ell^p \rightarrow \ell^p$ by $T : (x_j)_1^\infty \mapsto (x_j/j)_1^\infty$. Show that T is compact if
 - (i) $1 \leq p < \infty$,
 - (ii) $p = \infty$.
3. Suppose $A, B \in B(\mathcal{H})$ and AB is compact. Which statements are true?
 - (a) Both A and B are compact.
 - (b) At least A or B is compact.

4. Which of the following statements are true?
- There exists a compact operator with a closed image.
 - The image of any compact operator is closed.
 - The image of any compact operator is not closed.
 - There exists a compact operator with a non-closed image.
 - There exists a compact operator with a finite dimensional kernel.
 - The kernel of any compact operator is finite dimensional.
5. Which of the following operators $T : L^2[a, b] \rightarrow L^2[a, b]$ has a finite rank ?
- $(Tf)(t) = \sum_{j=1}^n \varphi_j(t) \int_a^b \psi_j(s) f(s) ds$, where $\varphi_j, \psi_j \in L^2[a, b]$.
 - $(Tf)(t) = \int_a^t f(s) ds$.
6. Consider the Hilbert space ℓ^2 and let $T : \ell^2 \rightarrow \ell^2$ be a linear operator defined by a matrix $(a_{i,j})_{i,j=1}^\infty$.
- Suppose $a_{i,j} = b_{i,j}c_{i,j}$, with $\sup_i \sum_j |b_{i,j}|^2 \leq C_1^2$ and $\sup_j \sum_i |c_{i,j}|^2 \leq C_2^2$. Show that T is bounded, with $\|T\| \leq C_1C_2$.
 - Suppose the $a_{i,j}$ satisfy $\sup_i \sum_j |a_{i,j}| \leq C_1^2$ and $\sup_j \sum_i |a_{i,j}| \leq C_2^2$. Show that T is bounded, with $\|T\| \leq C_1C_2$.
 - Suppose T is defined by the matrix

$$T = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots \\ a_2 & a_3 & \cdots & \\ a_3 & \cdots & & \\ \cdots & & & \end{pmatrix}$$

and suppose $\sum_{j=1}^\infty |a_j| < \infty$. Show that T is compact.

7. Let \mathcal{H} be a separable Hilbert space and $T \in B(\mathcal{H})$ a compact operator.
- Suppose T is self-adjoint. Show that there exist an orthonormal basis $\{e_n\}$ of \mathcal{H} and a sequence $\{\lambda_n\} \subset \mathbb{R}$ such that for each $x \in \mathcal{H}$, we have $x = \sum \langle x, e_k \rangle e_k$ and $Tx = \sum \lambda_k \langle x, e_k \rangle e_k$.
 - Show that even if T is not self-adjoint, there exist orthonormal sets $\{e_n\}$ and $\{f_n\}$ and a sequence $\{\mu_n\} \subset \mathbb{R}_+$ such that for each $x \in \mathcal{H}$, we have $x = \sum \langle x, e_k \rangle e_k$, and $Tx = \sum \mu_k \langle x, e_k \rangle f_k$.
*Hint: let $\{e_n\}$ correspond to T^*T as in (i), show that the corresponding λ_n must be positive, then take $\mu_n = \sqrt{\lambda_n}$.*
8. Let (X, μ) be a σ -finite measure space and let $K \in L^2(X \times X)$. Define the integral operator $T_K : L^2(\mu) \rightarrow L^2(\mu)$ by $(T_K f)(x) = \int_X K(x, y) f(y) d\mu(y)$. Show that $(T_K^* f)(x) = \int_X \overline{K(y, x)} f(y) d\mu(y)$.
9. Let $T_K : L^2[0, 1] \rightarrow L^2[0, 1]$ be the integral operator with kernel $K(x, y) = \min(x, y)$ for $0 \leq x, y \leq 1$.
- Show that T_K is a compact self-adjoint operator.
 - Suppose $T_K f = \lambda f$, where $0 \neq f \in L^2$ and $\lambda \in \mathbb{R}$. Differentiate this equation twice to obtain a differential equation for f . Obtain the boundary conditions for f and f' from the integral equations.
 - Multiply the differential equation of f by \bar{f} and integrate from 0 to 1. Deduce that $\lambda > 0$.

- (d) Find all the eigenvalues of T_K .
10. Let $T_K : L^2[0, 1] \rightarrow L^2[0, 1]$ be the integral operator with $K(x, y) = \chi_A(x, y)$, the characteristic function of $A = \{(x, y) : y \geq -x + 1\}$. Find all the eigenvalues of T_K .
11. Let $T : \ell^2 \rightarrow \ell^2$ be the right shift $T : (x_1, x_2, \dots) \mapsto (0, x_1, x_2, \dots)$. Show that 0 is not an eigenvalue of T , but $0 \in \sigma(T)$.
12. Let $T : \ell^2 \rightarrow \ell^2$ be defined by $T : (x_j)_1^\infty \mapsto (\alpha_j x_j)_1^\infty$, where (α_j) is dense in $[0, 1]$. Find the eigenvalues of T , and $\sigma(T)$.
13. Let $T : \ell^\infty \rightarrow \ell^\infty$ be the left shift $(x_1, x_2, \dots) \mapsto (x_2, x_3, \dots)$.
- (a) If $|\lambda| > 1$, show that $\lambda \in \rho(T)$.
- (b) If $|\lambda| \leq 1$, show that λ is an eigenvalue and find the eigenspace X_λ .
14. Let $T : \ell^p \rightarrow \ell^p$ be the left shift $(x_1, x_2, \dots) \mapsto (x_2, x_3, \dots)$. Suppose $1 \leq p < \infty$. If $|\lambda| = 1$, is λ an eigenvalue of T ?
15. Let T be a compact self-adjoint operator on a Hilbert space \mathcal{H} . Show that $\sigma(T) = \overline{\sigma_p(T)}$, where $\sigma_p(T)$ is the set of eigenvalues of T .
16. For the integral operator of Exercise 9, calculate
- (i) $\|T_K\|$ (*Hint: calculate $\|T_K\|_\sigma$*).
- (ii) $\|K\|_{L^2(X \times X)}$
- (iii) $\|T_K\|_2$ by two methods: first using (ii), next using the definition of the Hilbert-Schmidt norm, with an orthonormal basis of *eigenvectors*.
17. Same question as Exercise 16, for the integral operator of Exercise 10.
18. Let \mathcal{H} be a Hilbert space and let $T \in B(\mathcal{H})$ be self-adjoint. Show that $\sigma(T) \subseteq [m, M]$, where

$$m = \inf_{\|x\|=1} \langle Tx, x \rangle \quad \text{and} \quad M = \sup_{\|x\|=1} \langle Tx, x \rangle.$$

Hint: Show that if $\lambda = M + c$, $c > 0$, then $T - \lambda$ is bounded below. Argue similarly if $\lambda = m - c$.

19. Let \mathcal{H} be a Hilbert space.
- (i) Show that if $T \geq 0$, then $I + T$ is invertible.
- (ii) Show that if $T \in B(\mathcal{H})$, then $I + T^*T$ is invertible.
20. (McCarthy inequalities). Let T be a positive self-adjoint operator on \mathcal{H} and $x \in \mathcal{H}$. Show that
- (i) If $\alpha \geq 1$, then $\langle T^\alpha x, x \rangle \geq \langle Tx, x \rangle^\alpha \|x\|^{2-2\alpha}$.
- (ii) If $\alpha \in (0, 1]$, then $\langle T^\alpha x, x \rangle \leq \langle Tx, x \rangle^\alpha \|x\|^{2-2\alpha}$.
- Hint: for (i), apply the Hölder inequality to $\langle Tx, x \rangle = \int_{\sigma(T)} \lambda d\mu_x(\lambda)$. For (ii), apply (i) to T^α and α^{-1} .*

These inequalities remain true for unbounded operators, if $x \in D(T)$.

21. (Stone's formula). Let T be a self-adjoint operator. Show that for all $\psi \in \mathcal{H}$ we have

$$\lim_{\varepsilon \downarrow 0} \frac{1}{2\pi i} \int_a^b [(T - s - i\varepsilon)^{-1} \psi - (T - s + i\varepsilon)^{-1} \psi] ds = \frac{1}{2} [\chi_{[a,b]}(T)\psi + \chi_{(a,b)}(T)\psi].$$

Hint: let $f_\varepsilon(\lambda) = \frac{1}{2\pi i} \int_a^b \left(\frac{1}{\lambda - s - i\varepsilon} - \frac{1}{\lambda - s + i\varepsilon} \right) ds$. Show that

$$\lim_{\varepsilon \downarrow 0} f_\varepsilon(\lambda) = \begin{cases} 0 & \text{if } \lambda \notin [a, b], \\ \frac{1}{2} & \text{if } \lambda = a \text{ or } \lambda = b, \\ 1 & \text{if } \lambda \in (a, b). \end{cases}$$

Then use the fact that pointwise convergence of $f_\varepsilon(\lambda)$ implies strong convergence of $f_\varepsilon(T)$.

22. (One-parameter groups). Let X be a Banach space and suppose that to every $t \in \mathbb{R}$, there is an associated operator $Q(t)$ such that

- (i) $Q(0) = \text{id}$,
- (ii) $Q(s+t) = Q(s)Q(t)$ for all $s, t \in \mathbb{R}$,
- (iii) $\lim_{t \rightarrow 0} \|Q(t)x - x\| = 0$ for every $x \in X$.

Then we say that $\{Q(t)\}_{t \in \mathbb{R}}$ is a *strongly continuous one-parameter group*.

Let A be a self-adjoint operator on a Hilbert space \mathcal{H} . Show that the family $\{U(t)\}_{t \in \mathbb{R}}$ given by $U(t) = e^{-itA}$ is a strongly continuous one-parameter group of *unitary operators*.

Show moreover that the map $t \mapsto U(t)$ is strongly differentiable at $t = 0$, with derivative $-iA$. That is, show that $\lim_{t \rightarrow 0} \left\| \frac{U(t)\psi - \psi}{t} - (-iA\psi) \right\| = 0$ for any $\psi \in \mathcal{H}$.

Hint: for differentiability, calculate $\left\| \frac{U(t)\psi - \psi}{t} - (-iA\psi) \right\|^2$ using μ_ψ and use dominated convergence.

Bibliography

- [1] Robert A. Adams and John J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.
- [2] S. Attal. Lectures on Quantum Noise Theory. *author's homepage, Institut Camille Jordan*.
- [3] Rabi Bhattacharya and Edward C. Waymire. *A basic course in probability theory*. Universitext. Springer, New York, 2007.
- [4] Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.
- [5] N. L. Carothers. *A short course on Banach space theory*, volume 64 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2005.
- [6] J.-F. Colombeau. Multiplication of distributions. *Bull. Amer. Math. Soc. (N.S.)*, 23(2):251–268, 1990.
- [7] John B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.
- [8] Anton Deitmar. *A first course in harmonic analysis*. Universitext. Springer-Verlag, New York, second edition, 2005.
- [9] Nelson Dunford and Jacob T. Schwartz. *Linear Operators. I. General Theory*. With the assistance of W. G. Bade and R. G. Bartle. Pure and Applied Mathematics, Vol. 7. Interscience Publishers, Inc., New York; Interscience Publishers, Ltd., London, 1958.
- [10] Yuli Eidelman, Vitali Milman, and Antonis Tsolomitis. *Functional analysis*, volume 66 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2004. An introduction.
- [11] Gregory Eskin. *Lectures on linear partial differential equations*, volume 123 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2011.
- [12] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.
- [13] P. J. Fitzsimmons. Math 280c probability theory. *author's homepage, University of California*.
- [14] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.
- [15] T. Gantumur. Introduction to Distributions. *author's homepage, McGill University*.
- [16] Julio González-Díaz, Ignacio García-Jurado, and M. Gloria Fiestras-Janeiro. *An introductory course on mathematical game theory*, volume 115 of *Graduate Studies*

- in Mathematics*. American Mathematical Society, Providence, RI; Real Sociedad Matemática Española, Madrid, 2010.
- [17] A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo*, 8:1–79, 1953.
- [18] Alexandre Grothendieck. Produits tensoriels topologiques et espaces nucléaires. *Mem. Amer. Math. Soc.*, 1955(16):140, 1955.
- [19] Gerd Grubb. *Distributions and operators*, volume 252 of *Graduate Texts in Mathematics*. Springer, New York, 2009.
- [20] A. Ya. Helemskii. *Lectures and exercises on functional analysis*, volume 233 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2006. Translated from the 2004 Russian original by S. Akbarov.
- [21] Lars Hörmander. *The analysis of linear partial differential operators. I–IV*, volume 256 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1990.
- [22] Erwin Kreyszig. *Introductory functional analysis with applications*. John Wiley & Sons, New York-London-Sydney, 1978.
- [23] Giovanni Leoni. *A first course in Sobolev spaces*, volume 105 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2009.
- [24] Elliott H. Lieb and Michael Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2001.
- [25] Terry J. Morrison. *Functional analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York, 2001. An introduction to Banach space theory.
- [26] Gerard J. Murphy. *C^* -algebras and operator theory*. Academic Press, Inc., Boston, MA, 1990.
- [27] Jindřich Nečas. *Direct methods in the theory of elliptic equations*. Springer Monographs in Mathematics. Springer, Heidelberg, 2012. Translated from the 1967 French original by Gerard Tronel and Alois Kufner, Editorial coordination and preface by Šárka Nečasová and a contribution by Christian G. Simader.
- [28] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, second edition, 1980. Functional analysis.
- [29] J. R. Retherford. *Hilbert space: compact operators and the trace theorem*, volume 27 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1993.
- [30] H. L. Royden and P. M. Fitzpatrick. *Real analysis*. Pearson, fourth edition, 2010.
- [31] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
- [32] Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.
- [33] Xavier Saint Raymond. *Elementary introduction to the theory of pseudodifferential operators*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1991.
- [34] Konrad Schmüdgen. *Unbounded self-adjoint operators on Hilbert space*, volume 265 of *Graduate Texts in Mathematics*. Springer, Dordrecht, 2012.

-
- [35] Barry Simon. *Trace ideals and their applications*, volume 120 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, second edition, 2005.
- [36] Barry Simon. *Convexity*, volume 187 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2011. An analytic viewpoint.
- [37] François Trèves. *Topological vector spaces, distributions and kernels*. Academic Press, New York-London, 1967.
- [38] I. F. Vershynin. Lectures in Functional Analysis. *author's homepage, University of Michigan*.
- [39] Joachim Weidmann. *Linear operators in Hilbert spaces*, volume 68 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1980. Translated from the German by Joseph Szücs.
- [40] I. F. Wilde. Distribution Theory (Generalized Functions) Notes. *author's homepage, King's College London*.
- [41] I. F. Wilde. Functional Analysis (Topological Vectors Spaces version). *author's homepage, King's College London*.
- [42] Stephen Willard. *General topology*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1970.
- [43] M. W. Wong. *An introduction to pseudo-differential operators*. World Scientific Publishing Co., Inc., Teaneck, NJ, 1991.
- [44] Robert J. Zimmer. *Essential results of functional analysis*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1990.
- [45] Maciej Zworski. *Semiclassical analysis*, volume 138 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.