Topics in Real and Functional Analysis

Gerald Teschl

Gerald Teschl
Fakultät für Mathematik
Nordbergstraße 15
Universität Wien
1090 Wien, Austria

E-mail: Gerald.Teschl@univie.ac.at
URL: http://www.mat.univie.ac.at/~gerald/

2010 Mathematics subject classification. 46-01, 28-01, 46E30

Abstract. This manuscript provides a brief introduction to Real and Functional Analysis. It covers basic Hilbert and Banach space theory as well as basic measure theory including Lebesgue spaces and the Fourier transform.

Keywords and phrases. Functional Analysis, Banach space, Hilbert space, Measure theory, Lebesgue spaces, Fourier transform.

Typeset by $\mathcal{AMS}$-LaTeX and Makeindex.
Version: January 1, 2013

Contents

Preface vii

Part 1. Functional Analysis

Chapter 0. Introduction 3

§0.1. Linear partial differential equations 3

Chapter 1. A first look at Banach and Hilbert spaces 7

§1.1. Warm up: Metric and topological spaces 7

§1.2. The Banach space of continuous functions 17

§1.3. The geometry of Hilbert spaces 24

§1.4. Completeness 29

§1.5. Bounded operators 30

§1.6. Sums and quotients of Banach spaces 33

§1.7. Spaces of continuous and differentiable functions 34

Chapter 2. Hilbert spaces 39

§2.1. Orthonormal bases 39

§2.2. The projection theorem and the Riesz lemma 44

§2.3. Operators defined via forms 46

§2.4. Orthogonal sums and tensor products 48

Chapter 3. Compact operators 51

§3.1. Compact operators 51

§3.2. The spectral theorem for compact symmetric operators 53

§3.3. Applications to Sturm–Liouville operators 56

Chapter 4. The main theorems about Banach spaces 61

§4.1. The Baire theorem and its consequences 61

§4.2. The Hahn–Banach theorem and its consequences 66

§4.3. Weak convergence 72

Chapter 5. More on compact operators 79

§5.1. Canonical form of compact operators 79

§5.2. Hilbert–Schmidt and trace class operators 82

§5.3. Fredholm theory for compact operators 84

Chapter 6. Bounded linear operators 89

§6.1. Banach algebras 89

§6.2. The C^* algebra of operators and the spectral theorem 94

§6.3. Spectral measures 96

§6.4. The Stone–Weierstraß theorem 99

Part 2. Real Analysis

Chapter 7. Almost everything about Lebesgue integration 105

§7.1. Borel measures in a nut shell 105

§7.2. Extending a premeasure to a measure 112

§7.3. Measurable functions 116

§7.4. How wild are measurable objects 119

§7.5. Integration — Sum me up, Henri 121

§7.6. Product measures 126

§7.7. Transformation of measures and integrals 130

§7.8. Appendix: Transformation of Lebesgue–Stieltjes integrals 134

§7.9. Appendix: The connection with the Riemann integral 138

Chapter 8. The Lebesgue spaces L^p 143

§8.1. Functions almost everywhere 143

§8.2. Jensen $\leq$ Hölder $\leq$ Minkowski 145

§8.3. Nothing missing in L^p 148

§8.4. Integral operators 152

Chapter 9. More measure theory 155

§9.1. Decomposition of measures 155

§9.2. Derivatives of measures 158

§9.3. Complex measures 163
§9.4. Appendix: Functions of bounded variation and absolutely continuous functions 168

Chapter 10. The dual of L^p 175
§10.1. The dual of L^p, $p < \infty$ 175
§10.2. The dual of L^∞ and the Riesz representation theorem 177

Chapter 11. The Fourier transform 181
§11.1. The Fourier transform on L^2 181
§11.2. Interpolation and the Fourier transform on L^p 189
§11.3. Sobolev spaces 191

Appendix A. Zorn's lemma 195

Bibliography 197

Glossary of notation 199

Index 201

Preface

The present manuscript was written for my course *Functional Analysis* given at the University of Vienna in winter 2004 and 2009. It was adapted and extended for a course *Real Analysis* given in summer 2011. The two parts are to a large extent independent. In particular, the first part does not assume any knowledge from measure theory (at the expense of not mentioning L^p spaces).

It is updated whenever I find some errors. Hence you might want to make sure that you have the most recent version, which is available from

http://www.mat.univie.ac.at/~gerald/ftp/book-fa/

Please do not redistribute this file or put a copy on your personal webpage but link to the page above.

Acknowledgments

I wish to thank my students, Kerstin Ammann, Christian Ekstrand, Michael Fischer, Helge Krüger, Oliver Leingang, Alice Mikikits-Leitner, Caroline Moosmüller, Piotr Owczarek, Martina Pflegpeter, Tobias Preinerstorfer, Christian Schmid, Richard Welke, David Wimmesberger, Rudolf Zeidler, and colleagues, Heinz Hanßmann, Aleksey Kostenko, Wallace Lam, Daniel Lenz, Johanna Michor, and Alex Strohmaier, who have pointed out several typos and made useful suggestions for improvements.

Finally, no book is free of errors. So if you find one, or if you have comments or suggestions (no matter how small), please let me know.

Gerald Teschl

Vienna, Austria
February, 2011

Part 1

Functional Analysis

Chapter 0

Introduction

Functional analysis is an important tool in the investigation of all kind of problems in pure mathematics, physics, biology, economics, etc.. In fact, it is hard to find a branch in science where functional analysis is not used.

The main objects are (infinite dimensional) vector spaces with different concepts of convergence. The classical theory focuses on linear operators (i.e., functions) between these spaces but nonlinear operators are of course equally important. However, since one of the most important tools in investigating nonlinear mappings is linearization (differentiation), linear functional analysis will be our first topic in any case.

0.1. Linear partial differential equations

Rather than overwhelming you with a vast number of classical examples I want to focus on one: linear partial differential equations. We will use this example as a guide throughout this first chapter and will develop all necessary tools for a successful treatment of our particular problem.

In his investigation of heat conduction Fourier was lead to the (one dimensional) **heat** or diffusion equation

$$\frac{\partial}{\partial t} u(t,x) = \frac{\partial^2}{\partial x^2} u(t,x). \tag{0.1}$$

Here $u(t,x)$ is the temperature distribution at time t at the point x. It is usually assumed, that the temperature at $x = 0$ and $x = 1$ is fixed, say $u(t,0) = a$ and $u(t,1) = b$. By considering $u(t,x) \to u(t,x) - a - (b-a)x$ it is clearly no restriction to assume $a = b = 0$. Moreover, the initial temperature distribution $u(0,x) = u_0(x)$ is assumed to be known as well.

Since finding the solution seems at first sight not possible, we could try to find at least some solutions of (0.1) first. We could for example make an ansatz for $u(t, x)$ as a product of two functions, each of which depends on only one variable, that is,

$$u(t, x) = w(t)y(x). \tag{0.2}$$

This ansatz is called **separation of variables**. Plugging everything into the heat equation and bringing all t, x dependent terms to the left, right side, respectively, we obtain

$$\frac{\dot{w}(t)}{w(t)} = \frac{y''(x)}{y(x)}. \tag{0.3}$$

Here the dot refers to differentiation with respect to t and the prime to differentiation with respect to x.

Now if this equation should hold for all t and x, the quotients must be equal to a constant $-\lambda$ (we choose $-\lambda$ instead of λ for convenience later on). That is, we are lead to the equations

$$-\dot{w}(t) = \lambda w(t) \tag{0.4}$$

and

$$-y''(x) = \lambda y(x), \qquad y(0) = y(1) = 0, \tag{0.5}$$

which can easily be solved. The first one gives

$$w(t) = c_1 \mathrm{e}^{-\lambda t} \tag{0.6}$$

and the second one

$$y(x) = c_2 \cos(\sqrt{\lambda} x) + c_3 \sin(\sqrt{\lambda} x). \tag{0.7}$$

However, $y(x)$ must also satisfy the boundary conditions $y(0) = y(1) = 0$. The first one $y(0) = 0$ is satisfied if $c_2 = 0$ and the second one yields (c_3 can be absorbed by $w(t)$)

$$\sin(\sqrt{\lambda}) = 0, \tag{0.8}$$

which holds if $\lambda = (\pi n)^2$, $n \in \mathbb{N}$. In summary, we obtain the solutions

$$u_n(t, x) = c_n \mathrm{e}^{-(\pi n)^2 t} \sin(n\pi x), \qquad n \in \mathbb{N}. \tag{0.9}$$

So we have found a large number of solutions, but we still have not dealt with our initial condition $u(0, x) = u_0(x)$. This can be done using the superposition principle which holds since our equation is linear. Hence any finite linear combination of the above solutions will be again a solution. Moreover, under suitable conditions on the coefficients we can even consider infinite linear combinations. In fact, choosing

$$u(t, x) = \sum_{n=1}^{\infty} c_n \mathrm{e}^{-(\pi n)^2 t} \sin(n\pi x), \tag{0.10}$$

where the coefficients c_n decay sufficiently fast, we obtain further solutions of our equation. Moreover, these solutions satisfy

$$u(0,x) = \sum_{n=1}^{\infty} c_n \sin(n\pi x) \tag{0.11}$$

and expanding the initial conditions into a Fourier series

$$u_0(x) = \sum_{n=1}^{\infty} u_{0,n} \sin(n\pi x), \tag{0.12}$$

we see that the solution of our original problem is given by (0.10) if we choose $c_n = u_{0,n}$.

Of course for this last statement to hold we need to ensure that the series in (0.10) converges and that we can interchange summation and differentiation. You are asked to do so in Problem 0.1.

In fact many equations in physics can be solved in a similar way:

- **Reaction-Diffusion equation**:

$$\begin{aligned} &\frac{\partial}{\partial t}u(t,x) - \frac{\partial^2}{\partial x^2}u(t,x) + q(x)u(t,x) = 0,\\ &u(0,x) = u_0(x),\\ &u(t,0) = u(t,1) = 0. \end{aligned} \tag{0.13}$$

Here $u(t,x)$ could be the density of some gas in a pipe and $q(x) > 0$ describes that a certain amount per time is removed (e.g., by a chemical reaction).

- **Wave equation**:

$$\begin{aligned} &\frac{\partial^2}{\partial t^2}u(t,x) - \frac{\partial^2}{\partial x^2}u(t,x) = 0,\\ &u(0,x) = u_0(x), \quad \frac{\partial u}{\partial t}(0,x) = v_0(x)\\ &u(t,0) = u(t,1) = 0. \end{aligned} \tag{0.14}$$

Here $u(t,x)$ is the displacement of a vibrating string which is fixed at $x = 0$ and $x = 1$. Since the equation is of second order in time, both the initial displacement $u_0(x)$ and the initial velocity $v_0(x)$ of the string need to be known.

- **Schrödinger equation**:

$$\begin{aligned} &\mathrm{i}\frac{\partial}{\partial t}u(t,x) = -\frac{\partial^2}{\partial x^2}u(t,x) + q(x)u(t,x),\\ &u(0,x) = u_0(x),\\ &u(t,0) = u(t,1) = 0. \end{aligned} \tag{0.15}$$

Here $|u(t,x)|^2$ is the probability distribution of a particle trapped in a box $x \in [0,1]$ and $q(x)$ is a given external potential which describes the forces acting on the particle.

All these problems (and many others) lead to the investigation of the following problem

$$Ly(x) = \lambda y(x), \qquad L = -\frac{d^2}{dx^2} + q(x), \tag{0.16}$$

subject to the **boundary conditions**

$$y(a) = y(b) = 0. \tag{0.17}$$

Such a problem is called a **Sturm–Liouville boundary value problem**. Our example shows that we should prove the following facts about our Sturm–Liouville problems:

(i) The Sturm–Liouville problem has a countable number of eigenvalues E_n with corresponding eigenfunctions $u_n(x)$, that is, $u_n(x)$ satisfies the boundary conditions and $Lu_n(x) = E_n u_n(x)$.

(ii) The eigenfunctions u_n are complete, that is, any *nice* function $u(x)$ can be expanded into a generalized Fourier series

$$u(x) = \sum_{n=1}^{\infty} c_n u_n(x).$$

This problem is very similar to the eigenvalue problem of a matrix and we are looking for a generalization of the well-known fact that every symmetric matrix has an orthonormal basis of eigenvectors. However, our linear operator L is now acting on some space of functions which is not finite dimensional and it is not at all clear what even orthogonal should mean for functions. Moreover, since we need to handle infinite series, we need convergence and hence define the distance of two functions as well.

Hence our program looks as follows:

- What is the distance of two functions? This automatically leads us to the problem of convergence and completeness.
- If we additionally require the concept of orthogonality, we are lead to Hilbert spaces which are the proper setting for our eigenvalue problem.
- Finally, the spectral theorem for compact symmetric operators will be the solution of our above problem

Problem 0.1. *Find conditions for the initial distribution $u_0(x)$ such that (0.10) is indeed a solution (i.e., such that interchanging the order of summation and differentiation is admissible). (Hint: What is the connection between smoothness of a function and decay of its Fourier coefficients?)*

Chapter 1

A first look at Banach and Hilbert spaces

1.1. Warm up: Metric and topological spaces

Before we begin, I want to recall some basic facts from metric and topological spaces. I presume that you are familiar with these topics from your calculus course. As a general reference I can warmly recommend Kelly's classical book [**4**].

A **metric space** is a space X together with a distance function $d : X \times X \to \mathbb{R}$ such that

(i) $d(x, y) \geq 0$,

(ii) $d(x, y) = 0$ if and only if $x = y$,

(iii) $d(x, y) = d(y, x)$,

(iv) $d(x, z) \leq d(x, y) + d(y, z)$ (**triangle inequality**).

If (ii) does not hold, d is called a **pseudometric**. Moreover, it is straightforward to see the **inverse triangle inequality** (Problem 1.1)

$$|d(x, y) - d(z, y)| \leq d(x, z). \tag{1.1}$$

Example. Euclidean space $\mathbb{R}^n$ together with $d(x, y) = (\sum_{k=1}^n (x_k - y_k)^2)^{1/2}$ is a metric space and so is $\mathbb{C}^n$ together with $d(x, y) = (\sum_{k=1}^n |x_k - y_k|^2)^{1/2}$. ⋄

The set

$$B_r(x) = \{y \in X | d(x, y) < r\} \tag{1.2}$$

is called an **open ball** around x with radius $r > 0$. A point x of some set U is called an **interior point** of U if U contains some ball around x. If x

is an interior point of U, then U is also called a **neighborhood** of x. A point x is called a **limit point** of U (also **accumulation** or **cluster point**) if $(B_r(x)\backslash\{x\}) \cap U \neq \emptyset$ for every ball around x. Note that a limit point x need not lie in U, but U must contain points arbitrarily close to x. A point x is called an **isolated point** of U if there exists a neighborhood of x not containing any other points of U. A set which consists only of isolated points is called a **discrete set**.

Example. Consider $\mathbb{R}$ with the usual metric and let $U = (-1, 1)$. Then every point $x \in U$ is an interior point of U. The points ± 1 are limit points of U. $\diamond$

A set consisting only of interior points is called **open**. The family of open sets $\mathcal{O}$ satisfies the properties

(i) $\emptyset, X \in \mathcal{O}$,

(ii) $O_1, O_2 \in \mathcal{O}$ implies $O_1 \cap O_2 \in \mathcal{O}$,

(iii) $\{O_\alpha\} \subseteq \mathcal{O}$ implies $\bigcup_\alpha O_\alpha \in \mathcal{O}$.

That is, $\mathcal{O}$ is closed under finite intersections and arbitrary unions.

In general, a space X together with a family of sets $\mathcal{O}$, the open sets, satisfying (i)–(iii) is called a **topological space**. The notions of interior point, limit point, and neighborhood carry over to topological spaces if we replace open ball by open set.

There are usually different choices for the topology. Two not too interesting examples are the **trivial topology** $\mathcal{O} = \{\emptyset, X\}$ and the **discrete topology** $\mathcal{O} = \mathfrak{P}(X)$ (the powerset of X). Given two topologies $\mathcal{O}_1$ and $\mathcal{O}_2$ on X, $\mathcal{O}_1$ is called **weaker** (or **coarser**) than $\mathcal{O}_2$ if and only if $\mathcal{O}_1 \subseteq \mathcal{O}_2$.

Example. Note that different metrics can give rise to the same topology. For example, we can equip $\mathbb{R}^n$ (or $\mathbb{C}^n$) with the Euclidean distance $d(x, y)$ as before or we could also use

$$\tilde{d}(x, y) = \sum_{k=1}^{n} |x_k - y_k|. \tag{1.3}$$

Then

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} |x_k| \leq \sqrt{\sum_{k=1}^{n} |x_k|^2} \leq \sum_{k=1}^{n} |x_k| \tag{1.4}$$

shows $B_{r/\sqrt{n}}(x) \subseteq \tilde{B}_r(x) \subseteq B_r(x)$, where B, $\tilde{B}$ are balls computed using d, $\tilde{d}$, respectively. $\diamond$

Example. We can always replace a metric d by the bounded metric

$$\tilde{d}(x,y) = \frac{d(x,y)}{1+d(x,y)} \tag{1.5}$$

without changing the topology (since the open balls do not change). ⋄

Every subspace Y of a topological space X becomes a topological space of its own if we call $O \subseteq Y$ open if there is some open set $\tilde{O} \subseteq X$ such that $O = \tilde{O} \cap Y$. This natural topology $\mathcal{O} \cap Y$ is known as the **relative topology** (also **subspace**, **trace** or **induced topology**).

Example. The set $(0,1] \subseteq \mathbb{R}$ is not open in the topology of $X = \mathbb{R}$, but it is open in the relative topology when considered as a subset of $Y = [-1,1]$. ⋄

A family of open sets $\mathcal{B} \subseteq \mathcal{O}$ is called a **base** for the topology if for each x and each neighborhood $U(x)$, there is some set $O \in \mathcal{B}$ with $x \in O \subseteq U(x)$. Since an open set O is a neighborhood of every one of its points, it can be written as $O = \bigcup_{O \supseteq \tilde{O} \in \mathcal{B}} \tilde{O}$ and we have

Lemma 1.1. *If $\mathcal{B} \subseteq \mathcal{O}$ is a base for the topology, then every open set can be written as a union of elements from $\mathcal{B}$.*

If there exists a countable base, then X is called **second countable**.

Example. By construction the open balls $B_{1/n}(x)$ are a base for the topology in a metric space. In the case of $\mathbb{R}^n$ (or $\mathbb{C}^n$) it even suffices to take balls with rational center and hence $\mathbb{R}^n$ (as well as $\mathbb{C}^n$) is second countable. ⋄

A topological space is called a **Hausdorff space** if for two different points there are always two disjoint neighborhoods.

Example. Any metric space is a Hausdorff space: Given two different points x and y, the balls $B_{d/2}(x)$ and $B_{d/2}(y)$, where $d = d(x,y) > 0$, are disjoint neighborhoods (a pseudometric space will not be Hausdorff). ⋄

The complement of an open set is called a **closed set**. It follows from de Morgan's rules that the family of closed sets $\mathcal{C}$ satisfies

(i) $\emptyset, X \in \mathcal{C}$,

(ii) $C_1, C_2 \in \mathcal{C}$ implies $C_1 \cup C_2 \in \mathcal{C}$,

(iii) $\{C_\alpha\} \subseteq \mathcal{C}$ implies $\bigcap_\alpha C_\alpha \in \mathcal{C}$.

That is, closed sets are closed under finite unions and arbitrary intersections.

The smallest closed set containing a given set U is called the **closure**

$$\overline{U} = \bigcap_{C \in \mathcal{C}, U \subseteq C} C, \tag{1.6}$$

and the largest open set contained in a given set U is called the **interior**

$$U^\circ = \bigcup_{O \in \mathcal{O}, O \subseteq U} O. \tag{1.7}$$

It is not hard to see that the closure satisfies the following axioms (**Kuratowski closure axioms**):

(i) $\overline{\emptyset} = \emptyset$,
(ii) $U \subset \overline{U}$,
(iii) $\overline{\overline{U}} = \overline{U}$,
(iv) $\overline{U \cup V} = \overline{U} \cup \overline{V}$.

In fact, one can show that they can equivalently be used to define the topology by observing that the closed sets are precisely those which satisfy $\overline{A} = A$.

We can define interior and limit points as before by replacing the word ball by open set. Then it is straightforward to check

Lemma 1.2. *Let X be a topological space. Then the interior of U is the set of all interior points of U and the closure of U is the union of U with all limit points of U.*

A sequence $(x_n)_{n=1}^\infty \subseteq X$ is said to **converge** to some point $x \in X$ if $d(x, x_n) \to 0$. We write $\lim_{n\to\infty} x_n = x$ as usual in this case. Clearly the limit is unique if it exists (this is not true for a pseudometric).

Every convergent sequence is a **Cauchy sequence**; that is, for every $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that

$$d(x_n, x_m) \le \varepsilon, \qquad n, m \ge N. \tag{1.8}$$

If the converse is also true, that is, if every Cauchy sequence has a limit, then X is called **complete**.

Example. Both $\mathbb{R}^n$ and $\mathbb{C}^n$ are complete metric spaces. ⋄

Note that in a metric space x is a limit point of U if and only if there exists a sequence $(x_n)_{n=1}^\infty \subseteq U\backslash\{x\}$ with $\lim_{n\to\infty} x_n = x$. Hence U is closed if and only if for every convergent sequence the limit is in U. In particular,

Lemma 1.3. *A closed subset of a complete metric space is again a complete metric space.*

Note that convergence can also be equivalently formulated in terms of topological terms: A sequence x_n converges to x if and only if for every neighborhood U of x there is some $N \in \mathbb{N}$ such that $x_n \in U$ for $n \ge N$. In a Hausdorff space the limit is unique.

A set U is called **dense** if its closure is all of X, that is, if $\overline{U} = X$. A metric space is called **separable** if it contains a countable dense set.

Lemma 1.4. *A metric space is separable if and only if it is second countable as a topological space.*

Proof. From every dense set we get a countable base by considering open balls with rational radii and centers in the dense set. Conversely, from every countable base we obtain a dense set by choosing an element from each element of the base. □

Lemma 1.5. *Let X be a separable metric space. Every subset Y of X is again separable.*

Proof. Let $A = \{x_n\}_{n\in\mathbb{N}}$ be a dense set in X. The only problem is that $A \cap Y$ might contain no elements at all. However, some elements of A must be at least arbitrarily close: Let $J \subseteq \mathbb{N}^2$ be the set of all pairs (n, m) for which $B_{1/m}(x_n) \cap Y \neq \emptyset$ and choose some $y_{n,m} \in B_{1/m}(x_n) \cap Y$ for all $(n, m) \in J$. Then $B = \{y_{n,m}\}_{(n,m)\in J} \subseteq Y$ is countable. To see that B is dense, choose $y \in Y$. Then there is some sequence x_{n_k} with $d(x_{n_k}, y) < 1/k$. Hence $(n_k, k) \in J$ and $d(y_{n_k,k}, y) \leq d(y_{n_k,k}, x_{n_k}) + d(x_{n_k}, y) \leq 2/k \to 0$. □

Next we come to functions $f : X \to Y$, $x \mapsto f(x)$. We use the usual conventions $f(U) = \{f(x)|x \in U\}$ for $U \subseteq X$ and $f^{-1}(V) = \{x|f(x) \in V\}$ for $V \subseteq Y$. The set $\mathrm{Ran}(f) = f(X)$ is called the **range** of f and X is called the **domain** of f. A function f is called **injective** if for each $y \in Y$ there is at most one $x \in X$ with $f(x) = y$ (i.e., $f^{-1}(\{y\})$ contains at most one point) and **surjective** or **onto** if $\mathrm{Ran}(f) = Y$. A function f which is both injective and surjective is called **bijective**.

A function f between metric spaces X and Y is called continuous at a point $x \in X$ if for every $\varepsilon > 0$ we can find a $\delta > 0$ such that

$$d_Y(f(x), f(y)) \leq \varepsilon \quad \text{if} \quad d_X(x, y) < \delta. \tag{1.9}$$

If f is continuous at every point, it is called **continuous**.

Lemma 1.6. *Let X be a metric space. The following are equivalent:*

(i) *f is continuous at x (i.e, (1.9) holds).*

(ii) *$f(x_n) \to f(x)$ whenever $x_n \to x$.*

(iii) *For every neighborhood V of $f(x)$, $f^{-1}(V)$ is a neighborhood of x.*

Proof. (i) ⇒ (ii) is obvious. (ii) ⇒ (iii): If (iii) does not hold, there is a neighborhood V of $f(x)$ such that $B_\delta(x) \not\subseteq f^{-1}(V)$ for every δ. Hence we can choose a sequence $x_n \in B_{1/n}(x)$ such that $f(x_n) \notin f^{-1}(V)$. Thus $x_n \to x$ but $f(x_n) \not\to f(x)$. (iii) ⇒ (i): Choose $V = B_\varepsilon(f(x))$ and observe that by (iii), $B_\delta(x) \subseteq f^{-1}(V)$ for some δ. □

The last item implies that f is continuous if and only if the inverse image of every open set is again open (equivalently, the inverse image of every closed set is closed). If the image of every open set is open, then f is called **open**. A bijection f is called a **homeomorphism** if both f and its inverse f^{-1} are continuous. Note that if f is a bijection, then f^{-1} is continuous if and only if f is open.

In a topological space, (iii) is used as the definition for continuity. However, in general (ii) and (iii) will no longer be equivalent unless one uses generalized sequences, so-called nets, where the index set $\mathbb{N}$ is replaced by arbitrary directed sets.

The **support** of a function $f : X \to \mathbb{C}^n$ is the closure of all points x for which $f(x)$ does not vanish; that is,

$$\operatorname{supp}(f) = \overline{\{x \in X | f(x) \neq 0\}}. \tag{1.10}$$

If X and Y are metric spaces, then $X \times Y$ together with

$$d((x_1, y_1), (x_2, y_2)) = d_X(x_1, x_2) + d_Y(y_1, y_2) \tag{1.11}$$

is a metric space. A sequence (x_n, y_n) converges to (x, y) if and only if $x_n \to x$ and $y_n \to y$. In particular, the projections onto the first $(x, y) \mapsto x$, respectively, onto the second $(x, y) \mapsto y$, coordinate are continuous. Moreover, if X and Y are complete, so is $X \times Y$.

In particular, by the inverse triangle inequality (1.1),

$$|d(x_n, y_n) - d(x, y)| \leq d(x_n, x) + d(y_n, y), \tag{1.12}$$

we see that $d : X \times X \to \mathbb{R}$ is continuous.

Example. If we consider $\mathbb{R} \times \mathbb{R}$, we do not get the Euclidean distance of $\mathbb{R}^2$ unless we modify (1.11) as follows:

$$\tilde{d}((x_1, y_1), (x_2, y_2)) = \sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}. \tag{1.13}$$

As noted in our previous example, the topology (and thus also convergence/continuity) is independent of this choice. ⋄

If X and Y are just topological spaces, the **product topology** is defined by calling $O \subseteq X \times Y$ open if for every point $(x, y) \in O$ there are open neighborhoods U of x and V of y such that $U \times V \subseteq O$. In other words, the products of open sets from a basis of the product topology. In the case of metric spaces this clearly agrees with the topology defined via the product metric (1.11).

A **cover** of a set $Y \subseteq X$ is a family of sets $\{U_\alpha\}$ such that $Y \subseteq \bigcup_\alpha U_\alpha$. A cover is called open if all U_α are open. Any subset of $\{U_\alpha\}$ which still covers Y is called a **subcover**.

Lemma 1.7 (Lindelöf)**.** *If X is second countable, then every open cover has a countable subcover.*

Proof. Let $\{U_\alpha\}$ be an open cover for Y and let $\mathcal{B}$ be a countable base. Since every U_α can be written as a union of elements from $\mathcal{B}$, the set of all $B \in \mathcal{B}$ which satisfy $B \subseteq U_\alpha$ for some α form a countable open cover for Y. Moreover, for every B_n in this set we can find an α_n such that $B_n \subseteq U_{\alpha_n}$. By construction $\{U_{\alpha_n}\}$ is a countable subcover. □

A subset $K \subset X$ is called **compact** if every open cover has a finite subcover.

Lemma 1.8. *A topological space is compact if and only if it has the* **finite intersection property***: The intersection of a family of closed sets is empty if and only if the intersection of some finite subfamily is empty.*

Proof. By taking complements, to every family of open sets there is a corresponding family of closed sets and vice versa. Moreover, the open sets are a cover if and only if the corresponding closed sets have empty intersection. □

Lemma 1.9. *Let X be a topological space.*

(i) *The continuous image of a compact set is compact.*

(ii) *Every closed subset of a compact set is compact.*

(iii) *If X is Hausdorff, every compact set is closed.*

(iv) *The product of finitely many compact sets is compact.*

(v) *The finite union of compact sets is again compact.*

(vi) *If X is Hausdorff, any intersection of compact sets is again compact.*

Proof. (i) Observe that if $\{O_\alpha\}$ is an open cover for $f(Y)$, then $\{f^{-1}(O_\alpha)\}$ is one for Y.

(ii) Let $\{O_\alpha\}$ be an open cover for the closed subset Y (in the induced topology). Then there are open sets $\tilde{O}_\alpha$ with $O_\alpha = \tilde{O}_\alpha \cap Y$ and $\{\tilde{O}_\alpha\} \cup \{X \backslash Y\}$ is an open cover for X which has a finite subcover. This subcover induces a finite subcover for Y.

(iii) Let $Y \subseteq X$ be compact. We show that $X \backslash Y$ is open. Fix $x \in X \backslash Y$ (if $Y = X$, there is nothing to do). By the definition of Hausdorff, for every $y \in Y$ there are disjoint neighborhoods $V(y)$ of y and $U_y(x)$ of x. By compactness of Y, there are $y_1, \dots, y_n$ such that the $V(y_j)$ cover Y. But then $U(x) = \bigcap_{j=1}^{n} U_{y_j}(x)$ is a neighborhood of x which does not intersect Y.

(iv) Let $\{O_\alpha\}$ be an open cover for $X \times Y$. For every $(x, y) \in X \times Y$ there is some $\alpha(x, y)$ such that $(x, y) \in O_{\alpha(x,y)}$. By definition of the product topology there is some open rectangle $U(x, y) \times V(x, y) \subseteq O_{\alpha(x,y)}$. Hence for fixed x, $\{V(x, y)\}_{y \in Y}$ is an open cover of Y. Hence there are finitely many points $y_k(x)$ such that the $V(x, y_k(x))$ cover Y. Set $U(x) = \bigcap_k U(x, y_k(x))$. Since finite intersections of open sets are open, $\{U(x)\}_{x \in X}$ is an open cover and there are finitely many points x_j such that the $U(x_j)$ cover X. By construction, the $U(x_j) \times V(x_j, y_k(x_j)) \subseteq O_{\alpha(x_j, y_k(x_j))}$ cover $X \times Y$.

(v) Note that a cover of the union is a cover for each individual set and the union of the individual subcovers is the subcover we are looking for.

(vi) Follows from (ii) and (iii) since an intersection of closed sets is closed. □

As a consequence we obtain a simple criterion when a continuous function is a homeomorphism.

Corollary 1.10. *Let X and Y be topological spaces with X compact and Y Hausdorff. Then every continuous bijection $f : X \to Y$ is a homeomorphism.*

Proof. It suffices to show that f maps closed sets to closed sets. By (ii) every closed set is compact, by (i) its image is also compact, and by (iii) also closed. □

A subset $K \subset X$ is called **sequentially compact** if every sequence has a convergent subsequence. In a metric space compact and sequentially compact are equivalent.

Lemma 1.11. *Let X be a metric space. Then a subset is compact if and only if it is sequentially compact.*

Proof. Suppose X is compact and let x_n be a sequence which has no convergent subsequence. Then $K = \{x_n\}$ has no limit points and is hence compact by Lemma 1.9 (ii). For every n there is a ball $B_{\varepsilon_n}(x_n)$ which contains only finitely many elements of K. However, finitely many suffice to cover K, a contradiction.

Conversely, suppose X is sequentially compact and let $\{O_\alpha\}$ be some open cover which has no finite subcover. For every $x \in X$ we can choose some $\alpha(x)$ such that if $B_r(x)$ is the largest ball contained in $O_{\alpha(x)}$, then either $r \geq 1$ or there is no β with $B_{2r}(x) \subset O_\beta$ (show that this is possible). Now choose a sequence x_n such that $x_n \notin \bigcup_{m<n} O_{\alpha(x_m)}$. Note that by construction the distance $d = d(x_m, x_n)$ to every successor of x_m is either larger than 1 or the ball $B_{2d}(x_m)$ will not fit into any of the O_α.

Now let y be the limit of some convergent subsequence and fix some $r \in (0,1)$ such that $B_r(y) \subseteq O_{\alpha(y)}$. Then this subsequence must eventually be in $B_{r/5}(y)$, but this is impossible since if $d = d(x_{n_1}, x_{n_2})$ is the distance between two consecutive elements of this subsequence, then $B_{2d}(x_{n_1})$ cannot fit into $O_{\alpha(y)}$ by construction whereas on the other hand $B_{2d}(x_{n_1}) \subseteq B_{4r/5}(a) \subseteq O_{\alpha(y)}$. □

In a metric space, a set is called **bounded** if it is contained inside some ball. Note that compact sets are always bounded (show this!). In $\mathbb{R}^n$ (or $\mathbb{C}^n$) the converse also holds.

Theorem 1.12 (Heine–Borel)**.** *In $\mathbb{R}^n$ (or $\mathbb{C}^n$) a set is compact if and only if it is bounded and closed.*

Proof. By Lemma 1.9 (ii) and (iii) it suffices to show that a closed interval in $I \subseteq \mathbb{R}$ is compact. Moreover, by Lemma 1.11 it suffices to show that every sequence in $I = [a, b]$ has a convergent subsequence. Let x_n be our sequence and divide $I = [a, \frac{a+b}{2}] \cup [\frac{a+b}{2}, b]$. Then at least one of these two intervals, call it I_1, contains infinitely many elements of our sequence. Let $y_1 = x_{n_1}$ be the first one. Subdivide I_1 and pick $y_2 = x_{n_2}$, with $n_2 > n_1$ as before. Proceeding like this, we obtain a Cauchy sequence y_n (note that by construction $I_{n+1} \subseteq I_n$ and hence $|y_n - y_m| \le \frac{b-a}{n}$ for $m \ge n$). □

By Lemma 1.11 this is equivalent to

Theorem 1.13 (Bolzano–Weierstraß)**.** *Every bounded infinite subset of $\mathbb{R}^n$ (or $\mathbb{C}^n$) has at least one limit point.*

Combining Theorem 1.12 with Lemma 1.9 (i) we also obtain the **extreme value theorem**.

Theorem 1.14 (Weierstraß)**.** *Let K be compact. Every continuous function $f : K \to \mathbb{R}$ attains its maximum and minimum.*

A topological space is called **locally compact** if every point has a compact neighborhood.

Example. $\mathbb{R}^n$ is locally compact. ⋄

The **distance** between a point $x \in X$ and a subset $Y \subseteq X$ is

$$\mathrm{dist}(x, Y) = \inf_{y \in Y} d(x, y). \tag{1.14}$$

Note that x is a limit point of Y if and only if $\mathrm{dist}(x, Y) = 0$.

Lemma 1.15. *Let X be a metric space. Then*

$$|\,\mathrm{dist}(x, Y) - \mathrm{dist}(z, Y)| \le d(x, z). \tag{1.15}$$

In particular, $x \mapsto \mathrm{dist}(x, Y)$ is continuous.

Proof. Taking the infimum in the triangle inequality $d(x,y) \le d(x,z) + d(z,y)$ shows $\mathrm{dist}(x,Y) \le d(x,z) + \mathrm{dist}(z,Y)$. Hence $\mathrm{dist}(x,Y) - \mathrm{dist}(z,Y) \le d(x,z)$. Interchanging x and z shows $\mathrm{dist}(z,Y) - \mathrm{dist}(x,Y) \le d(x,z)$. □

Lemma 1.16 (Urysohn). *Suppose C_1 and C_2 are disjoint closed subsets of a metric space X. Then there is a continuous function $f : X \to [0,1]$ such that f is zero on C_2 and one on C_1.*

If X is locally compact and C_1 is compact, one can choose f with compact support.

Proof. To prove the first claim, set $f(x) = \frac{\mathrm{dist}(x,C_2)}{\mathrm{dist}(x,C_1)+\mathrm{dist}(x,C_2)}$. For the second claim, observe that there is an open set O such that $\overline{O}$ is compact and $C_1 \subset O \subset \overline{O} \subset X \backslash C_2$. In fact, for every $x \in C_1$, there is a ball $B_\varepsilon(x)$ such that $\overline{B_\varepsilon(x)}$ is compact and $\overline{B_\varepsilon(x)} \subset X \backslash C_2$. Since C_1 is compact, finitely many of them cover C_1 and we can choose the union of those balls to be O. Now replace C_2 by $X \backslash O$. □

Note that Urysohn's lemma implies that a metric space is **normal**; that is, for any two disjoint closed sets C_1 and C_2, there are disjoint open sets O_1 and O_2 such that $C_j \subseteq O_j$, $j = 1, 2$. In fact, choose f as in Urysohn's lemma and set $O_1 = f^{-1}([0,1/2))$, respectively, $O_2 = f^{-1}((1/2,1])$.

Lemma 1.17. *Let X be a locally compact metric space. Suppose K is a compact set and $\{O_j\}_{j=1}^n$ an open cover. Then there is a* **partition of unity** *for K subordinate to this cover; that is, there are continuous functions $h_j : X \to [0,1]$ such that h_j has compact support contained in O_j and*

$$\sum_{j=1}^{n} h_j(x) \le 1 \tag{1.16}$$

with equality for $x \in K$.

Proof. For every $x \in K$ there is some ε and some j such that $\overline{B_\varepsilon(x)} \subseteq O_j$. By compactness of K, finitely many of these balls cover K. Let K_j be the union of those balls which lie inside O_j. By Urysohn's lemma there are continuous functions $g_j : X \to [0,1]$ such that $g_j = 1$ on K_j and $g_j = 0$ on $X \backslash O_j$. Now set

$$h_j = g_j \prod_{k=1}^{j-1} (1 - g_k).$$

Then $h_j : X \to [0,1]$ has compact support contained in O_j and

$$\sum_{j=1}^{n} h_j(x) = 1 - \prod_{j=1}^{n} (1 - g_j(x))$$

shows that the sum is one for $x \in K$, since $x \in K_j$ for some j implies $g_j(x) = 1$ and causes the product to vanish. □

Problem 1.1. *Show that* $|d(x,y) - d(z,y)| \le d(x,z)$.

Problem 1.2. *Show the* **quadrangle inequality** $|d(x,y) - d(x',y')| \le d(x,x') + d(y,y')$.

Problem 1.3. *Show that the closure satisfies the Kuratowski closure axioms.*

Problem 1.4. *Show that the closure and interior operators are dual in the sense that*

$$X \backslash \overline{A} = (X \backslash A)^\circ \qquad \text{and} \qquad X \backslash A^\circ = \overline{(X \backslash A)}.$$

(Hint: De Morgan's laws.)

Problem 1.5. *Let* $U \subseteq V$ *be subsets of a metric space* X. *Show that if* U *is dense in* V *and* V *is dense in* X, *then* U *is dense in* X.

Problem 1.6. *Show that every open set* $O \subseteq \mathbb{R}$ *can be written as a countable union of disjoint intervals. (Hint: Let* $\{I_\alpha\}$ *be the set of all maximal open subintervals of* O*; that is,* $I_\alpha \subseteq O$ *and there is no other subinterval of* O *which contains* I_α. *Then this is a cover of disjoint open intervals which has a countable subcover.)*

1.2. The Banach space of continuous functions

Now let us have a first look at Banach spaces by investigating the set of continuous functions $C(I)$ on a compact interval $I = [a,b] \subset \mathbb{R}$. Since we want to handle complex models, we will always consider complex-valued functions!

One way of declaring a distance, well-known from calculus, is the **maximum norm**:

$$\|f\|_\infty = \max_{x \in I} |f(x)|. \tag{1.17}$$

It is not hard to see that with this definition $C(I)$ becomes a normed vector space:

A **normed vector space** X is a vector space X over $\mathbb{C}$ (or $\mathbb{R}$) with a nonnegative function (the **norm**) $\|.\|$ such that

- $\|f\| > 0$ for $f \neq 0$ (**positive definiteness**),
- $\|\alpha f\| = |\alpha| \, \|f\|$ for all $\alpha \in \mathbb{C}$, $f \in X$ (**positive homogeneity**), and
- $\|f + g\| \le \|f\| + \|g\|$ for all $f, g \in X$ (**triangle inequality).**

If positive definiteness is dropped from the requirements one calls $\|.\|$ a **seminorm**.

From the triangle inequality we also get the **inverse triangle inequality** (Problem 1.7)

$$|\|f\| - \|g\|| \le \|f - g\|. \tag{1.18}$$

Once we have a norm, we have a **distance** $d(f, g) = \|f-g\|$ and hence we know when a sequence of vectors f_n **converges** to a vector f. We will write $f_n \to f$ or $\lim_{n\to\infty} f_n = f$, as usual, in this case. Moreover, a mapping $F : X \to Y$ between two normed spaces is called **continuous** if $f_n \to f$ implies $F(f_n) \to F(f)$. In fact, it is not hard to see that the norm, vector addition, and multiplication by scalars are continuous (Problem 1.8).

In addition to the concept of convergence we have also the concept of a **Cauchy sequence** and hence the concept of completeness: A normed space is called **complete** if every Cauchy sequence has a limit. A complete normed space is called a **Banach space**.

Example. The space $\ell^1(\mathbb{N})$ of all complex-valued sequences $x = (x_j)_{j=1}^{\infty}$ for which the norm

$$\|x\|_1 = \sum_{j=1}^{\infty} |a_j| \tag{1.19}$$

is finite is a Banach space.

To show this, we need to verify three things: (i) $\ell^1(\mathbb{N})$ is a vector space that is closed under addition and scalar multiplication, (ii) $\|.\|_1$ satisfies the three requirements for a norm, and (iii) $\ell^1(\mathbb{N})$ is complete.

First of all observe

$$\sum_{j=1}^{k} |x_j + y_j| \le \sum_{j=1}^{k} |x_j| + \sum_{j=1}^{k} |y_j| \le \|x\|_1 + \|y\|_1 \tag{1.20}$$

for every finite k. Letting $k \to \infty$, we conclude that $\ell^1(\mathbb{N})$ is closed under addition and that the triangle inequality holds. That $\ell^1(\mathbb{N})$ is closed under scalar multiplication together with homogeneity as well as definiteness are straightforward. It remains to show that $\ell^1(\mathbb{N})$ is complete. Let $x^n = (x_j^n)_{j=1}^{\infty}$ be a Cauchy sequence; that is, for given $\varepsilon > 0$ we can find an N_ε such that $\|x^m - x^n\|_1 \le \varepsilon$ for $m, n \ge N_\varepsilon$. This implies in particular $|x_j^m - x_j^n| \le \varepsilon$ for every fixed j. Thus x_j^n is a Cauchy sequence for fixed j and by completeness of $\mathbb{C}$ has a limit: $\lim_{n\to\infty} x_j^n = x_j$. Now consider

$$\sum_{j=1}^{k} |x_j^m - x_j^n| \le \varepsilon \tag{1.21}$$

and take $m \to \infty$:

$$\sum_{j=1}^{k} |x_j - x_j^n| \le \varepsilon. \tag{1.22}$$

Since this holds for all finite k, we even have $\|x - x^n\|_1 \le \varepsilon$. Hence $(x - x^n) \in \ell^1(\mathbb{N})$ and since $x^n \in \ell^1(\mathbb{N})$, we also conclude $x = x^n + (x - x^n) \in \ell^1(\mathbb{N})$. $\diamond$

Example. The previous example can be generalized by considering the space $\ell^p(\mathbb{N})$ of all complex-valued sequences $x = (x_j)_{j=1}^\infty$ for which the norm

$$\|x\|_p = \left(\sum_{j=1}^{\infty} |x_j|^p\right)^{1/p}, \qquad p \in [1, \infty), \tag{1.23}$$

is finite. By $|x_j + y_j|^p \le 2^p \max(|x_j|, |y_j|)^p = 2^p \max(|x_j|^p, |y_j|^p) \le 2^p(|x_j|^p + |y_j|^p)$ it is a vector space, but the triangle inequality is only easy to see in the case $p = 1$.

To prove it we need the elementary inequality (Problem 1.11)

$$a^{1/p} b^{1/q} \le \frac{1}{p} a + \frac{1}{q} b, \qquad \frac{1}{p} + \frac{1}{q} = 1, \quad a, b \ge 0, \tag{1.24}$$

which implies **Hölder's inequality**

$$\|xy\|_1 \le \|x\|_p \|y\|_q \tag{1.25}$$

for $x \in \ell^p(\mathbb{N})$, $y \in \ell^q(\mathbb{N})$. In fact, by homogeneity of the norm it suffices to prove the case $\|x\|_p = \|y\|_q = 1$. But this case follows by choosing $a = |x_j|^p$ and $b = |y_j|^q$ in (1.24) and summing over all j.

Now using $|x_j + y_j|^p \le |x_j|\, |x_j + y_j|^{p-1} + |y_j|\, |x_j + y_j|^{p-1}$, we obtain from Hölder's inequality (note $(p-1)q = p$)

$$\begin{aligned} \|x + y\|_p^p &\le \|x\|_p \|(x + y)^{p-1}\|_q + \|y\|_p \|(x + y)^{p-1}\|_q \\ &= (\|x\|_p + \|y\|_p) \|(x + y)\|_p^{p-1}. \end{aligned}$$

Hence ℓ^p is a normed space. That it is complete can be shown as in the case $p = 1$ (Problem 1.12). $\diamond$

Example. The space $\ell^\infty(\mathbb{N})$ of all complex-valued bounded sequences $x = (x_j)_{j=1}^\infty$ together with the norm

$$\|x\|_\infty = \sup_{j \in \mathbb{N}} |x_j| \tag{1.26}$$

is a Banach space (Problem 1.13). Note that with this definition Hölder's inequality (1.25) remains true for the cases $p = 1$, $q = \infty$ and $p = \infty$, $q = 1$. The reason for the notation is explained in Problem 1.15. $\diamond$

Example. Every closed subspace of a Banach space is again a Banach space. For example, the space $c_0(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$ of all sequences converging to zero is a closed subspace. In fact, if $x \in \ell^\infty(\mathbb{N}) \setminus c_0(\mathbb{N})$, then $\liminf_{j\to\infty} |x_j| \ge \varepsilon > 0$ and thus $\|x - y\|_\infty \ge \varepsilon$ for every $y \in c_0(\mathbb{N})$. ⋄

Now what about convergence in the space $C(I)$? A sequence of functions $f_n(x)$ converges to f if and only if

$$\lim_{n\to\infty} \|f - f_n\|_\infty = \lim_{n\to\infty} \sup_{x\in I} |f_n(x) - f(x)| = 0. \tag{1.27}$$

That is, in the language of real analysis, f_n converges uniformly to f. Now let us look at the case where f_n is only a Cauchy sequence. Then $f_n(x)$ is clearly a Cauchy sequence of real numbers for every fixed $x \in I$. In particular, by completeness of $\mathbb{C}$, there is a limit $f(x)$ for each x. Thus we get a limiting function $f(x)$. Moreover, letting $m \to \infty$ in

$$|f_m(x) - f_n(x)| \le \varepsilon \qquad \forall m, n > N_\varepsilon,\ x \in I, \tag{1.28}$$

we see

$$|f(x) - f_n(x)| \le \varepsilon \qquad \forall n > N_\varepsilon,\ x \in I; \tag{1.29}$$

that is, $f_n(x)$ converges uniformly to $f(x)$. However, up to this point we do not know whether it is in our vector space $C(I)$ or not, that is, whether it is continuous or not. Fortunately, there is a well-known result from real analysis which tells us that the uniform limit of continuous functions is again continuous: Fix $x \in I$ and $\varepsilon > 0$. To show that f is continuous we need to find a δ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$. Pick n so that $\|f_n - f\|_\infty < \varepsilon/3$ and δ so that $|x - y| < \delta$ implies $|f_n(x) - f_n(y)| < \varepsilon/3$. Then $|x - y| < \delta$ implies

$$|f(x)-f(y)| \le |f(x)-f_n(x)|+|f_n(x)-f_n(y)|+|f_n(y)-f(y)| < \frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3} = \varepsilon$$

as required. Hence $f(x) \in C(I)$ and thus every Cauchy sequence in $C(I)$ converges. Or, in other words

Theorem 1.18. *$C(I)$ with the maximum norm is a Banach space.*

Next we want to look at *countable bases*. To this end we introduce a few definitions first.

The set of all finite linear combinations of a set of vectors $\{u_n\} \subset X$ is called the **span** of $\{u_n\}$ and denoted by $\operatorname{span}\{u_n\}$. A set of vectors $\{u_n\} \subset X$ is called **linearly independent** if every finite subset is. If $\{u_n\}_{n=1}^N \subset X$, $N \in \mathbb{N} \cup \{\infty\}$, is countable, we can throw away all elements which can be expressed as linear combinations of the previous ones to obtain a subset of linearly independent vectors which have the same span.

We will call a countable set of linearly independent vectors $\{u_n\}_{n=1}^N \subset X$ a **Schauder basis** if every element $f \in X$ can be uniquely written as a countable linear combination of the basis elements:

$$f = \sum_{n=1}^{N} c_n u_n, \qquad c_n = c_n(f) \in \mathbb{C}, \tag{1.30}$$

where the sum has to be understood as a limit if $N = \infty$ (the sum is not required to converge unconditionally). Since we have assumed the set to be linearly independent, the coefficients $c_n(f)$ are uniquely determined.

Example. The set of vectors δ^n, with $\delta^n_n = 1$ and $\delta^n_m = 0$, $n \neq m$, is a Schauder Basis for the Banach space $\ell^1(\mathbb{N})$.

Let $x = (x_j)_{j=1}^\infty \in \ell^1(\mathbb{N})$ be given and set $x^n = \sum_{j=1}^n x_j \delta^j$. Then

$$\|x - x^n\|_1 = \sum_{j=n+1}^{\infty} |x_j| \to 0$$

since $x^n_j = x_j$ for $1 \le j \le n$ and $x^n_j = 0$ for $j > n$. Hence

$$x = \sum_{j=1}^{\infty} x_j \delta^j$$

and $\{\delta^n\}_{n=1}^\infty$ is a Schauder basis (linear independence is left as an exercise). ◇

A set whose span is dense is called **total** and if we have a countable total set, we also have a countable dense set (consider only linear combinations with rational coefficients — show this). A normed vector space containing a countable dense set is called **separable**.

Example. Every Schauder basis is total and thus every Banach space with a Schauder basis is separable (the converse is not true). In particular, the Banach space $\ell^1(\mathbb{N})$ is separable. ◇

While we will not give a Schauder basis for $C(I)$, we will at least show that it is separable. In order to prove this we need a lemma first.

Lemma 1.19 (Smoothing). *Let $u_n(x)$ be a sequence of nonnegative continuous functions on $[-1,1]$ such that*

$$\int_{|x|\le 1} u_n(x)dx = 1 \quad \textit{and} \quad \int_{\delta \le |x| \le 1} u_n(x)dx \to 0, \quad \delta > 0. \tag{1.31}$$

(In other words, u_n has mass one and concentrates near $x = 0$ as $n \to \infty$.)

Then for every $f \in C[-\frac{1}{2}, \frac{1}{2}]$ which vanishes at the endpoints, $f(-\frac{1}{2}) = f(\frac{1}{2}) = 0$, we have that

$$f_n(x) = \int_{-1/2}^{1/2} u_n(x-y) f(y) dy \tag{1.32}$$

converges uniformly to $f(x)$.

Proof. Since f is uniformly continuous, for given ε we can find a $\delta < 1/2$ (independent of x) such that $|f(x) - f(y)| \le \varepsilon$ whenever $|x - y| \le \delta$. Moreover, we can choose n such that $\int_{\delta \le |y| \le 1} u_n(y)dy \le \varepsilon$. Now abbreviate $M = \max_{x\in[-1/2,1/2]}\{1, |f(x)|\}$ and note

$$|f(x) - \int_{-1/2}^{1/2} u_n(x-y)f(x)dy| = |f(x)|\,|1 - \int_{-1/2}^{1/2} u_n(x-y)dy| \le M\varepsilon.$$

In fact, either the distance of x to one of the boundary points $\pm\frac{1}{2}$ is smaller than δ and hence $|f(x)| \le \varepsilon$ or otherwise $[-\delta,\delta] \subset [x-1/2, x+1/2]$ and the difference between one and the integral is smaller than ε.

Using this, we have

$$\begin{aligned}
|f_n(x) - f(x)| \le & \int_{-1/2}^{1/2} u_n(x-y)|f(y) - f(x)|dy + M\varepsilon \\
= & \int_{|y|\le 1/2, |x-y|\le\delta} u_n(x-y)|f(y) - f(x)|dy \\
& + \int_{|y|\le 1/2, |x-y|\ge\delta} u_n(x-y)|f(y) - f(x)|dy + M\varepsilon \\
\le & \varepsilon + 2M\varepsilon + M\varepsilon = (1+3M)\varepsilon, \qquad (1.33)
\end{aligned}$$

which proves the claim. □

Note that f_n will be as smooth as u_n, hence the title smoothing lemma. Moreover, f_n will be a polynomial if u_n is. The same idea is used to approximate noncontinuous functions by smooth ones (of course the convergence will no longer be uniform in this case).

Now we are ready to show:

Theorem 1.20 (Weierstraß)**.** *Let I be a compact interval. Then the set of polynomials is dense in $C(I)$.*

Proof. Let $f(x) \in C(I)$ be given. By considering $f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x - a)$ it is no loss to assume that f vanishes at the boundary points. Moreover, without restriction we only consider $I = [\frac{-1}{2}, \frac{1}{2}]$ (why?).

Now the claim follows from Lemma 1.19 using

$$u_n(x) = \frac{1}{I_n}(1 - x^2)^n,$$

where

$$I_n = \int_{-1}^{1} (1-x^2)^n dx = \frac{n}{n+1} \int_{-1}^{1} (1-x)^{n-1}(1+x)^{n+1} dx$$
$$= \cdots = \frac{n!}{(n+1)\cdots(2n+1)} 2^{2n+1} = \frac{(n!)^2 2^{2n+1}}{(2n+1)!} = \frac{n!}{\frac{1}{2}(\frac{1}{2}+1)\cdots(\frac{1}{2}+n)}.$$

Indeed, the first part of (1.31) holds by construction and the second part follows from the elementary estimate

$$\frac{2}{2n+1} \le I_n < 2.$$

□

Corollary 1.21. *C(I) is separable.*

However, $\ell^\infty(\mathbb{N})$ is not separable (Problem 1.14)!

Problem 1.7. *Show that* $|\|f\| - \|g\|| \le \|f-g\|$.

Problem 1.8. *Let X be a Banach space. Show that the norm, vector addition, and multiplication by scalars are continuous. That is, if $f_n \to f$, $g_n \to g$, and $\alpha_n \to \alpha$, then $\|f_n\| \to \|f\|$, $f_n + g_n \to f+g$, and $\alpha_n g_n \to \alpha g$.*

Problem 1.9. *Let X be a Banach space. Show that $\sum_{j=1}^\infty \|f_j\| < \infty$ implies that*

$$\sum_{j=1}^\infty f_j = \lim_{n\to\infty} \sum_{j=1}^n f_j$$

exists. The series is called **absolutely convergent** *in this case.*

Problem 1.10. *While $\ell^1(\mathbb{N})$ is separable, it still has room for an uncountable set of linearly independent vectors. Show this by considering vectors of the form*

$$x^a = (1, a, a^2, \dots), \qquad a \in (0,1).$$

(Hint: Take n such vectors and cut them off after $n+1$ terms. If the cut off vectors are linearly independent, so are the original ones. Recall the Vandermonde determinant.)

Problem 1.11. *Prove* (1.24). *(Hint: Take logarithms on both sides.)*

Problem 1.12. *Show that $\ell^p(\mathbb{N})$ is a separable Banach space.*

Problem 1.13. *Show that $\ell^\infty(\mathbb{N})$ is a Banach space.*

Problem 1.14. *Show that $\ell^\infty(\mathbb{N})$ is not separable. (Hint: Consider sequences which take only the value one and zero. How many are there? What is the distance between two such sequences?)*

Problem 1.15. *Show that if* $x \in \ell^{p_0}(\mathbb{N})$ *for some* $p_0 \in [1,\infty)$*, then* $x \in \ell^p(\mathbb{N})$ *for* $p \geq p_0$ *and*

$$\lim_{p\to\infty} \|x\|_p = \|x\|_\infty.$$

1.3. The geometry of Hilbert spaces

So it looks like $C(I)$ has all the properties we want. However, there is still one thing missing: How should we define orthogonality in $C(I)$? In Euclidean space, two vectors are called **orthogonal** if their scalar product vanishes, so we would need a scalar product:

Suppose $\mathfrak{H}$ is a vector space. A map $\langle .,..\rangle : \mathfrak{H}\times\mathfrak{H} \to \mathbb{C}$ is called a **sesquilinear form** if it is conjugate linear in the first argument and linear in the second; that is,

$$\begin{aligned} \langle \alpha_1 f_1 + \alpha_2 f_2, g\rangle &= \alpha_1^*\langle f_1, g\rangle + \alpha_2^*\langle f_2, g\rangle, \\ \langle f, \alpha_1 g_1 + \alpha_2 g_2\rangle &= \alpha_1\langle f, g_1\rangle + \alpha_2\langle f, g_2\rangle, \end{aligned} \qquad \alpha_1,\alpha_2 \in \mathbb{C}, \tag{1.34}$$

where '$*$' denotes complex conjugation. A sesquilinear form satisfying the requirements

(i) $\langle f,f\rangle > 0$ for $f \neq 0$ (positive definite),

(ii) $\langle f,g\rangle = \langle g,f\rangle^*$ (symmetry)

is called an **inner product** or **scalar product**. Associated with every scalar product is a norm

$$\|f\| = \sqrt{\langle f,f\rangle}. \tag{1.35}$$

Only the triangle inequality is nontrivial. It will follow from the Cauchy–Schwarz inequality below. Until then, just regard (1.35) as a convenient short hand notation.

The pair $(\mathfrak{H}, \langle .,..\rangle)$ is called an **inner product space**. If $\mathfrak{H}$ is complete (with respect to the norm (1.35)), it is called a **Hilbert space**.

Example. Clearly $\mathbb{C}^n$ with the usual scalar product

$$\langle x,y\rangle = \sum_{j=1}^n x_j^* y_j \tag{1.36}$$

is a (finite dimensional) Hilbert space. ◇

Example. A somewhat more interesting example is the Hilbert space $\ell^2(\mathbb{N})$, that is, the set of all complex-valued sequences

$$\Big\{ (x_j)_{j=1}^\infty \Big| \sum_{j=1}^\infty |x_j|^2 < \infty \Big\} \tag{1.37}$$

with scalar product

$$\langle x, y\rangle = \sum_{j=1}^{\infty} x_j^* y_j. \tag{1.38}$$

(Show that this is in fact a separable Hilbert space — Problem 1.12.) ⋄

A vector $f \in \mathfrak{H}$ is called **normalized** or a **unit vector** if $\|f\| = 1$. Two vectors $f, g \in \mathfrak{H}$ are called **orthogonal** or **perpendicular** ($f \perp g$) if $\langle f, g\rangle = 0$ and **parallel** if one is a multiple of the other.

If f and g are orthogonal, we have the **Pythagorean theorem**:

$$\|f + g\|^2 = \|f\|^2 + \|g\|^2, \qquad f \perp g, \tag{1.39}$$

which is one line of computation (do it!).

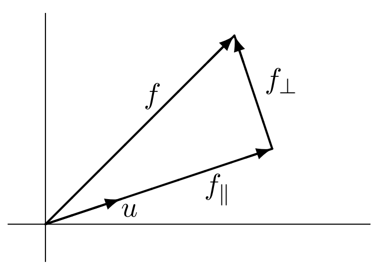
Suppose u is a unit vector. Then the projection of f in the direction of u is given by

$$f_\parallel = \langle u, f\rangle u \tag{1.40}$$

and $f_\perp$ defined via

$$f_\perp = f - \langle u, f\rangle u \tag{1.41}$$

is perpendicular to u since $\langle u, f_\perp\rangle = \langle u, f - \langle u, f\rangle u\rangle = \langle u, f\rangle - \langle u, f\rangle\langle u, u\rangle = 0$.



Taking any other vector parallel to u, we obtain from (1.39)

$$\|f - \alpha u\|^2 = \|f_\perp + (f_\parallel - \alpha u)\|^2 = \|f_\perp\|^2 + |\langle u, f\rangle - \alpha|^2 \tag{1.42}$$

and hence $f_\parallel = \langle u, f\rangle u$ is the unique vector parallel to u which is closest to f.

As a first consequence we obtain the **Cauchy–Schwarz–Bunjakowski** inequality:

Theorem 1.22 (Cauchy–Schwarz–Bunjakowski)**.** *Let $\mathfrak{H}_0$ be an inner product space. Then for every $f, g \in \mathfrak{H}_0$ we have*

$$|\langle f, g\rangle| \le \|f\|\,\|g\| \tag{1.43}$$

with equality if and only if f and g are parallel.

Proof. It suffices to prove the case $\|g\| = 1$. But then the claim follows from $\|f\|^2 = |\langle g, f\rangle|^2 + \|f_\perp\|^2$. □

Note that the Cauchy–Schwarz inequality entails that the scalar product is continuous in both variables; that is, if $f_n \to f$ and $g_n \to g$, we have $\langle f_n, g_n\rangle \to \langle f, g\rangle$.

As another consequence we infer that the map $\|.\|$ is indeed a norm. In fact,

$$\|f+g\|^2 = \|f\|^2 + \langle f, g\rangle + \langle g, f\rangle + \|g\|^2 \le (\|f\| + \|g\|)^2. \tag{1.44}$$

But let us return to $C(I)$. Can we find a scalar product which has the maximum norm as associated norm? Unfortunately the answer is no! The reason is that the maximum norm does not satisfy the parallelogram law (Problem 1.18).

Theorem 1.23 (Jordan–von Neumann)**.** *A norm is associated with a scalar product if and only if the* **parallelogram law**

$$\|f+g\|^2 + \|f-g\|^2 = 2\|f\|^2 + 2\|g\|^2 \tag{1.45}$$

holds.

In this case the scalar product can be recovered from its norm by virtue of the **polarization identity**

$$\langle f, g\rangle = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2 + \mathrm{i}\|f-\mathrm{i}g\|^2 - \mathrm{i}\|f+\mathrm{i}g\|^2\right). \tag{1.46}$$

Proof. If an inner product space is given, verification of the parallelogram law and the polarization identity is straightforward (Problem 1.20).

To show the converse, we define

$$s(f,g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2 + \mathrm{i}\|f-\mathrm{i}g\|^2 - \mathrm{i}\|f+\mathrm{i}g\|^2\right).$$

Then $s(f,f) = \|f\|^2$ and $s(f,g) = s(g,f)^*$ are straightforward to check. Moreover, another straightforward computation using the parallelogram law shows

$$s(f,g) + s(f,h) = 2s(f, \frac{g+h}{2}).$$

Now choosing $h = 0$ (and using $s(f,0) = 0$) shows $s(f,g) = 2s(f,\frac{g}{2})$ and thus $s(f,g) + s(f,h) = s(f, g+h)$. Furthermore, by induction we infer $\frac{m}{2^n} s(f,g) = s(f, \frac{m}{2^n} g)$; that is, $\alpha\, s(f,g) = s(f, \alpha g)$ for every positive rational α. By continuity (which follows from continuity of $\|.\|$) this holds for all $\alpha > 0$ and $s(f,-g) = -s(f,g)$, respectively, $s(f,\mathrm{i}g) = \mathrm{i}\, s(f,g)$, finishes the proof. □

Note that the parallelogram law and the polarization identity even hold for sesquilinear forms (Problem 1.20).

But how do we define a scalar product on $C(I)$? One possibility is

$$\langle f, g\rangle = \int_a^b f^*(x)g(x)dx. \tag{1.47}$$

The corresponding inner product space is denoted by $\mathcal{L}^2_{cont}(I)$. Note that we have

$$\|f\| \le \sqrt{|b-a|}\|f\|_\infty \tag{1.48}$$

and hence the maximum norm is stronger than the $\mathcal{L}^2_{cont}$ norm.

Suppose we have two norms $\|.\|_1$ and $\|.\|_2$ on a vector space X. Then $\|.\|_2$ is said to be **stronger** than $\|.\|_1$ if there is a constant $m > 0$ such that

$$\|f\|_1 \le m\|f\|_2. \tag{1.49}$$

It is straightforward to check the following.

Lemma 1.24. *If $\|.\|_2$ is stronger than $\|.\|_1$, then every $\|.\|_2$ Cauchy sequence is also a $\|.\|_1$ Cauchy sequence.*

Hence if a function $F : X \to Y$ is continuous in $(X, \|.\|_1)$, it is also continuous in $(X, \|.\|_2)$ and if a set is dense in $(X, \|.\|_2)$, it is also dense in $(X, \|.\|_1)$.

In particular, $\mathcal{L}^2_{cont}$ is separable. But is it also complete? Unfortunately the answer is no:

Example. Take $I = [0, 2]$ and define

$$f_n(x) = \begin{cases} 0, & 0 \le x \le 1 - \frac{1}{n}, \\ 1 + n(x-1), & 1 - \frac{1}{n} \le x \le 1, \\ 1, & 1 \le x \le 2. \end{cases} \tag{1.50}$$

Then $f_n(x)$ is a Cauchy sequence in $\mathcal{L}^2_{cont}$, but there is no limit in $\mathcal{L}^2_{cont}$! Clearly the limit should be the step function which is 0 for $0 \le x < 1$ and 1 for $1 \le x \le 2$, but this step function is discontinuous (Problem 1.23)! ⋄

This shows that in infinite dimensional vector spaces different norms will give rise to different convergent sequences! In fact, the key to solving problems in infinite dimensional spaces is often finding the right norm! This is something which cannot happen in the finite dimensional case.

Theorem 1.25. *If X is a finite dimensional vector space, then all norms are equivalent. That is, for any two given norms $\|.\|_1$ and $\|.\|_2$, there are positive constants m_1 and m_2 such that*

$$\frac{1}{m_2}\|f\|_1 \le \|f\|_2 \le m_1\|f\|_1. \tag{1.51}$$

Proof. Since equivalence of norms is an equivalence relation (check this!) we can assume that $\|.\|_2$ is the usual Euclidean norm. Moreover, we can choose an orthogonal basis u_j, $1 \le j \le n$, such that $\|\sum_j \alpha_j u_j\|_2^2 = \sum_j |\alpha_j|^2$. Let $f = \sum_j \alpha_j u_j$. Then by the triangle and Cauchy–Schwarz inequalities

$$\|f\|_1 \le \sum_j |\alpha_j| \|u_j\|_1 \le \sqrt{\sum_j \|u_j\|_1^2}\, \|f\|_2$$

and we can choose $m_2 = \sqrt{\sum_j \|u_j\|_1}$.

In particular, if f_n is convergent with respect to $\|.\|_2$, it is also convergent with respect to $\|.\|_1$. Thus $\|.\|_1$ is continuous with respect to $\|.\|_2$ and attains its minimum $m > 0$ on the unit sphere (which is compact by the Heine-Borel theorem, Theorem 1.12). Now choose $m_1 = 1/m$. □

Problem 1.16. *Show that the norm in a Hilbert space satisfies* $\|f + g\| = \|f\| + \|g\|$ *if and only if* $f = \alpha g$, $\alpha \ge 0$, *or* $g = 0$.

Problem 1.17 (Generalized parallelogram law)**.** *Show that in a Hilbert space*

$$\sum_{1 \le j < k \le n} \|x_j - x_k\|^2 + \|\sum_{1 \le j \le n} x_j\|^2 = n \sum_{1 \le j \le n} \|x_j\|^2.$$

The case $n = 2$ *is* (1.45).

Problem 1.18. *Show that the maximum norm on* $C[0,1]$ *does not satisfy the parallelogram law.*

Problem 1.19. *In a Banach space the unit ball is convex by the triangle inequality. A Banach space* X *is called* **uniformly convex** *if for every* $\varepsilon > 0$ *there is some* δ *such that* $\|x\| \le 1$, $\|y\| \le 1$, *and* $\|\frac{x+y}{2}\| \ge 1 - \delta$ *imply* $\|x - y\| \le \varepsilon$.

Geometrically this implies that if the average of two vectors inside the closed unit ball is close to the boundary, then they must be close to each other.

Show that a Hilbert space is uniformly convex and that one can choose $\delta(\varepsilon) = 1 - \sqrt{1 - \frac{\varepsilon^2}{4}}$. *Draw the unit ball for* $\mathbb{R}^2$ *for the norms* $\|x\|_1 = |x_1| + |x_2|$, $\|x\|_2 = \sqrt{|x_1|^2 + |x_2|^2}$, *and* $\|x\|_\infty = \max(|x_1|, |x_2|)$. *Which of these norms makes* $\mathbb{R}^2$ *uniformly convex?*

(Hint: For the first part use the parallelogram law.)

Problem 1.20. *Suppose* $\mathfrak{Q}$ *is a vector space. Let* $s(f,g)$ *be a sesquilinear form on* $\mathfrak{Q}$ *and* $q(f) = s(f,f)$ *the associated quadratic form. Prove the* **parallelogram law**

$$q(f+g) + q(f-g) = 2q(f) + 2q(g) \tag{1.52}$$

and the **polarization identity**

$$s(f,g) = \frac{1}{4}\left(q(f+g) - q(f-g) + \mathrm{i}\,q(f-\mathrm{i}g) - \mathrm{i}\,q(f+\mathrm{i}g)\right). \tag{1.53}$$

Conversely, show that every quadratic form $q(f) : \mathfrak{Q} \to \mathbb{R}$ satisfying $q(\alpha f) = |\alpha|^2 q(f)$ and the parallelogram law gives rise to a sesquilinear form via the polarization identity.

Show that $s(f,g)$ is symmetric if and only if $q(f)$ is real-valued.

Problem 1.21. *A sesquilinear form is called* **bounded** *if*

$$\|s\| = \sup_{\|f\|=\|g\|=1} |s(f,g)|$$

is finite. Similarly, the associated quadratic form q is **bounded** *if*

$$\|q\| = \sup_{\|f\|=1} |q(f)|$$

is finite. Show

$$\|q\| \le \|s\| \le 2\|q\|.$$

(Hint: Use the parallelogram law and the polarization identity from the previous problem.)

Problem 1.22. *Suppose $\mathfrak{Q}$ is a vector space. Let $s(f,g)$ be a sesquilinear form on $\mathfrak{Q}$ and $q(f) = s(f,f)$ the associated quadratic form. Show that the Cauchy–Schwarz inequality*

$$|s(f,g)| \le q(f)^{1/2} q(g)^{1/2} \tag{1.54}$$

holds if $q(f) \ge 0$.

(Hint: Consider $0 \le q(f+\alpha g) = q(f) + 2\mathrm{Re}(\alpha\, s(f,g)) + |\alpha|^2 q(g)$ and choose $\alpha = t\, s(f,g)^*/|s(f,g)|$ with $t \in \mathbb{R}$.)

Problem 1.23. *Prove the claims made about f_n, defined in (1.50), in the last example.*

1.4. Completeness

Since $\mathcal{L}^2_{cont}$ is not complete, how can we obtain a Hilbert space from it? Well, the answer is simple: take the **completion**.

If X is an (incomplete) normed space, consider the set of all Cauchy sequences $\tilde{X}$. Call two Cauchy sequences equivalent if their difference converges to zero and denote by $\bar{X}$ the set of all equivalence classes. It is easy to see that $\bar{X}$ (and $\tilde{X}$) inherit the vector space structure from X. Moreover,

Lemma 1.26. *If x_n is a Cauchy sequence, then $\|x_n\|$ converges.*

Consequently the norm of a Cauchy sequence $(x_n)_{n=1}^{\infty}$ can be defined by $\|(x_n)_{n=1}^{\infty}\| = \lim_{n\to\infty} \|x_n\|$ and is independent of the equivalence class (show this!). Thus $\bar{X}$ is a normed space ($\tilde{X}$ is not! Why?).

Theorem 1.27. *$\bar{X}$ is a Banach space containing X as a dense subspace if we identify $x \in X$ with the equivalence class of all sequences converging to x.*

Proof. (Outline) It remains to show that $\bar{X}$ is complete. Let $\xi_n = [(x_{n,j})_{j=1}^{\infty}]$ be a Cauchy sequence in $\bar{X}$. Then it is not hard to see that $\xi = [(x_{j,j})_{j=1}^{\infty}]$ is its limit. □

Let me remark that the completion $\bar{X}$ is unique. More precisely every other complete space which contains X as a dense subset is isomorphic to $\bar{X}$. This can for example be seen by showing that the identity map on X has a unique extension to $\bar{X}$ (compare Theorem 1.29 below).

In particular it is no restriction to assume that a normed vector space or an inner product space is complete. However, in the important case of $\mathcal{L}^2_{cont}$ it is somewhat inconvenient to work with equivalence classes of Cauchy sequences and hence we will give a different characterization using the Lebesgue integral later.

Problem 1.24. *Provide a detailed proof of Theorem 1.27.*

1.5. Bounded operators

A linear map A between two normed spaces X and Y will be called a **(linear) operator**

$$A : \mathfrak{D}(A) \subseteq X \to Y. \tag{1.55}$$

The linear subspace $\mathfrak{D}(A)$ on which A is defined is called the **domain** of A and is usually required to be dense. The **kernel** (also **null space**)

$$\operatorname{Ker}(A) = \{f \in \mathfrak{D}(A) | Af = 0\} \subseteq X \tag{1.56}$$

and **range**

$$\operatorname{Ran}(A) = \{Af | f \in \mathfrak{D}(A)\} = A\mathfrak{D}(A) \subseteq Y \tag{1.57}$$

are defined as usual. The operator A is called **bounded** if the operator norm

$$\|A\| = \sup_{f\in\mathfrak{D}(A), \|f\|_X=1} \|Af\|_Y \tag{1.58}$$

is finite.

By construction, a bounded operator is Lipschitz continuous,

$$\|Af\|_Y \le \|A\|\|f\|_X, \qquad f \in \mathfrak{D}(A), \tag{1.59}$$

and hence continuous. The converse is also true

Theorem 1.28. *An operator A is bounded if and only if it is continuous.*

Proof. Suppose A is continuous but not bounded. Then there is a sequence of unit vectors u_n such that $\|Au_n\| \geq n$. Then $f_n = \frac{1}{n} u_n$ converges to 0 but $\|Af_n\| \geq 1$ does not converge to 0. □

In particular, if X is finite dimensional, then every operator is bounded. Note that in general one and the same operation might be bounded (i.e. continuous) or unbounded, depending on the norm chosen.

Example. Consider the vector space of differentiable functions $X = C^1[0,1]$ and equip it with the norm (cf. Problem 1.27)

$$\|f\|_{\infty,1} = \max_{x\in[0,1]} |f(x)| + \max_{x\in[0,1]} |f'(x)|$$

Let $Y = C[0,1]$ and observe that the differential operator $A = \frac{d}{dx} : X \to Y$ is bounded since

$$\|Af\|_\infty = \max_{x\in[0,1]} |f'(x)| \leq \max_{x\in[0,1]} |f(x)| + \max_{x\in[0,1]} |f'(x)| = \|f\|_{\infty,1}.$$

However, if we consider $A = \frac{d}{dx} : \mathfrak{D}(A) \subseteq Y \to Y$ defined on $\mathfrak{D}(A) = C^1[0,1]$, then we have an unbounded operator. Indeed, choose

$$u_n(x) = \sin(n\pi x)$$

which is normalized, $\|u_n\|_\infty = 1$, and observe that

$$Au_n(x) = u_n'(x) = n\pi\cos(n\pi x)$$

is unbounded, $\|Au_n\|_\infty = n\pi$. Note that $\mathfrak{D}(A)$ contains the set of polynomials and thus is dense by the Weierstraß approximation theorem (Theorem 1.20). ⋄

If A is bounded and densely defined, it is no restriction to assume that it is defined on all of X.

Theorem 1.29 (B.L.T. theorem)**.** *Let $A : \mathfrak{D}(A) \subseteq X \to Y$ be an bounded linear operator and let Y be a Banach space. If $\mathfrak{D}(A)$ is dense, there is a unique (continuous) extension of A to X which has the same operator norm.*

Proof. Since a bounded operator maps Cauchy sequences to Cauchy sequences, this extension can only be given by

$$\overline{A}f = \lim_{n\to\infty} Af_n, \qquad f_n \in \mathfrak{D}(A), \quad f \in X.$$

To show that this definition is independent of the sequence $f_n \to f$, let $g_n \to f$ be a second sequence and observe

$$\|Af_n - Ag_n\| = \|A(f_n - g_n)\| \leq \|A\|\|f_n - g_n\| \to 0.$$

Since for $f \in \mathfrak{D}(A)$ we can choose $f_n = f$, we see that $\overline{A}f = Af$ in this case, that is, $\overline{A}$ is indeed an extension. From continuity of vector addition and scalar multiplication it follows that $\overline{A}$ is linear. Finally, from continuity of the norm we conclude that the operator norm does not increase. □

The set of all bounded linear operators from X to Y is denoted by $\mathfrak{L}(X,Y)$. If $X = Y$, we write $\mathfrak{L}(X,X) = \mathfrak{L}(X)$. An operator in $\mathfrak{L}(X,\mathbb{C})$ is called a **bounded linear functional** and the space $X^* = \mathfrak{L}(X,\mathbb{C})$ is called the dual space of X.

Theorem 1.30. *The space $\mathfrak{L}(X,Y)$ together with the operator norm (1.58) is a normed space. It is a Banach space if Y is.*

Proof. That (1.58) is indeed a norm is straightforward. If Y is complete and A_n is a Cauchy sequence of operators, then $A_n f$ converges to an element g for every f. Define a new operator A via $Af = g$. By continuity of the vector operations, A is linear and by continuity of the norm $\|Af\| = \lim_{n\to\infty} \|A_n f\| \le (\lim_{n\to\infty} \|A_n\|)\|f\|$, it is bounded. Furthermore, given $\varepsilon > 0$ there is some N such that $\|A_n - A_m\| \le \varepsilon$ for $n, m \ge N$ and thus $\|A_n f - A_m f\| \le \varepsilon\|f\|$. Taking the limit $m \to \infty$, we see $\|A_n f - Af\| \le \varepsilon\|f\|$; that is, $A_n \to A$. □

The Banach space of bounded linear operators $\mathfrak{L}(X)$ even has a multiplication given by composition. Clearly this multiplication satisfies

$$(A+B)C = AC + BC, \quad A(B+C) = AB + BC, \quad A, B, C \in \mathfrak{L}(X) \tag{1.60}$$

and

$$(AB)C = A(BC), \qquad \alpha\,(AB) = (\alpha A)B = A\,(\alpha B), \quad \alpha \in \mathbb{C}. \tag{1.61}$$

Moreover, it is easy to see that we have

$$\|AB\| \le \|A\|\|B\|. \tag{1.62}$$

In other words, $\mathfrak{L}(X)$ is a so-called **Banach algebra**. However, note that our multiplication is not commutative (unless X is one-dimensional). We even have an **identity**, the identity operator $\mathbb{I}$ satisfying $\|\mathbb{I}\| = 1$.

Problem 1.25. *Consider $X = \mathbb{C}^n$ and let $A : X \to X$ be a matrix. Equip X with the norm (show that this is a norm)*

$$\|x\|_\infty = \max_{1\le j\le n} |x_j|$$

and compute the operator norm $\|A\|$ with respect to this matrix in terms of the matrix entries. Do the same with respect to the norm

$$\|x\|_1 = \sum_{1\le j\le n} |x_j|.$$

Problem 1.26. *Show that the integral operator*

$$(Kf)(x) = \int_0^1 K(x,y) f(y) dy,$$

where $K(x,y) \in C([0,1] \times [0,1])$, defined on $\mathfrak{D}(K) = C[0,1]$ is a bounded operator both in $X = C[0,1]$ (max norm) and $X = \mathcal{L}^2_{cont}(0,1)$.

Problem 1.27. *Let I be a compact interval. Show that the set of differentiable functions $C^1(I)$ becomes a Banach space if we set $\|f\|_{\infty,1} = \max_{x\in I} |f(x)| + \max_{x\in I} |f'(x)|$.*

Problem 1.28. *Show that $\|AB\| \le \|A\| \|B\|$ for every $A, B \in \mathfrak{L}(X)$. Conclude that the multiplication is continuous: $A_n \to A$ and $B_n \to B$ imply $A_n B_n \to AB$.*

Problem 1.29. *Let $A \in \mathfrak{L}(X)$ be a bijection. Show*

$$\|A^{-1}\|^{-1} = \inf_{f\in X, \|f\|=1} \|Af\|.$$

Problem 1.30. *Let*

$$f(z) = \sum_{j=0}^{\infty} f_j z^j, \qquad |z| < R,$$

be a convergent power series with convergence radius $R > 0$. Suppose A is a bounded operator with $\|A\| < R$. Show that

$$f(A) = \sum_{j=0}^{\infty} f_j A^j$$

exists and defines a bounded linear operator (cf. Problem 1.9).

1.6. Sums and quotients of Banach spaces

Given two Banach spaces X_1 and X_2 we can define their **(direct) sum** $X = X_1 \oplus X_2$ as the cartesian product $X_1 \times X_2$ together with the norm $\|(x_1, x_2)\| = \|x_1\| + \|x_2\|$. In fact, since all norms on $\mathbb{R}^2$ are equivalent (Theorem 1.25), we could as well take $\|(x_1, x_2)\| = (\|x_1\|^p + \|x_2\|^p)^{1/p}$ or $\|(x_1, x_2)\| = \max(\|x_1\|, \|x_2\|)$. In particular, in the case of Hilbert spaces the choice $p = 2$ will ensure that X is again a Hilbert space. Note that X_1 and X_2 can be regarded as subspaces of $X_1 \times X_2$ by virtue of the obvious embeddings $x_1 \hookrightarrow (x_1, 0)$ and $x_2 \hookrightarrow (0, x_2)$. It is straightforward to show that X is again a Banach space and to generalize this concept to finitely many spaces (Problem 1.31).

Moreover, given a closed subspace M of a Banach space X we can define the **quotient space** X/M as the set of all equivalence classes $[x] = x + M$ with respect to the equivalence relation $x \equiv y$ if $x - y \in M$. It is straightforward to see that X/M is a vector space when defining $[x]+[y] = [x+y]$ and $\alpha[x] = [\alpha x]$ (show that these definitions are independent of the representative of the equivalence class).

Lemma 1.31. *Let M be a closed subspace of a Banach space X. Then X/M together with the norm*

$$\|[x]\| = \inf_{y\in M} \|x+y\|. \tag{1.63}$$

is a Banach space.

Proof. First of all we need to show that (1.63) is indeed a norm. If $\|[x]\| = 0$ we must have a sequence $x_j \in M$ with $x_j \to x$ and since M is closed we conclude $x \in M$, that is $[x] = [0]$ as required. To see $\|\alpha[x]\| = |\alpha|\|[x]\|$ we use again the definition

$$\begin{aligned}\|\alpha[x]\| = \|[\alpha x]\| = \inf_{y\in M} \|\alpha x + y\| = \inf_{y\in M} \|\alpha x + \alpha y\| \\ = |\alpha| \inf_{y\in M} \|x+y\| = |\alpha|\|[x]\|.\end{aligned}$$

The triangle inequality follows with a similar argument and is left as an exercise.

Thus (1.63) is a norm and it remains to show that X/M is complete. To this end let $[x_n]$ be a Cauchy sequence. Since it suffices to show that some subsequence has a limit, we can assume $\|[x_{n+1}]-[x_n]\| < 2^{-n}$ without loss of generality. Moreover, by definition of (1.63) we can chose the representatives x_n such that $\|x_{n+1}-x_n\| < 2^{-n}$ (start with x_1 and then chose the remaining ones inductively). By construction x_n is a Cauchy sequence which has a limit $x \in X$ since X is complete. Moreover, it is straightforward to check that $[x]$ is the limit of $[x_n]$. □

Problem 1.31. *Let X_j, $j = 1, \dots, n$ be Banach spaces. Let X be the cartesian product $X_1 \times \cdots \times X_n$ together with the norm*

$$\|(x_1, \dots, x_n)\|_p = \begin{cases} \left(\sum_{j=1}^n \|x_j\|^p\right)^{1/p}, & 1 \le p < \infty, \\ \max_{j=1,\dots,n} \|x_j\|, & p = \infty. \end{cases}$$

Show that X is a Banach space. Show that all norms are equivalent.

Problem 1.32. *Suppose $A \in \mathfrak{L}(X,Y)$. Show that $\mathrm{Ker}(A)$ is closed. Show that A is well defined on $X/\mathrm{Ker}(A)$ and that this new operator is again bounded (with the same norm) and injective.*

1.7. Spaces of continuous and differentiable functions

In this section we introduce a few further sets of continuous and differentiable functions which are of interest in applications.

First, for any set $U \subseteq \mathbb{R}^m$ the set of all bounded continuous functions $C_b(U)$ together with the sup norm

$$\|f\|_\infty = \sup_{x\in U} |f(x)| \tag{1.64}$$

is a Banach space and this can be shown as in Section 1.2. Moreover, the above norm can be augmented to handle differentiable functions by considering the space $C_b^1(U)$ of all continuously differentiable functions for which the following norm is

$$\|f\|_{\infty,1} = \|f\|_\infty + \sum_{j=1}^{m} \|\partial_j f\|_\infty \tag{1.65}$$

finite, where $\partial_j = \frac{\partial}{\partial x_j}$. Note that $\|\partial_j f\|$ for one j (or all j) is not sufficient as it is only a seminorm (it vanishes for every constant function). However, since the sum of seminorms is again a seminorm (Problem 1.330 the above expression defines indeed a norm. It is also not hard to see that $C_b^1(U,\mathbb{C}^n)$ is complete. In fact, let f^k be a Cauchy sequence, then $f^k(x)$ converges uniformly to some continuous function $f(x)$ and the same is true for the partial derivatives $\partial_j f^k(x) \to g_j(x)$. Moreover, since $f^k(x) = f^k(c, x_2, \dots, x_m) + \int_c^{x_1} \partial_j f^k(t, x_2, \dots, x_m)dt \to f(x) = f(c, x_2, \dots, x_m) + \int_c^{x_1} g_j(t, x_2, \dots, x_m)$ we obtain $\partial_j f(x) = g_j(x)$. The remaining derivatives follow analogously and thus $f^k \to f$ in $C_b^1(U,\mathbb{C}^n)$.

Clearly this approach extends to higher derivatives. To this end let $C^k(U)$ be the set of all complex-valued functions which have partial derivatives of order up to k. For $f \in C^k(U)$ and $\alpha \in \mathbb{N}_0^n$ we set

$$\partial_\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}, \qquad |\alpha| = \alpha_1 + \cdots + \alpha_n. \tag{1.66}$$

An element $\alpha \in \mathbb{N}_0^n$ is called a **multi-index** and $|\alpha|$ is called its **order**.

Theorem 1.32. *The space $C_b^k(U)$ of all functions whose partial derivatives up to order k are bounded and continuous form a Banach space with norm*

$$\|f\|_{\infty,k} = \sum_{|\alpha|\le k} \sup_{x\in U} |\partial_\alpha f(x)|. \tag{1.67}$$

An important subspace is $C_\infty^k(\mathbb{R}^n)$, the set of all functions in $C_b^k(\mathbb{R}^n)$ for which $\lim_{|x|\to\infty} |\partial_\alpha f(x)| = 0$ for all $\alpha \le k$. For any function f not in $C_\infty^k(\mathbb{R}^n)$ there must be a sequence $|x_j| \to \infty$ and some α such that $|\partial_\alpha f(x_j)| \ge \varepsilon > 0$. But then $\|f - g\|_{\infty,k} \ge \varepsilon$ for every g in $C_\infty^k(\mathbb{R}^n)$ and thus $C_\infty^k(\mathbb{R}^n)$ is a closed subspace. In particular, it is a Banach space of its own.

Note that the space $C_b^k(U)$ could be further refined by requiring the highest derivatives to be Hölder continuous. Recall that a function $f : U \to$

$\mathbb{C}$ is called uniformly **Hölder continuous** with exponent $\gamma \in (0,1]$ if

$$[f]_\gamma = \sup_{x \neq y \in U} \frac{|f(x) - f(y)|}{|x-y|} \tag{1.68}$$

is finite. Clearly any Hölder continuous function is continuous and in the special case $\gamma = 1$ we obtain the **Lipschitz continuous** functions.

It is easy to verify that this is a seminorm and that the corresponding space is complete.

Theorem 1.33. *The space $C_b^{k,\gamma}(U)$ of all functions whose partial derivatives up to order k are bounded and Hölder continuous with exponent $\gamma \in (0,1]$ form a Banach space with norm*

$$\|f\|_{\infty,k,\gamma} = \|f\|_{\infty,k} + \sum_{|\alpha|=k} [\partial_\alpha f]_\gamma. \tag{1.69}$$

Note that by the mean value theorem all derivatives up to order lower than k are automatically Lipschitz continuous.

While the above spaces are able to cover a wide variety of situations, there are still cases where the above definitions are not suitable. For example, suppose you want to consider continuous function $C(\mathbb{R})$ but you do not want to assume boundedness. Then you could try to use local uniform convergence as follows. Consider the seminorms

$$\|f\|_j = \sup_{|x| \le j} |f(x)|, \qquad j \in \mathbb{N}, \tag{1.70}$$

and introduce a corresponding metric

$$d(f,g) = \sum_{j \in \mathbb{N}} \frac{1}{2^j} \frac{\|f-g\|_j}{1 + \|f-g\|_j}. \tag{1.71}$$

Then $f_k \to f$ if and only if $\|f_k - f\|_j \to 0$ for all $j \in \mathbb{N}$ (show this). Moreover, as above it follows that $C(\mathbb{R})$ is complete, but nevertheless we do not get a norm and hence no Banach space from this approach. However, the metric (1.71) has a a few additional properties: It is translation invariant

$$d(f,g) = d(f-h, g-h) \tag{1.72}$$

and convex

$$d(x, \lambda y + (1-\lambda)z) \le \lambda d(x,y) + (1-\lambda) d(x,z), \qquad \lambda \in [0,1]. \tag{1.73}$$

In general, a vector space X with a countable family of seminorms $\|.\|_j$ such that $\|f\|_j = 0$ for all j implies $f = 0$ is called a **Fréchet space** if it is complete with respect to the metric (1.71).

Example. The space $C^\infty(\mathbb{R}^m)$ together with the seminorms

$$\|f\|_{j,k} = \sup_{|x| \le k} |f^{(j)}(x)|, \qquad j \in \mathbb{N}_0,\ k \in \mathbb{N}, \tag{1.74}$$

is a **Fréchet space**. ◇

Example. The space of all entire functions $f(z)$ (i.e. functions which are holomorphic on all of $\mathbb{C}$) together with the seminorms $\|f\|_k = \sup_{|z|\le k} |f(x)|$, $k \in \mathbb{N}$, is a Fréchet space. Completeness follows from the Weierstraß convergence theorem which states that a limit of holomorphic functions which is uniform on every compact subset is again holomorphic. ◇

Note that in all the above spaces we could replace complex-valued by $\mathbb{C}^n$-valued functions.

Problem 1.33. *Suppose X is a vector space and $\|.\|_j$, $1 \le j \le n$ is a finite family of seminorms. Show that $\|x\| = \sum_{j=1}^{b} \|x\|_j$ is a seminorm. It is a norm if and only if $\|x\|_j = 0$ for all j implies $x = 0$.*

Problem 1.34. *Show $C_b^{k,\gamma_2}(U) \subseteq C_b^{k,\gamma_1}(U) \subseteq C_b^k(U)$ for $0 < \gamma_1 < \gamma_2 \le 1$.*

Problem 1.35. *Suppose X is a metric vector space whose metric is convex. Then every ball is convex (i.e, $x, y \in B$ implies , $\lambda x + (1-\lambda) y \in B$ for every $\lambda \in [0,1]$).*

Problem 1.36. *Show that if d is a pseudometric, then so is $\frac{d}{1+d}$. Show that if d is convex so is $\frac{d}{1+d}$. (Hint: Note that $f(x) = x/(1+x)$ is monotone.)*

Problem 1.37. *Let X be some space together with a sequence of pseudometrics d_j, $j \in \mathbb{N}$. Show that*

$$d(x,y) = \sum_{j\in\mathbb{N}} \frac{1}{2^j} \frac{d_j(x,y)}{1 + d_j(x,y)}$$

is again a pseudometric. It is a metric if and only if $d_j(x,y) = 0$ for all j implies $x = y$.

Chapter 2

Hilbert spaces

2.1. Orthonormal bases

In this section we will investigate orthonormal series and you will notice hardly any difference between the finite and infinite dimensional cases. Throughout this chapter $\mathfrak{H}$ will be a Hilbert space.

As our first task, let us generalize the projection into the direction of one vector:

A set of vectors $\{u_j\}$ is called an **orthonormal set** if $\langle u_j, u_k\rangle = 0$ for $j \neq k$ and $\langle u_j, u_j\rangle = 1$. Note that every orthonormal set is linearly independent (show this).

Lemma 2.1. *Suppose $\{u_j\}_{j=1}^n$ is an orthonormal set. Then every $f \in \mathfrak{H}$ can be written as*

$$f = f_\parallel + f_\perp, \qquad f_\parallel = \sum_{j=1}^n \langle u_j, f\rangle u_j, \tag{2.1}$$

where $f_\parallel$ and $f_\perp$ are orthogonal. Moreover, $\langle u_j, f_\perp\rangle = 0$ for all $1 \le j \le n$. In particular,

$$\|f\|^2 = \sum_{j=1}^n |\langle u_j, f\rangle|^2 + \|f_\perp\|^2. \tag{2.2}$$

Moreover, every $\hat{f}$ in the span of $\{u_j\}_{j=1}^n$ satisfies

$$\|f - \hat{f}\| \ge \|f_\perp\| \tag{2.3}$$

with equality holding if and only if $\hat{f} = f_\parallel$. In other words, $f_\parallel$ is uniquely characterized as the vector in the span of $\{u_j\}_{j=1}^n$ closest to f.

Proof. A straightforward calculation shows $\langle u_j, f - f_\parallel\rangle = 0$ and hence $f_\parallel$ and $f_\perp = f - f_\parallel$ are orthogonal. The formula for the norm follows by applying (1.39) iteratively.

Now, fix a vector

$$\hat{f} = \sum_{j=1}^{n} \alpha_j u_j$$

in the span of $\{u_j\}_{j=1}^n$. Then one computes

$$\begin{aligned}\|f - \hat{f}\|^2 &= \|f_\parallel + f_\perp - \hat{f}\|^2 = \|f_\perp\|^2 + \|f_\parallel - \hat{f}\|^2\\ &= \|f_\perp\|^2 + \sum_{j=1}^{n} |\alpha_j - \langle u_j, f\rangle|^2\end{aligned}$$

from which the last claim follows. □

From (2.2) we obtain **Bessel's inequality**

$$\sum_{j=1}^{n} |\langle u_j, f\rangle|^2 \le \|f\|^2 \tag{2.4}$$

with equality holding if and only if f lies in the span of $\{u_j\}_{j=1}^n$.

Of course, since we cannot assume $\mathfrak{H}$ to be a finite dimensional vector space, we need to generalize Lemma 2.1 to arbitrary orthonormal sets $\{u_j\}_{j\in J}$. We start by assuming that J is countable. Then Bessel's inequality (2.4) shows that

$$\sum_{j\in J} |\langle u_j, f\rangle|^2 \tag{2.5}$$

converges absolutely. Moreover, for any finite subset $K \subset J$ we have

$$\|\sum_{j\in K} \langle u_j, f\rangle u_j\|^2 = \sum_{j\in K} |\langle u_j, f\rangle|^2 \tag{2.6}$$

by the Pythagorean theorem and thus $\sum_{j\in J}\langle u_j, f\rangle u_j$ is a Cauchy sequence if and only if $\sum_{j\in J} |\langle u_j, f\rangle|^2$ is. Now let J be arbitrary. Again, Bessel's inequality shows that for any given $\varepsilon > 0$ there are at most finitely many j for which $|\langle u_j, f\rangle| \ge \varepsilon$ (namely at most $\|f\|/\varepsilon$). Hence there are at most countably many j for which $|\langle u_j, f\rangle| > 0$. Thus it follows that

$$\sum_{j\in J} |\langle u_j, f\rangle|^2 \tag{2.7}$$

is well-defined and (by completeness) so is

$$\sum_{j\in J} \langle u_j, f\rangle u_j. \tag{2.8}$$

Furthermore, it is also independent of the order of summation.

In particular, by continuity of the scalar product we see that Lemma 2.1 can be generalized to arbitrary orthonormal sets.

Theorem 2.2. *Suppose $\{u_j\}_{j\in J}$ is an orthonormal set in a Hilbert space $\mathfrak{H}$. Then every $f \in \mathfrak{H}$ can be written as*

$$f = f_\parallel + f_\perp, \qquad f_\parallel = \sum_{j\in J} \langle u_j, f\rangle u_j, \tag{2.9}$$

where $f_\parallel$ and $f_\perp$ are orthogonal. Moreover, $\langle u_j, f_\perp\rangle = 0$ for all $j \in J$. In particular,

$$\|f\|^2 = \sum_{j\in J} |\langle u_j, f\rangle|^2 + \|f_\perp\|^2. \tag{2.10}$$

Furthermore, every $\hat{f} \in \overline{\operatorname{span}\{u_j\}_{j\in J}}$ satisfies

$$\|f - \hat{f}\| \geq \|f_\perp\| \tag{2.11}$$

with equality holding if and only if $\hat{f} = f_\parallel$. In other words, $f_\parallel$ is uniquely characterized as the vector in $\overline{\operatorname{span}\{u_j\}_{j\in J}}$ closest to f.

Proof. The first part follows as in Lemma 2.1 using continuity of the scalar product. The same is true for the last part except for the fact that every $f \in \overline{\operatorname{span}\{u_j\}_{j\in J}}$ can be written as $f = \sum_{j\in J} \alpha_j u_j$ (i.e., $f = f_\parallel$). To see this let $f_n \in \operatorname{span}\{u_j\}_{j\in J}$ converge to f. Then $\|f - f_n\|^2 = \|f_\parallel - f_n\|^2 + \|f_\perp\|^2 \to 0$ implies $f_n \to f_\parallel$ and $f_\perp = 0$. □

Note that from Bessel's inequality (which of course still holds) it follows that the map $f \to f_\parallel$ is continuous.

Of course we are particularly interested in the case where every $f \in \mathfrak{H}$ can be written as $\sum_{j\in J} \langle u_j, f\rangle u_j$. In this case we will call the orthonormal set $\{u_j\}_{j\in J}$ an **orthonormal basis** (ONB).

If $\mathfrak{H}$ is separable it is easy to construct an orthonormal basis. In fact, if $\mathfrak{H}$ is separable, then there exists a countable total set $\{f_j\}_{j=1}^N$. Here $N \in \mathbb{N}$ if $\mathfrak{H}$ is finite dimensional and $N = \infty$ otherwise. After throwing away some vectors, we can assume that f_{n+1} cannot be expressed as a linear combination of the vectors $f_1, \dots, f_n$. Now we can construct an orthonormal set as follows: We begin by normalizing f_1:

$$u_1 = \frac{f_1}{\|f_1\|}. \tag{2.12}$$

Next we take f_2 and remove the component parallel to u_1 and normalize again:

$$u_2 = \frac{f_2 - \langle u_1, f_2\rangle u_1}{\|f_2 - \langle u_1, f_2\rangle u_1\|}. \tag{2.13}$$

Proceeding like this, we define recursively

$$u_n = \frac{f_n - \sum_{j=1}^{n-1} \langle u_j, f_n\rangle u_j}{\|f_n - \sum_{j=1}^{n-1} \langle u_j, f_n\rangle u_j\|}. \tag{2.14}$$

This procedure is known as **Gram–Schmidt orthogonalization**. Hence we obtain an orthonormal set $\{u_j\}_{j=1}^N$ such that $\mathrm{span}\{u_j\}_{j=1}^n = \mathrm{span}\{f_j\}_{j=1}^n$ for any finite n and thus also for $n = N$ (if $N = \infty$). Since $\{f_j\}_{j=1}^N$ is total, so is $\{u_j\}_{j=1}^N$. Now suppose there is some $f = f_\| + f_\perp \in \mathfrak{H}$ for which $f_\perp \neq 0$. Since $\{u_j\}_{j=1}^N$ is total, we can find a $\hat{f}$ in its span, such that $\|f - \hat{f}\| < \|f_\perp\|$ contradicting (2.11). Hence we infer that $\{u_j\}_{j=1}^N$ is an orthonormal basis.

Theorem 2.3. *Every separable Hilbert space has a countable orthonormal basis.*

Example. In $\mathcal{L}^2_{cont}(-1,1)$ we can orthogonalize the polynomial $f_n(x) = x^n$ (which are total by the Weierstraß approximation theorem — Theorem 1.20) The resulting polynomials are up to a normalization equal to the Legendre polynomials

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{3\,x^2 - 1}{2}, \quad \dots \tag{2.15}$$

(which are normalized such that $P_n(1) = 1$). ⋄

Example. The set of functions

$$u_n(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\mathrm{i}nx}, \qquad n \in \mathbb{Z}, \tag{2.16}$$

forms an orthonormal basis for $\mathfrak{H} = \mathcal{L}^2_{cont}(0, 2\pi)$. The corresponding orthogonal expansion is just the ordinary Fourier series. (That these functions are total will follow from the Stone–Weierstraß theorem — Problem 6.10) ⋄

The following equivalent properties also characterize a basis.

Theorem 2.4. *For an orthonormal set $\{u_j\}_{j\in J}$ in a Hilbert space $\mathfrak{H}$ the following conditions are equivalent:*

(i) *$\{u_j\}_{j\in J}$ is a maximal orthogonal set.*

(ii) *For every vector $f \in \mathfrak{H}$ we have*

$$f = \sum_{j\in J} \langle u_j, f\rangle u_j. \tag{2.17}$$

(iii) *For every vector $f \in \mathfrak{H}$ we have* **Parseval's relation**

$$\|f\|^2 = \sum_{j\in J} |\langle u_j, f\rangle|^2. \tag{2.18}$$

(iv) *$\langle u_j, f\rangle = 0$ for all $j \in J$ implies $f = 0$.*

Proof. We will use the notation from Theorem 2.2.
(i) ⇒ (ii): If $f_\perp \neq 0$, then we can normalize $f_\perp$ to obtain a unit vector $\tilde{f}_\perp$ which is orthogonal to all vectors u_j. But then $\{u_j\}_{j\in J} \cup \{\tilde{f}_\perp\}$ would be a larger orthonormal set, contradicting the maximality of $\{u_j\}_{j\in J}$.
(ii) ⇒ (iii): This follows since (ii) implies $f_\perp = 0$.
(iii) ⇒ (iv): If $\langle f, u_j\rangle = 0$ for all $j \in J$, we conclude $\|f\|^2 = 0$ and hence $f = 0$.
(iv) ⇒ (i): If $\{u_j\}_{j\in J}$ were not maximal, there would be a unit vector g such that $\{u_j\}_{j\in J} \cup \{g\}$ is a larger orthonormal set. But $\langle u_j, g\rangle = 0$ for all $j \in J$ implies $g = 0$ by (iv), a contradiction. □

By continuity of the norm it suffices to check (iii), and hence also (ii), for f in a dense set.

It is not surprising that if there is one countable basis, then it follows that every other basis is countable as well.

Theorem 2.5. *In a Hilbert space $\mathfrak{H}$ every orthonormal basis has the same cardinality.*

Proof. Without loss of generality we assume that $\mathfrak{H}$ is infinite dimensional. Let $\{u_j\}_{j\in J}$ and $\{v_k\}_{k\in K}$ be two orthonormal bases. Set $K_j = \{k \in K | \langle v_k, u_j\rangle \neq 0\}$. Since these are the expansion coefficients of u_j with respect to $\{v_k\}_{k\in K}$, this set is countable. Hence the set $\tilde{K} = \bigcup_{j\in J} K_j$ has the same cardinality as J. But $k \in K\backslash\tilde{K}$ implies $v_k = 0$ and hence $\tilde{K} = K$. □

The cardinality of an orthonormal basis is also called the Hilbert space **dimension** of $\mathfrak{H}$.

It even turns out that, up to unitary equivalence, there is only one separable infinite dimensional Hilbert space:

A bijective linear operator $U \in \mathfrak{L}(\mathfrak{H}_1, \mathfrak{H}_2)$ is called **unitary** if U preserves scalar products:

$$\langle Ug, Uf\rangle_2 = \langle g, f\rangle_1, \qquad g, f \in \mathfrak{H}_1. \tag{2.19}$$

By the polarization identity (1.46) this is the case if and only if U preserves norms: $\|Uf\|_2 = \|f\|_1$ for all $f \in \mathfrak{H}_1$ (note the a norm preserving linear operator is automatically injective). The two Hilbert space $\mathfrak{H}_1$ and $\mathfrak{H}_2$ are called **unitarily equivalent** in this case.

Let $\mathfrak{H}$ be an infinite dimensional Hilbert space and let $\{u_j\}_{j\in\mathbb{N}}$ be any orthogonal basis. Then the map $U : \mathfrak{H} \to \ell^2(\mathbb{N})$, $f \mapsto (\langle u_j, f\rangle)_{j\in\mathbb{N}}$ is unitary (by Theorem 2.4 (ii) it is onto and by (iii) it is norm preserving). In particular,

Theorem 2.6. *Any separable infinite dimensional Hilbert space is unitarily equivalent to $\ell^2(\mathbb{N})$.*

Finally we briefly turn to the case where $\mathfrak{H}$ is not separable.

Theorem 2.7. *Every Hilbert space has an orthonormal basis.*

Proof. To prove this we need to resort to Zorn's lemma (see Appendix A): The collection of all orthonormal sets in $\mathfrak{H}$ can be partially ordered by inclusion. Moreover, every linearly ordered chain has an upper bound (the union of all sets in the chain). Hence a fundamental result from axiomatic set theory, Zorn's lemma, implies the existence of a maximal element, that is, an orthonormal set which is not a proper subset of every other orthonormal set. □

Hence, if $\{u_j\}_{j\in J}$ is an orthogonal basis, we can show that $\mathfrak{H}$ is unitarily equivalent to $\ell^2(J)$ and, by prescribing J, we can find an Hilbert space of any given dimension.

Problem 2.1. *Let $\{u_j\}$ be some orthonormal basis. Show that a bounded linear operator A is uniquely determined by its matrix elements $A_{jk} = \langle u_j, Au_k\rangle$ with respect to this basis.*

2.2. The projection theorem and the Riesz lemma

Let $M \subseteq \mathfrak{H}$ be a subset. Then $M^\perp = \{f | \langle g, f\rangle = 0, \forall g \in M\}$ is called the **orthogonal complement** of M. By continuity of the scalar product it follows that $M^\perp$ is a closed linear subspace and by linearity that $(\overline{\mathrm{span}(M)})^\perp = M^\perp$. For example we have $\mathfrak{H}^\perp = \{0\}$ since any vector in $\mathfrak{H}^\perp$ must be in particular orthogonal to all vectors in some orthonormal basis.

Theorem 2.8 (Projection theorem)**.** *Let M be a closed linear subspace of a Hilbert space $\mathfrak{H}$. Then every $f \in \mathfrak{H}$ can be uniquely written as $f = f_\parallel + f_\perp$ with $f_\parallel \in M$ and $f_\perp \in M^\perp$. One writes*

$$M \oplus M^\perp = \mathfrak{H} \tag{2.20}$$

in this situation.

Proof. Since M is closed, it is a Hilbert space and has an orthonormal basis $\{u_j\}_{j\in J}$. Hence the existence part follows from Theorem 2.2. To see uniqueness suppose there is another decomposition $f = \tilde{f}_\parallel + \tilde{f}_\perp$. Then $f_\parallel - \tilde{f}_\parallel = \tilde{f}_\perp - f_\perp \in M \cap M^\perp = \{0\}$. □

Corollary 2.9. *Every orthogonal set $\{u_j\}_{j\in J}$ can be extended to an orthogonal basis.*

Proof. Just add an orthogonal basis for $(\{u_j\}_{j\in J})^\perp$. □

Moreover, Theorem 2.8 implies that to every $f \in \mathfrak{H}$ we can assign a unique vector $f_\parallel$ which is the vector in M closest to f. The rest, $f - f_\parallel$, lies in $M^\perp$. The operator $P_M f = f_\parallel$ is called the **orthogonal projection** corresponding to M. Note that we have

$$P_M^2 = P_M \qquad \text{and} \qquad \langle P_M g, f\rangle = \langle g, P_M f\rangle \tag{2.21}$$

since $\langle P_M g, f\rangle = \langle g_\parallel, f_\parallel\rangle = \langle g, P_M f\rangle$. Clearly we have $P_{M^\perp} f = f - P_M f = f_\perp$. Furthermore, (2.21) uniquely characterizes orthogonal projections (Problem 2.4).

Moreover, if M is a closed subspace we have $P_{M^{\perp\perp}} = \mathbb{I} - P_{M^\perp} = \mathbb{I} - (\mathbb{I} - P_M) = P_M$, that is, $M^{\perp\perp} = M$. If M is an arbitrary subset, we have at least

$$M^{\perp\perp} = \overline{\operatorname{span}(M)}. \tag{2.22}$$

Note that by $\mathfrak{H}^\perp = \{0\}$ we see that $M^\perp = \{0\}$ if and only if M is total.

Finally we turn to **linear functionals**, that is, to operators $\ell : \mathfrak{H} \to \mathbb{C}$. By the Cauchy–Schwarz inequality we know that $\ell_g : f \mapsto \langle g, f\rangle$ is a bounded linear functional (with norm $\|g\|$). In turns out that in a Hilbert space every bounded linear functional can be written in this way.

Theorem 2.10 (Riesz lemma)**.** *Suppose ℓ is a bounded linear functional on a Hilbert space $\mathfrak{H}$. Then there is a unique vector $g \in \mathfrak{H}$ such that $\ell(f) = \langle g, f\rangle$ for all $f \in \mathfrak{H}$.*

In other words, a Hilbert space is equivalent to its own dual space $\mathfrak{H}^* \cong \mathfrak{H}$ via the map $f \mapsto \langle f, .\rangle$ which is a conjugate linear isometric bijection between $\mathfrak{H}$ and $\mathfrak{H}^*$.

Proof. If $\ell \equiv 0$, we can choose $g = 0$. Otherwise $\operatorname{Ker}(\ell) = \{f | \ell(f) = 0\}$ is a proper subspace and we can find a unit vector $\tilde{g} \in \operatorname{Ker}(\ell)^\perp$. For every $f \in \mathfrak{H}$ we have $\ell(f)\tilde{g} - \ell(\tilde{g})f \in \operatorname{Ker}(\ell)$ and hence

$$0 = \langle \tilde{g}, \ell(f)\tilde{g} - \ell(\tilde{g})f\rangle = \ell(f) - \ell(\tilde{g})\langle \tilde{g}, f\rangle.$$

In other words, we can choose $g = \ell(\tilde{g})^*\tilde{g}$. To see uniqueness, let g_1, g_2 be two such vectors. Then $\langle g_1 - g_2, f\rangle = \langle g_1, f\rangle - \langle g_2, f\rangle = \ell(f) - \ell(f) = 0$ for every $f \in \mathfrak{H}$, which shows $g_1 - g_2 \in \mathfrak{H}^\perp = \{0\}$. □

Problem 2.2. *Suppose $U : \mathfrak{H} \to \mathfrak{H}$ is unitary and $M \subseteq \mathfrak{H}$. Show that $UM^\perp = (UM)^\perp$.*

Problem 2.3. *Show that an orthogonal projection $P_M \neq 0$ has norm one.*

Problem 2.4. *Suppose $P \in \mathfrak{L}(\mathfrak{H})$ satisfies*

$$P^2 = P \qquad \textit{and} \qquad \langle Pf, g\rangle = \langle f, Pg\rangle$$

and set $M = \operatorname{Ran}(P)$. Show

- *$Pf = f$ for $f \in M$ and M is closed,*
- *$g \in M^\perp$ implies $Pg \in M^\perp$ and thus $Pg = 0$,*

and conclude $P = P_M$.

2.3. Operators defined via forms

In many situations operators are not given explicitly, but implicitly via their associated sesquilinear forms $\langle f, Ag\rangle$. As an easy consequence of the Riesz lemma, Theorem 2.10, we obtain that there is a one to one correspondence between bounded operators and bounded sesquilinear forms:

Lemma 2.11. *Suppose s is a bounded sesquilinear form; that is,*

$$|s(f,g)| \le C\|f\|\,\|g\|. \tag{2.23}$$

Then there is a unique bounded operator A such that

$$s(f,g) = \langle f, Ag\rangle. \tag{2.24}$$

Moreover, the norm of A is given by

$$\|A\| = \sup_{\|f\|=\|g\|=1} |\langle f, Ag\rangle| \le C. \tag{2.25}$$

Proof. For every $g \in \mathfrak{H}$ we have an associated bounded linear functional $\ell_g(f) = s(f,g)^*$. By Theorem 2.10 there is a corresponding $h \in \mathfrak{H}$ (depending on g) such that $\ell_g(f) = \langle h, f\rangle$, that is $s(f,g) = \langle f, h\rangle$ and we can define A via $Ag = h$. It is not hard to check that A is linear and from

$$\|Af\|^2 = \langle Af, Af\rangle = s(Af, f) \le C\|Af\|\|f\|$$

we infer $\|Af\| \le C\|f\|$, which shows that A is bounded with $\|A\| \le C$. Equation (2.25) is left as an exercise (Problem 2.6). □

Note that by the polarization identity (Problem 1.20), A is already uniquely determined by its quadratic form $q_A(f) = \langle f, Af\rangle$.

As a first application we introduce the **adjoint operator** via Lemma 2.11 as the operator associated with the sesquilinear form $s(f,g) = \langle Af, g\rangle$.

Theorem 2.12. *For every bounded operator $A \in \mathfrak{L}(\mathfrak{H})$ there is a unique bounded operator A^* defined via*

$$\langle f, A^* g\rangle = \langle Af, g\rangle. \tag{2.26}$$

Example. If $\mathfrak{H} = \mathbb{C}^n$ and $A = (a_{jk})_{1\le j,k\le n}$, then $A^* = (a_{kj}^*)_{1\le j,k\le n}$. ⋄

Example. Suppose $U \in \mathfrak{L}(\mathfrak{H})$ is unitary. Then $U^* = U^{-1}$. This follows from Lemma 2.11 since $\langle f, g\rangle = \langle Uf, Ug\rangle = \langle f, U^*Ug\rangle$ implies $U^*U = \mathbb{I}$. Since U is bijective we can multiply this last equation from the right with U^{-1} to obtain the claim. ⋄

A few simple properties of taking adjoints are listed below.

Lemma 2.13. *Let $A, B \in \mathfrak{L}(\mathfrak{H})$ and $\alpha \in \mathbb{C}$. Then*

(i) $(A+B)^* = A^* + B^*$, $\quad (\alpha A)^* = \alpha^* A^*$,

(ii) $A^{**} = A$,

(iii) $(AB)^* = B^* A^*$,

(iv) $\|A^*\| = \|A\|$ *and* $\|A\|^2 = \|A^* A\| = \|A A^*\|$.

Proof. (i) is obvious. (ii) follows from $\langle f, A^{**} g\rangle = \langle A^* f, g\rangle = \langle f, Ag\rangle$. (iii) follows from $\langle f, (AB)g\rangle = \langle A^* f, Bg\rangle = \langle B^* A^* f, g\rangle$. (iv) follows using (2.25) from

$$\begin{aligned} \|A^*\| &= \sup_{\|f\|=\|g\|=1} |\langle f, A^* g\rangle| = \sup_{\|f\|=\|g\|=1} |\langle Af, g\rangle| \\ &= \sup_{\|f\|=\|g\|=1} |\langle g, Af\rangle| = \|A\| \end{aligned}$$

and

$$\begin{aligned} \|A^* A\| &= \sup_{\|f\|=\|g\|=1} |\langle f, A^* A g\rangle| = \sup_{\|f\|=\|g\|=1} |\langle Af, Ag\rangle| \\ &= \sup_{\|f\|=1} \|Af\|^2 = \|A\|^2, \end{aligned}$$

where we have used that $|\langle Af, Ag\rangle|$ attains its maximum when Af and Ag are parallel (compare Theorem 1.22). □

Note that $\|A\| = \|A^*\|$ implies that taking adjoints is a continuous operation. For later use also note that (Problem 2.8)

$$\operatorname{Ker}(A^*) = \operatorname{Ran}(A)^\perp. \tag{2.27}$$

A sesquilinear form is called nonnegative if $s(f,f) \geq 0$ and we will call A **nonnegative**, $A \geq 0$, if its associated sesquilinear form is. We will write $A \geq B$ if $A - B \geq 0$.

Lemma 2.14. *Suppose $A \geq \varepsilon \mathbb{I}$ for some $\varepsilon > 0$. Then A is a bijection with bounded inverse, $\|A^{-1}\| \leq \frac{1}{\varepsilon}$.*

Proof. By definition $\varepsilon\|f\|^2 \leq \langle f, Af\rangle \leq \|f\|\|Af\|$ and thus $\varepsilon\|f\| \leq \|Af\|$. In particular, $Af = 0$ implies $f = 0$ and thus for every $g \in \operatorname{Ran}(A)$ there is a unique $f = A^{-1} g$. Moreover, by $\|A^{-1} g\| = \|f\| \leq \varepsilon^{-1}\|Af\| = \varepsilon^{-1}\|g\|$ the operator A^{-1} is bounded. So if $g_n \in \operatorname{Ran}(A)$ converges to some $g \in \mathfrak{H}$, then $f_n = A^{-1} g_n$ converges to some f. Taking limits in $g_n = A f_n$ shows that $g = Af$ is in the range of A, that is, the range of A is closed. To show that $\operatorname{Ran}(A) = \mathfrak{H}$ we pick $h \in \operatorname{Ran}(A)^\perp$. Then $0 = \langle h, Ah\rangle \geq \varepsilon\|h\|^2$ shows $h = 0$ and thus $\operatorname{Ran}(A)^\perp = \{0\}$. □

Combining the last two results we obtain the famous Lax–Milgram theorem which plays an important role in theory of elliptic partial differential equations.

Theorem 2.15 (Lax–Milgram). *Let s be a sesquilinear form which is*

- *bounded, $|s(f,g)| \le C\|f\|\,\|g\|$, and*
- *coercive, $s(f,f) \ge \varepsilon\|f\|^2$.*

Then for every $g \in \mathfrak{H}$ there is a unique $f \in \mathfrak{H}$ such that

$$s(h,f) = \langle h, g\rangle, \qquad \forall h \in \mathfrak{H}. \tag{2.28}$$

Proof. Let A be the operator associated with s. Then $A \ge \varepsilon$ and $f = A^{-1}g$. □

Problem 2.5. *Let $\mathfrak{H}$ a Hilbert space and let $u, v \in \mathfrak{H}$. Show that the operator*

$$Af = \langle u, f\rangle v$$

is bounded and compute its norm. Compute the adjoint of A.

Problem 2.6. *Prove (2.25). (Hint: Use $\|f\| = \sup_{\|g\|=1} |\langle g, f\rangle|$ — compare Theorem 1.22.)*

Problem 2.7. *Suppose A has a bounded inverse A^{-1}. Show $(A^{-1})^* = (A^*)^{-1}$.*

Problem 2.8. *Show (2.27).*

2.4. Orthogonal sums and tensor products

Given two Hilbert spaces $\mathfrak{H}_1$ and $\mathfrak{H}_2$, we define their **orthogonal sum** $\mathfrak{H}_1 \oplus \mathfrak{H}_2$ to be the set of all pairs $(f_1, f_2) \in \mathfrak{H}_1 \times \mathfrak{H}_2$ together with the scalar product

$$\langle (g_1, g_2), (f_1, f_2)\rangle = \langle g_1, f_1\rangle_1 + \langle g_2, f_2\rangle_2. \tag{2.29}$$

It is left as an exercise to verify that $\mathfrak{H}_1 \oplus \mathfrak{H}_2$ is again a Hilbert space. Moreover, $\mathfrak{H}_1$ can be identified with $\{(f_1, 0) | f_1 \in \mathfrak{H}_1\}$ and we can regard $\mathfrak{H}_1$ as a subspace of $\mathfrak{H}_1 \oplus \mathfrak{H}_2$, and similarly for $\mathfrak{H}_2$. It is also customary to write $f_1 + f_2$ instead of (f_1, f_2).

More generally, let $\mathfrak{H}_j$, $j \in \mathbb{N}$, be a countable collection of Hilbert spaces and define

$$\bigoplus_{j=1}^{\infty} \mathfrak{H}_j = \{\sum_{j=1}^{\infty} f_j |\, f_j \in \mathfrak{H}_j,\ \sum_{j=1}^{\infty} \|f_j\|_j^2 < \infty\}, \tag{2.30}$$

which becomes a Hilbert space with the scalar product

$$\langle \sum_{j=1}^{\infty} g_j, \sum_{j=1}^{\infty} f_j\rangle = \sum_{j=1}^{\infty} \langle g_j, f_j\rangle_j. \tag{2.31}$$

Example. $\bigoplus_{j=1}^{\infty} \mathbb{C} = \ell^2(\mathbb{N})$. ◇

Similarly, if $\mathfrak{H}$ and $\tilde{\mathfrak{H}}$ are two Hilbert spaces, we define their tensor product as follows: The elements should be products $f \otimes \tilde{f}$ of elements $f \in \mathfrak{H}$ and $\tilde{f} \in \tilde{\mathfrak{H}}$. Hence we start with the set of all finite linear combinations of elements of $\mathfrak{H} \times \tilde{\mathfrak{H}}$

$$\mathcal{F}(\mathfrak{H}, \tilde{\mathfrak{H}}) = \{\sum_{j=1}^{n} \alpha_j (f_j, \tilde{f}_j) | (f_j, \tilde{f}_j) \in \mathfrak{H} \times \tilde{\mathfrak{H}},\ \alpha_j \in \mathbb{C}\}. \tag{2.32}$$

Since we want $(f_1 + f_2) \otimes \tilde{f} = f_1 \otimes \tilde{f} + f_2 \otimes \tilde{f}$, $f \otimes (\tilde{f}_1 + \tilde{f}_2) = f \otimes \tilde{f}_1 + f \otimes \tilde{f}_2$, and $(\alpha f) \otimes \tilde{f} = f \otimes (\alpha \tilde{f})$ we consider $\mathcal{F}(\mathfrak{H}, \tilde{\mathfrak{H}})/\mathcal{N}(\mathfrak{H}, \tilde{\mathfrak{H}})$, where

$$\mathcal{N}(\mathfrak{H}, \tilde{\mathfrak{H}}) = \operatorname{span}\{\sum_{j,k=1}^{n} \alpha_j \beta_k (f_j, \tilde{f}_k) - (\sum_{j=1}^{n} \alpha_j f_j, \sum_{k=1}^{n} \beta_k \tilde{f}_k)\} \tag{2.33}$$

and write $f \otimes \tilde{f}$ for the equivalence class of $(f, \tilde{f})$.

Next we define

$$\langle f \otimes \tilde{f}, g \otimes \tilde{g} \rangle = \langle f, g \rangle \langle \tilde{f}, \tilde{g} \rangle \tag{2.34}$$

which extends to a sesquilinear form on $\mathcal{F}(\mathfrak{H}, \tilde{\mathfrak{H}})/\mathcal{N}(\mathfrak{H}, \tilde{\mathfrak{H}})$. To show that we obtain a scalar product, we need to ensure positivity. Let $f = \sum_i \alpha_i f_i \otimes \tilde{f}_i \neq 0$ and pick orthonormal bases u_j, $\tilde{u}_k$ for $\operatorname{span}\{f_i\}$, $\operatorname{span}\{\tilde{f}_i\}$, respectively. Then

$$f = \sum_{j,k} \alpha_{jk} u_j \otimes \tilde{u}_k, \quad \alpha_{jk} = \sum_i \alpha_i \langle u_j, f_i \rangle \langle \tilde{u}_k, \tilde{f}_i \rangle \tag{2.35}$$

and we compute

$$\langle f, f \rangle = \sum_{j,k} |\alpha_{jk}|^2 > 0. \tag{2.36}$$

The completion of $\mathcal{F}(\mathfrak{H}, \tilde{\mathfrak{H}})/\mathcal{N}(\mathfrak{H}, \tilde{\mathfrak{H}})$ with respect to the induced norm is called the **tensor product** $\mathfrak{H} \otimes \tilde{\mathfrak{H}}$ of $\mathfrak{H}$ and $\tilde{\mathfrak{H}}$.

Lemma 2.16. *If u_j, $\tilde{u}_k$ are orthonormal bases for $\mathfrak{H}$, $\tilde{\mathfrak{H}}$, respectively, then $u_j \otimes \tilde{u}_k$ is an orthonormal basis for $\mathfrak{H} \otimes \tilde{\mathfrak{H}}$.*

Proof. That $u_j \otimes \tilde{u}_k$ is an orthonormal set is immediate from (2.34). Moreover, since $\operatorname{span}\{u_j\}$, $\operatorname{span}\{\tilde{u}_k\}$ are dense in $\mathfrak{H}$, $\tilde{\mathfrak{H}}$, respectively, it is easy to see that $u_j \otimes \tilde{u}_k$ is dense in $\mathcal{F}(\mathfrak{H}, \tilde{\mathfrak{H}})/\mathcal{N}(\mathfrak{H}, \tilde{\mathfrak{H}})$. But the latter is dense in $\mathfrak{H} \otimes \tilde{\mathfrak{H}}$. □

Example. We have $\mathfrak{H} \otimes \mathbb{C}^n = \mathfrak{H}^n$. ◇

It is straightforward to extend the tensor product to any finite number of Hilbert spaces. We even note

$$(\bigoplus_{j=1}^{\infty} \mathfrak{H}_j) \otimes \mathfrak{H} = \bigoplus_{j=1}^{\infty} (\mathfrak{H}_j \otimes \mathfrak{H}), \tag{2.37}$$

where equality has to be understood in the sense that both spaces are unitarily equivalent by virtue of the identification

$$(\sum_{j=1}^{\infty} f_j) \otimes f = \sum_{j=1}^{\infty} f_j \otimes f. \tag{2.38}$$

Problem 2.9. *Show that* $f \otimes \tilde{f} = 0$ *if and only if* $f = 0$ *or* $\tilde{f} = 0$.

Problem 2.10. *We have* $f \otimes \tilde{f} = g \otimes \tilde{g} \neq 0$ *if and only if there is some* $\alpha \in \mathbb{C}\backslash\{0\}$ *such that* $f = \alpha g$ *and* $\tilde{f} = \alpha^{-1} \tilde{g}$.

Problem 2.11. *Show* (2.37)

Chapter 3

Compact operators

3.1. Compact operators

A linear operator A defined on a normed space X is called **compact** if every sequence Af_n has a convergent subsequence whenever f_n is bounded. The set of all compact operators is denoted by $\mathfrak{C}(X)$. It is not hard to see that the set of compact operators is an ideal of the set of bounded operators (Problem 3.1):

Theorem 3.1. *Every compact linear operator is bounded. Linear combinations of compact operators are compact and the product of a bounded and a compact operator is again compact.*

If X is a Banach space then this ideal is even closed:

Theorem 3.2. *Let X be a Banach space, and let A_n be a convergent sequence of compact operators. Then the limit A is again compact.*

Proof. Let $f_j^{(0)}$ be a bounded sequence. Choose a subsequence $f_j^{(1)}$ such that $A_1 f_j^{(1)}$ converges. From $f_j^{(1)}$ choose another subsequence $f_j^{(2)}$ such that $A_2 f_j^{(2)}$ converges and so on. Since $f_j^{(n)}$ might disappear as $n \to \infty$, we consider the diagonal sequence $f_j = f_j^{(j)}$. By construction, f_j is a subsequence of $f_j^{(n)}$ for $j \geq n$ and hence $A_n f_j$ is Cauchy for every fixed n. Now

$$\begin{aligned} \|Af_j - Af_k\| &= \|(A - A_n)(f_j - f_k) + A_n(f_j - f_k)\| \\ &\leq \|A - A_n\| \|f_j - f_k\| + \|A_n f_j - A_n f_k\| \end{aligned}$$

shows that Af_j is Cauchy since the first term can be made arbitrary small by choosing n large and the second by the Cauchy property of $A_n f_j$. □

Note that it suffices to verify compactness on a dense set.

Theorem 3.3. *Let X be a normed space and $A \in \mathfrak{C}(X)$. Let $\overline{X}$ be its completion, then $\overline{A} \in \mathfrak{C}(\overline{X})$, where $\overline{A}$ is the unique extension of A.*

Proof. Let $f_n \in \overline{X}$ be a given bounded sequence. We need to show that $\overline{A} f_n$ has a convergent subsequence. Pick $f_n^j \in X$ such that $\|f_n^j - f_n\| \leq \frac{1}{j}$ and by compactness of A we can assume that $A f_n^n \to g$. But then $\|\overline{A} f_n - g\| \leq \|A\| \|f_n - f_n^n\| + \|A f_n^n - g\|$ shows that $\overline{A} f_n \to g$. □

One of the most important examples of compact operators are integral operators:

Lemma 3.4. *The integral operator*

$$(Kf)(x) = \int_a^b K(x,y) f(y) dy, \tag{3.1}$$

where $K(x,y) \in C([a,b] \times [a,b])$, defined on $\mathcal{L}^2_{cont}(a,b)$ is compact.

Proof. First of all note that $K(.,..)$ is continuous on $[a,b] \times [a,b]$ and hence uniformly continuous. In particular, for every $\varepsilon > 0$ we can find a $\delta > 0$ such that $|K(y,t) - K(x,t)| \leq \varepsilon$ whenever $|y - x| \leq \delta$. Let $g(x) = Kf(x)$. Then

$$\begin{aligned} |g(x) - g(y)| &\leq \int_a^b |K(y,t) - K(x,t)| \, |f(t)| dt \\ &\leq \varepsilon \int_a^b |f(t)| dt \leq \varepsilon \|1\| \, \|f\|, \end{aligned}$$

whenever $|y - x| \leq \delta$. Hence, if $f_n(x)$ is a bounded sequence in $\mathcal{L}^2_{cont}(a,b)$, then $g_n(x) = K f_n(x)$ is equicontinuous and has a uniformly convergent subsequence by the Arzelà–Ascoli theorem (Theorem 3.5 below). But a uniformly convergent sequence is also convergent in the norm induced by the scalar product. Therefore K is compact. □

Note that (almost) the same proof shows that K is compact when defined on $C[a,b]$.

Theorem 3.5 (Arzelà–Ascoli)**.** *Suppose the sequence of functions $f_n(x)$, $n \in \mathbb{N}$, in $C[a,b]$ is (uniformly) equicontinuous, that is, for every $\varepsilon > 0$ there is a $\delta > 0$ (independent of n) such that*

$$|f_n(x) - f_n(y)| \leq \varepsilon \quad \text{if} \quad |x - y| < \delta. \tag{3.2}$$

If the sequence f_n is bounded, then there is a uniformly convergent subsequence.

Proof. Let $\{x_j\}_{j=1}^\infty$ be a dense subset of our interval (e.g., all rational numbers in $[a,b]$). Since $f_n(x_1)$ is bounded, we can choose a subsequence $f_n^{(1)}(x)$ such that $f_n^{(1)}(x_1)$ converges (Bolzano–Weierstraß). Similarly we can extract a subsequence $f_n^{(2)}(x)$ from $f_n^{(1)}(x)$ which converges at x_2 (and hence also at x_1 since it is a subsequence of $f_n^{(1)}(x)$). By induction we get a sequence $f_n^{(j)}(x)$ converging at $x_1, \dots, x_j$. The diagonal sequence $\tilde{f}_n(x) = f_n^{(n)}(x)$ will hence converge for all $x = x_j$ (why?). We will show that it converges uniformly for all x:

Fix $\varepsilon > 0$ and chose δ such that $|f_n(x) - f_n(y)| \le \frac{\varepsilon}{3}$ for $|x-y| < \delta$. The balls $B_\delta(x_j)$ cover $[a,b]$ and by compactness even finitely many, say $1 \le j \le p$, suffice. Furthermore, choose N_ε such that $|\tilde{f}_m(x_j) - \tilde{f}_n(x_j)| \le \frac{\varepsilon}{3}$ for $n, m \ge N_\varepsilon$ and $1 \le j \le p$.

Now pick x and note that $x \in B_\delta(x_j)$ for some j. Thus

$$\begin{aligned} |\tilde{f}_m(x) - \tilde{f}_n(x)| \le & |\tilde{f}_m(x) - \tilde{f}_m(x_j)| + |\tilde{f}_m(x_j) - \tilde{f}_n(x_j)| \\ & + |\tilde{f}_n(x_j) - \tilde{f}_n(x)| \le \varepsilon \end{aligned}$$

for $n, m \ge N_\varepsilon$, which shows that $\tilde{f}_n$ is Cauchy with respect to the maximum norm. By completeness of $C[a,b]$ it has a limit. □

Compact operators are very similar to (finite) matrices as we will see in the next section.

Problem 3.1. *Show that compact operators form an ideal.*

Problem 3.2. *Show that adjoint of the integral operator from Lemma 3.4 is the integral operator with kernel $K(y,x)^*$.*

3.2. The spectral theorem for compact symmetric operators

Let $\mathfrak{H}$ be a Hilbert space. A linear operator A is called **symmetric** if its domain is dense and if

$$\langle g, Af\rangle = \langle Ag, f\rangle \qquad f, g \in \mathfrak{D}(A). \tag{3.3}$$

If A is bounded (with $\mathfrak{D}(A) = \mathfrak{H}$), then A is symmetric precisely if $A = A^*$, that is, if A is **self-adjoint**. However, for unbounded operators there is a subtle but important difference between symmetry and self-adjointness.

A number $z \in \mathbb{C}$ is called **eigenvalue** of A if there is a nonzero vector $u \in \mathfrak{D}(A)$ such that

$$Au = zu. \tag{3.4}$$

The vector u is called a corresponding **eigenvector** in this case. The set of all eigenvectors corresponding to z is called the **eigenspace**

$$\operatorname{Ker}(A - z) \tag{3.5}$$

corresponding to z. Here we have used the shorthand notation $A-z$ for $A-z\mathbb{I}$. An eigenvalue is called **simple** if there is only one linearly independent eigenvector.

Theorem 3.6. *Let A be symmetric. Then all eigenvalues are real and eigenvectors corresponding to different eigenvalues are orthogonal.*

Proof. Suppose λ is an eigenvalue with corresponding normalized eigenvector u. Then $\lambda = \langle u, Au\rangle = \langle Au, u\rangle = \lambda^*$, which shows that λ is real. Furthermore, if $Au_j = \lambda_j u_j$, $j = 1, 2$, we have

$$(\lambda_1 - \lambda_2)\langle u_1, u_2\rangle = \langle Au_1, u_2\rangle - \langle u_1, Au_2\rangle = 0$$

finishing the proof. □

Note that while eigenvectors corresponding to the same eigenvalue λ will in general not automatically be orthogonal, we can of course replace each set of eigenvectors corresponding to λ by an set of orthonormal eigenvectors having the same linear span (e.g. using Gram–Schmidt orthogonalization).

Now we show that A has an eigenvalue at all (which is not clear in the infinite dimensional case)!

Theorem 3.7. *A symmetric compact operator A has an eigenvalue α_1 which satisfies $|\alpha_1| = \|A\|$.*

Proof. We set $\alpha = \|A\|$ and assume $\alpha \neq 0$ (i.e, $A \neq 0$) without loss of generality. Since

$$\|A\|^2 = \sup_{f:\|f\|=1} \|Af\|^2 = \sup_{f:\|f\|=1} \langle Af, Af\rangle = \sup_{f:\|f\|=1} \langle f, A^2 f\rangle$$

there exists a normalized sequence u_n such that

$$\lim_{n\to\infty} \langle u_n, A^2 u_n\rangle = \alpha^2.$$

Since A is compact, it is no restriction to assume that $A^2 u_n$ converges, say $\lim_{n\to\infty} A^2 u_n = \alpha^2 u$. Now

$$\begin{aligned}\|(A^2 - \alpha^2)u_n\|^2 &= \|A^2 u_n\|^2 - 2\alpha^2\langle u_n, A^2 u_n\rangle + \alpha^4\\ &\leq 2\alpha^2(\alpha^2 - \langle u_n, A^2 u_n\rangle)\end{aligned}$$

(where we have used $\|A^2 u_n\| \leq \|A\|\|Au_n\| \leq \|A\|^2\|u_n\| = \alpha^2$) implies $\lim_{n\to\infty}(A^2 u_n - \alpha^2 u_n) = 0$ and hence $\lim_{n\to\infty} u_n = u$. In addition, u is a normalized eigenvector of A^2 since $(A^2 - \alpha^2)u = 0$. Factorizing this last equation according to $(A - \alpha)u = v$ and $(A + \alpha)v = 0$ show that either $v \neq 0$ is an eigenvector corresponding to $-\alpha$ or $v = 0$ and hence $u \neq 0$ is an eigenvector corresponding to α. □

Note that for a bounded operator A, there cannot be an eigenvalue with absolute value larger than $\|A\|$, that is, the set of eigenvalues is bounded by $\|A\|$ (Problem 3.3).

Now consider a symmetric compact operator A with eigenvalue α_1 (as above) and corresponding normalized eigenvector u_1. Setting

$$\mathfrak{H}_1 = \{u_1\}^\perp = \{f \in \mathfrak{H} | \langle u_1, f\rangle = 0\} \tag{3.6}$$

we can restrict A to $\mathfrak{H}_1$ since $f \in \mathfrak{H}_1$ implies

$$\langle u_1, Af\rangle = \langle Au_1, f\rangle = \alpha_1 \langle u_1, f\rangle = 0 \tag{3.7}$$

and hence $Af \in \mathfrak{H}_1$. Denoting this restriction by A_1, it is not hard to see that A_1 is again a symmetric compact operator. Hence we can apply Theorem 3.7 iteratively to obtain a sequence of eigenvalues α_j with corresponding normalized eigenvectors u_j. Moreover, by construction, u_j is orthogonal to all u_k with $k < j$ and hence the eigenvectors $\{u_j\}$ form an orthonormal set. This procedure will not stop unless $\mathfrak{H}$ is finite dimensional. However, note that $\alpha_j = 0$ for $j \geq n$ might happen if $A_n = 0$.

Theorem 3.8. *Suppose $\mathfrak{H}$ is an infinite dimensional Hilbert space and $A : \mathfrak{H} \to \mathfrak{H}$ is a compact symmetric operator. Then there exists a sequence of real eigenvalues α_j converging to 0. The corresponding normalized eigenvectors u_j form an orthonormal set and every $f \in \mathfrak{H}$ can be written as*

$$f = \sum_{j=1}^{\infty} \langle u_j, f\rangle u_j + h, \tag{3.8}$$

where h is in the kernel of A, that is, $Ah = 0$.

In particular, if 0 is not an eigenvalue, then the eigenvectors form an orthonormal basis (in addition, $\mathfrak{H}$ need not be complete in this case).

Proof. Existence of the eigenvalues α_j and the corresponding eigenvectors u_j has already been established. If the eigenvalues should not converge to zero, there is a subsequence such that $|\alpha_{j_k}| \geq \varepsilon$. Hence $v_k = \alpha_{j_k}^{-1} u_{j_k}$ is a bounded sequence ($\|v_k\| \leq \frac{1}{\varepsilon}$) for which Av_k has no convergent subsequence since $\|Av_k - Av_l\|^2 = \|u_{j_k} - u_{j_l}\|^2 = 2$, a contradiction.

Next, setting

$$f_n = \sum_{j=1}^{n} \langle u_j, f\rangle u_j,$$

we have

$$\|A(f - f_n)\| \leq |\alpha_n| \|f - f_n\| \leq |\alpha_n| \|f\|$$

since $f - f_n \in \mathfrak{H}_n$ and $\|A_n\| = |\alpha_n|$. Letting $n \to \infty$ shows $A(f_\infty - f) = 0$ proving (3.8). □

By applying A to (3.8) we obtain the following canonical form of compact symmetric operators.

Corollary 3.9. *Every compact symmetric operator A can be written as*

$$Af = \sum_{j=1}^{\infty} \alpha_j \langle u_j, f\rangle u_j, \tag{3.9}$$

where α_j are the nonzero eigenvalues with corresponding eigenvectors u_j from the previous theorem.

Remark: There are two cases where our procedure might fail to construct an orthonormal basis of eigenvectors. One case is where there is an infinite number of nonzero eigenvalues. In this case α_n never reaches 0 and all eigenvectors corresponding to 0 are missed. In the other case, 0 is reached, but there might not be a countable basis and hence again some of the eigenvectors corresponding to 0 are missed. In any case by adding vectors from the kernel (which are automatically eigenvectors), one can always extend the eigenvectors u_j to an orthonormal basis of eigenvectors.

Corollary 3.10. *Every compact symmetric operator has an associated orthonormal basis of eigenvectors.*

This is all we need and it remains to apply these results to Sturm–Liouville operators.

Problem 3.3. *Show that if A is bounded, then every eigenvalue α satisfies $|\alpha| \le \|A\|$.*

Problem 3.4. *Find the eigenvalues and eigenfunctions of the integral operator*

$$(Kf)(x) = \int_0^1 u(x)v(y)f(y)dy$$

in $\mathcal{L}^2_{cont}(0,1)$, where $u(x)$ and $v(x)$ are some given continuous functions.

Problem 3.5. *Find the eigenvalues and eigenfunctions of the integral operator*

$$(Kf)(x) = 2\int_0^1 (2xy - x - y + 1)f(y)dy$$

in $\mathcal{L}^2_{cont}(0,1)$.

3.3. Applications to Sturm–Liouville operators

Now, after all this hard work, we can show that our Sturm–Liouville operator

$$L = -\frac{d^2}{dx^2} + q(x), \tag{3.10}$$

where q is continuous and real, defined on

$$\mathfrak{D}(L) = \{f \in C^2[0,1] | f(0) = f(1) = 0\} \subset \mathcal{L}^2_{cont}(0,1), \tag{3.11}$$

has an orthonormal basis of eigenfunctions.

The corresponding eigenvalue equation $Lu = zu$ explicitly reads

$$-u''(x) + q(x)u(x) = zu(x). \tag{3.12}$$

It is a second order homogenous linear ordinary differential equations and hence has two linearly independent solutions. In particular, specifying two initial conditions, e.g. $u(0) = 0, u'(0) = 1$ determines the solution uniquely. Hence, if we require $u(0) = 0$, the solution is determined up to a multiple and consequently the additional requirement $u(1) = 0$ cannot be satisfied by a nontrivial solution in general. However, there might be some $z \in \mathbb{C}$ for which the solution corresponding to the initial conditions $u(0) = 0, u'(0) = 1$ happens to satisfy $u(1) = 0$ and these are precisely the eigenvalues we are looking for.

Note that the fact that $\mathcal{L}^2_{cont}(0,1)$ is not complete causes no problems since we can always replace it by its completion $\mathfrak{H} = L^2(0,1)$. A thorough investigation of this completion will be given later, at this point this is not essential.

We first verify that L is symmetric:

$$\begin{aligned}
\langle f, Lg\rangle &= \int_0^1 f(x)^*(-g''(x) + q(x)g(x))dx \\
&= \int_0^1 f'(x)^* g'(x)dx + \int_0^1 f(x)^* q(x)g(x)dx \\
&= \int_0^1 -f''(x)^* g(x)dx + \int_0^1 f(x)^* q(x)g(x)dx \\
&= \langle Lf, g\rangle.
\end{aligned} \tag{3.13}$$

Here we have used integration by part twice (the boundary terms vanish due to our boundary conditions $f(0) = f(1) = 0$ and $g(0) = g(1) = 0$).

Of course we want to apply Theorem 3.8 and for this we would need to show that L is compact. But this task is bound to fail, since L is not even bounded (see the example in Section 1.5)!

So here comes the trick: If L is unbounded its inverse L^{-1} might still be bounded. Moreover, L^{-1} might even be compact and this is the case here! Since L might not be injective (0 might be an eigenvalue), we consider $R_L(z) = (L - z)^{-1}$, $z \in \mathbb{C}$, which is also known as the **resolvent** of L.

In order to compute the resolvent, we need to solve the inhomogeneous equation $(L - z)f = g$. This can be done using the variation of constants formula from ordinary differential equations which determines the solution up to an arbitrary solution of the homogenous equation. This homogenous equation has to be chosen such that $f \in \mathfrak{D}(L)$, that is, such that $f(0) = f(1) = 0$.

Define

$$\begin{aligned} f(x) &= \frac{u_+(z,x)}{W(z)}\Big(\int_0^x u_-(z,t)g(t)dt\Big) \\ &\quad + \frac{u_-(z,x)}{W(z)}\Big(\int_x^1 u_+(z,t)g(t)dt\Big), \end{aligned} \tag{3.14}$$

where $u_\pm(z,x)$ are the solutions of the homogenous differential equation $-u_\pm''(z,x)+(q(x)-z)u_\pm(z,x) = 0$ satisfying the initial conditions $u_-(z,0) = 0$, $u_-'(z,0) = 1$ respectively $u_+(z,1) = 0$, $u_+'(z,1) = 1$ and

$$W(z) = W(u_+(z), u_-(z)) = u_-'(z,x)u_+(z,x) - u_-(z,x)u_+'(z,x) \tag{3.15}$$

is the Wronski determinant, which is independent of x (check this!).

Then clearly $f(0) = 0$ since $u_-(z,0) = 0$ and similarly $f(1) = 0$ since $u_+(z,1) = 0$. Furthermore, f is differentiable and a straightforward computation verifies

$$\begin{aligned} f'(x) &= \frac{u_+(z,x)'}{W(z)}\Big(\int_0^x u_-(z,t)g(t)dt\Big) \\ &\quad + \frac{u_-(z,x)'}{W(z)}\Big(\int_x^1 u_+(z,t)g(t)dt\Big). \end{aligned} \tag{3.16}$$

Thus we can differentiate once more giving

$$\begin{aligned} f''(x) &= \frac{u_+(z,x)''}{W(z)}\Big(\int_0^x u_-(z,t)g(t)dt\Big) \\ &\quad + \frac{u_-(z,x)''}{W(z)}\Big(\int_x^1 u_+(z,t)g(t)dt\Big) - g(x) \\ &= (q(x) - z)f(x) - g(x). \end{aligned} \tag{3.17}$$

In summary, f is in the domain of L and satisfies $(L - z)f = g$.

Note that z is an eigenvalue if and only if $W(z) = 0$. In fact, in this case $u_+(z,x)$ and $u_-(z,x)$ are linearly dependent and hence $u_-(z,1) = c\,u_+(z,1) = 0$ which shows that $u_-(z,x)$ satisfies both boundary conditions and is thus an eigenfunction.

Introducing the **Green function**

$$G(z,x,t) = \frac{1}{W(u_+(z),u_-(z))}\begin{cases} u_+(z,x)u_-(z,t), & x \ge t \\ u_+(z,t)u_-(z,x), & x \le t \end{cases} \tag{3.18}$$

we see that $(L - z)^{-1}$ is given by

$$(L - z)^{-1}g(x) = \int_0^1 G(z,x,t)g(t)dt. \tag{3.19}$$

Moreover, from $G(z,x,t) = G(z,t,x)$ it follows that $(L-z)^{-1}$ is symmetric for $z \in \mathbb{R}$ (Problem 3.6) and from Lemma 3.4 it follows that it is compact. Hence Theorem 3.8 applies to $(L-z)^{-1}$ and we obtain:

Theorem 3.11. *The Sturm–Liouville operator L has a countable number of eigenvalues E_n. All eigenvalues are discrete and simple. The corresponding normalized eigenfunctions u_n form an orthonormal basis for $\mathcal{L}^2_{cont}(0,1)$.*

Proof. Pick a value $\lambda \in \mathbb{R}$ such that $R_L(\lambda)$ exists. By Lemma 3.4 $R_L(\lambda)$ is compact and by Theorem 3.3 this remains true if we replace $\mathcal{L}^2_{cont}(0,1)$ by its completion. By Theorem 3.8 there are eigenvalues α_n of $R_L(\lambda)$ with corresponding eigenfunctions u_n. Moreover, $R_L(\lambda)u_n = \alpha_n u_n$ is equivalent to $Lu_n = (\lambda + \frac{1}{\alpha_n})u_n$, which shows that $E_n = \lambda + \frac{1}{\alpha_n}$ are eigenvalues of L with corresponding eigenfunctions u_n. Now everything follows from Theorem 3.8 except that the eigenvalues are simple. To show this, observe that if u_n and v_n are two different eigenfunctions corresponding to E_n, then $u_n(0) = v_n(0) = 0$ implies $W(u_n, v_n) = 0$ and hence u_n and v_n are linearly dependent. □

Problem 3.6. *Show that for our Sturm–Liouville operator $u_\pm(z,x)^* = u_\pm(z^*,x)$. Conclude $R_L(z)^* = R_L(z^*)$. (Hint: Problem 3.2.)*

Problem 3.7. *Show that the resolvent $R_A(z) = (A-z)^{-1}$ (provided it exists and is densely defined) of a symmetric operator A is again symmetric for $z \in \mathbb{R}$. (Hint: $g \in \mathfrak{D}(R_A(z))$ if and only if $g = (A-z)f$ for some $f \in \mathfrak{D}(A)$).*

Chapter 4

The main theorems about Banach spaces

4.1. The Baire theorem and its consequences

Recall that the interior of a set is the largest open subset (that is, the union of all open subsets). A set is called **nowhere dense** if its closure has empty interior. The key to several important theorems about Banach spaces is the observation that a Banach space cannot be the countable union of nowhere dense sets.

Theorem 4.1 (Baire category theorem)**.** *Let X be a complete metric space. Then X cannot be the countable union of nowhere dense sets.*

Proof. Suppose $X = \bigcup_{n=1}^{\infty} X_n$. We can assume that the sets X_n are closed and none of them contains a ball; that is, $X \backslash X_n$ is open and nonempty for every n. We will construct a Cauchy sequence x_n which stays away from all X_n.

Since $X \backslash X_1$ is open and nonempty, there is a closed ball $B_{r_1}(x_1) \subseteq X \backslash X_1$. Reducing r_1 a little, we can even assume $\overline{B_{r_1}(x_1)} \subseteq X \backslash X_1$. Moreover, since X_2 cannot contain $B_{r_1}(x_1)$, there is some $x_2 \in B_{r_1}(x_1)$ that is not in X_2. Since $B_{r_1}(x_1) \cap (X \backslash X_2)$ is open, there is a closed ball $\overline{B_{r_2}(x_2)} \subseteq B_{r_1}(x_1) \cap (X \backslash X_2)$. Proceeding by induction, we obtain a sequence of balls such that

$$\overline{B_{r_n}(x_n)} \subseteq B_{r_{n-1}}(x_{n-1}) \cap (X \backslash X_n).$$

Now observe that in every step we can choose r_n as small as we please; hence without loss of generality $r_n \to 0$. Since by construction $x_n \in \overline{B_{r_N}(x_N)}$ for $n \geq N$, we conclude that x_n is Cauchy and converges to some point $x \in X$.

But $x \in \overline{B_{r_n}(x_n)} \subseteq X\backslash X_n$ for every n, contradicting our assumption that the X_n cover X. □

Remark: The set of rational numbers $\mathbb{Q}$ can be written as a countable union of its elements. This shows that completeness assumption is crucial.

(Sets which can be written as the countable union of nowhere dense sets are said to be of first category. All other sets are second category. Hence we have the name category theorem.)

In other words, if $X_n \subseteq X$ is a sequence of closed subsets which cover X, at least one X_n contains a ball of radius $\varepsilon > 0$.

Since a closed set is nowhere dense if and only if its complement is open and dense (cf. Problem 1.4), there is a reformulation which is also worthwhile noting:

Corollary 4.2. *Let X be a complete metric space. Then any countable intersection of open dense sets is again dense.*

Proof. Let O_n be open dense sets whose intersection is not dense. Then this intersection must be missing some closed ball $\overline{B_\varepsilon}$. This ball will lie in $\bigcup_n X_n$, where $X_n = X\backslash O_n$ are closed and nowhere dense. Now note that $\tilde{X}_n = X_n \cup \overline{B}$ are closed nowhere dense sets in $\overline{B_\varepsilon}$. But $\overline{B_\varepsilon}$ is a complete metric space, a contradiction. □

Now we come to the first important consequence, the **uniform boundedness principle**.

Theorem 4.3 (Banach–Steinhaus)**.** *Let X be a Banach space and Y some normed vector space. Let $\{A_\alpha\} \subseteq \mathfrak{L}(X,Y)$ be a family of bounded operators. Suppose $\|A_\alpha x\| \le C(x)$ is bounded for fixed $x \in X$. Then $\{A_\alpha\}$ is uniformly bounded, $\|A_\alpha\| \le C$.*

Proof. Let

$$X_n = \{x|\, \|A_\alpha x\| \le n \text{ for all } \alpha\} = \bigcap_\alpha \{x|\, \|A_\alpha x\| \le n\}.$$

Then $\bigcup_n X_n = X$ by assumption. Moreover, by continuity of A_α and the norm, each X_n is an intersection of closed sets and hence closed. By Baire's theorem at least one contains a ball of positive radius: $\overline{B_\varepsilon(x_0)} \subset X_n$. Now observe

$$\|A_\alpha y\| \le \|A_\alpha(y + x_0)\| + \|A_\alpha x_0\| \le n + C(x_0)$$

for $\|y\| \le \varepsilon$. Setting $y = \varepsilon \frac{x}{\|x\|}$, we obtain

$$\|A_\alpha x\| \le \frac{n + C(x_0)}{\varepsilon} \|x\|$$

for every x. □

The next application is

Theorem 4.4 (Open mapping). *Let $A \in \mathfrak{L}(X, Y)$ be a bounded linear operator from one Banach space onto another. Then A is open (i.e., maps open sets to open sets).*

Proof. Denote by $B_r^X(x) \subseteq X$ the open ball with radius r centered at x and let $B_r^X = B_r^X(0)$. Similarly for $B_r^Y(y)$. By scaling and translating balls (using linearity of A), it suffices to prove $B_\varepsilon^Y \subseteq A(B_1^X)$ for some $\varepsilon > 0$. Since A is surjective we have

$$Y = \bigcup_{n=1}^{\infty} A(B_n^X)$$

and the Baire theorem implies that for some n, $\overline{A(B_n^X)}$ contains a ball $B_\varepsilon^Y(y)$. Without restriction $n = 1$ (just scale the balls). Since $-\overline{A(B_1^X)} = \overline{A(-B_1^X)} = \overline{A(B_1^X)}$ we see $B_\varepsilon^Y(-y) \subseteq \overline{A(B_1^X)}$ and by convexity of $\overline{A(B_1^X)}$ we also have $B_\varepsilon^Y \subseteq \overline{A(B_1^X)}$.

So we have $B_\varepsilon^Y \subseteq \overline{A(B_1^X)}$, but we would need $B_\varepsilon^Y \subseteq A(B_1^X)$. To complete the proof we will show $\overline{A(B_1^X)} \subseteq A(B_2^X)$ which implies $B_{\varepsilon/2}^Y \subseteq A(B_1^X)$.

For every $y \in \overline{A(B_1^X)}$ we can choose some sequence $y_n \in A(B_1^X)$ with $y_n \to y$. Moreover, there even is some $x_n \in B_1^X$ with $y_n = A(x_n)$. However x_n might not converge, so we need to argue more carefully and ensure convergence along the way: start with $x_1 \in B_1^X$ such that $y - Ax_1 \in B_{\varepsilon/2}^Y$. Scaling the relation $B_\varepsilon^Y \subset \overline{A(B_1^X)}$ we have $B_{\varepsilon/2}^Y \subset \overline{A(B_{1/2}^X)}$ and hence we can choose $x_2 \in B_{1/2}^X$ such that $(y - Ax_1) - Ax_2 \in B_{\varepsilon/4}^Y \subset \overline{A(B_{1/4}^X)}$. Proceeding like this we obtain a sequence of points $x_n \in B_{2^{1-n}}^X$ such that

$$y - \sum_{k=1}^{n} Ax_k \in B_{\varepsilon 2^{-n}}^Y.$$

By $\|x_k\| < 2^{1-k}$ the limit $x = \sum_{k=1}^{\infty} x_k$ exists and satisfies $\|x\| < 2$. Hence $y = Ax \in A(B_2^X)$ as desired. □

Remark: The requirement that A is onto is crucial (just look at the one-dimensional case $X = \mathbb{C}$). Moreover, the converse is also true: If A is open, then the image of the unit ball contains again some ball $B_\varepsilon^Y \subseteq A(B_1^X)$. Hence by scaling $B_{r\varepsilon}^Y \subseteq A(B_r^X)$ and letting $r \to \infty$ we see that A is onto: $Y = A(X)$.

As an immediate consequence we get the inverse mapping theorem:

Theorem 4.5 (Inverse mapping). *Let $A \in \mathfrak{L}(X, Y)$ be a bounded linear bijection between Banach spaces. Then A^{-1} is continuous.*

Example. Consider the operator $(Aa)_{j=1}^n = (\frac{1}{j} a_j)_{j=1}^n$ in $\ell^2(\mathbb{N})$. Then its inverse $(A^{-1}a)_{j=1}^n = (j\, a_j)_{j=1}^n$ is unbounded (show this!). This is in agreement with our theorem since its range is dense (why?) but not all of $\ell^2(\mathbb{N})$: For example $(b_j = \frac{1}{j})_{j=1}^\infty \notin \operatorname{Ran}(A)$ since $b = Aa$ gives the contradiction

$$\infty = \sum_{j=1}^\infty 1 = \sum_{j=1}^\infty |j b_j|^2 = \sum_{j=1}^\infty |a_j|^2 < \infty.$$

In fact, for an injective operator the range is closed if and only if the inverse is bounded (Problem 4.2). ⋄

Another important consequence is the closed graph theorem. The **graph** of an operator A is just

$$\Gamma(A) = \{(x, Ax) | x \in \mathfrak{D}(A)\}. \tag{4.1}$$

If A is linear, the graph is a subspace of the Banach space $X \oplus Y$ (provided X and Y are Banach spaces), which is just the cartesian product together with the norm

$$\|(x, y)\|_{X\oplus Y} = \|x\|_X + \|y\|_Y \tag{4.2}$$

(check this). Note that $(x_n, y_n) \to (x, y)$ if and only if $x_n \to x$ and $y_n \to y$. We say that A has a closed graph if $\Gamma(A)$ is a closed set in $X \oplus Y$.

Theorem 4.6 (Closed graph)**.** *Let $A : X \to Y$ be a linear map from a Banach space X to another Banach space Y. Then A is bounded if and only if its graph is closed.*

Proof. If $\Gamma(A)$ is closed, then it is again a Banach space. Now the projection $\pi_1(x, Ax) = x$ onto the first component is a continuous bijection onto X. So by the inverse mapping theorem its inverse π_1^{-1} is again continuous. Moreover, the projection $\pi_2(x, Ax) = Ax$ onto the second component is also continuous and consequently so is $A = \pi_2 \circ \pi_1^{-1}$. The converse is easy. □

Remark: The crucial fact here is that A is defined on *all* of X!

Operators whose graphs are closed are called **closed operators**. Being closed is the next option you have once an operator turns out to be unbounded. If A is closed, then $x_n \to x$ does not guarantee you that Ax_n converges (like continuity would), but it at least guarantees that if Ax_n converges, it converges to the right thing, namely Ax:

- A bounded: $x_n \to x$ implies $Ax_n \to Ax$.
- A closed: $x_n \to x$ and $Ax_n \to y$ implies $y = Ax$.

If an operator is not closed, you can try to take the closure of its graph, to obtain a closed operator. If A is bounded this always works (which is just the contents of Theorem 1.29). However, in general, the closure of the graph might not be the graph of an operator as we might pick up points $(x, y_{1,2}) \in \overline{\Gamma(A)}$ with $y_1 \neq y_2$. Since $\overline{\Gamma(A)}$ is a subspace, we also have $(x, y_2) - (x, y_1) = (0, y_2 - y_1) \in \overline{\Gamma(A)}$ in this case and thus $\overline{\Gamma(A)}$ is the graph of some operator if and only if

$$\overline{\Gamma(A)} \cap \{(0, y) | y \in Y\} = \{(0, 0)\}. \tag{4.3}$$

If this is the case, A is called **closable** and the operator $\overline{A}$ associated with $\overline{\Gamma(A)}$ is called the **closure** of A.

In particular, A is closable if and only if $x_n \to 0$ and $Ax_n \to y$ implies $y = 0$. In this case

$$\begin{aligned} \mathfrak{D}(\overline{A}) &= \{x \in X | \exists x_n \in \mathfrak{D}(A),\ y \in Y : x_n \to x \text{ and } Ax_n \to y\}, \\ \overline{A}x &= y. \end{aligned} \tag{4.4}$$

For yet another way of defining the closure see Problem 4.5.

Example. Consider the operator A in $\ell^p(\mathbb{N})$ defined by $Aa_j = ja_j$ on $\mathfrak{D}(A) = \{a \in \ell^p(\mathbb{N}) | a_j \neq 0 \text{ for finitely many } j\}$.

(i). A is closable. In fact, if $a^n \to 0$ and $Aa^n \to b$ then we have $a^n_j \to 0$ and thus $ja^n_j \to 0 = b_j$ for any $j \in \mathbb{N}$.

(ii). The closure of A is given by

$$\mathfrak{D}(\overline{A}) = \{a \in \ell^p(\mathbb{N}) | (ja_j)_{j=1}^\infty \in \ell^p(\mathbb{N})\}$$

and $\overline{A}a_j = ja_j$. In fact, if $a^n \to a$ and $Aa^n \to b$ then we have $a^n_j \to a_j$ and $ja^n_j \to b_j$ for any $j \in \mathbb{N}$ and thus $b_j = ja_j$ for any $j \in \mathbb{N}$. In particular $(ja_j)_{j=1}^\infty = (b_j)_{j=1}^\infty \in \ell^p(\mathbb{N})$. Conversely, suppose $(ja_j)_{j=1}^\infty \in \ell^p(\mathbb{N})$ and consider

$$a^n_j = \begin{cases} a_j, & j \leq n, \\ 0, & j > n. \end{cases}$$

Then $a^n \to a$ and $Aa^n \to (ja_j)_{j=1}^\infty$.

(iii). Note that the inverse of $\overline{A}$ is the bounded operator $\overline{A}^{-1} a_j = j^{-1} a_j$ defined on all of $\ell^p(\mathbb{N})$. Thus $\overline{A}^{-1}$ is closed. However, since its range $\operatorname{Ran}(\overline{A}^{-1}) = \mathfrak{D}(\overline{A})$ is dense but not all of $\ell^p(\mathbb{N})$, $\overline{A}^{-1}$ does not map closed sets to closed sets in general. In particular, the concept of a closed operator should not be confused with the concept of a closed map in topology! ⋄

The closed graph theorem tells us that closed linear operators can be defined on all of X if and only if they are bounded. So if we have an unbounded operator we cannot have both! That is, if we want our operator to be at least closed, we have to live with domains. This is the reason why in quantum mechanics most operators are defined on domains. In fact, there is another important property which does not allow unbounded operators to be defined on the entire space:

Theorem 4.7 (Hellinger–Toeplitz). *Let $A : \mathfrak{H} \to \mathfrak{H}$ be a linear operator on some Hilbert space $\mathfrak{H}$. If A is symmetric, that is $\langle g, Af\rangle = \langle Ag, f\rangle$, $f, g \in \mathfrak{H}$, then A is bounded.*

Proof. It suffices to prove that A is closed. In fact, $f_n \to f$ and $Af_n \to g$ implies

$$\langle h, g\rangle = \lim_{n\to\infty} \langle h, Af_n\rangle = \lim_{n\to\infty} \langle Ah, f_n\rangle = \langle Ah, f\rangle = \langle h, Af\rangle$$

for every $h \in \mathfrak{H}$. Hence $Af = g$. □

Problem 4.1. *Is the sum of two closed operators also closed? (Here $A + B$ is defined on $\mathfrak{D}(A + B) = \mathfrak{D}(A) \cap \mathfrak{D}(B)$.)*

Problem 4.2. *Suppose $A : \mathfrak{D}(A) \to \mathrm{Ran}(A)$ is closed and injective. Show that A^{-1} defined on $\mathfrak{D}(A^{-1}) = \mathrm{Ran}(A)$ is closed as well.*

Conclude that in this case $\mathrm{Ran}(A)$ is closed if and only if A^{-1} is bounded.

Problem 4.3. *Show that the differential operator $A = \frac{d}{dx}$ defined on $\mathfrak{D}(A) = C^1[0,1] \subset C[0,1]$ (sup norm) is a closed operator. (Compare the example in Section 1.5.)*

Problem 4.4. *Consider $A = \frac{d}{dx}$ defined on $\mathfrak{D}(A) = C^1[0,1] \subset L^2(0,1)$. Show that its closure is given by*

$$\mathfrak{D}(\overline{A}) = \{f \in L^2(0,1) | \exists g \in L^2(0,1),\, c \in \mathbb{C} : f(x) = c + \int_0^x g(y)dy\}$$

and $\overline{A}f = g$.

Problem 4.5. *Consider a linear operator $A : \mathfrak{D}(A) \subseteq X \to Y$, where X and Y are Banach spaces. Define the* **graph norm** *associated with A by*

$$\|x\|_A = \|x\|_X + \|Ax\|_Y. \tag{4.5}$$

Show that $A : \mathfrak{D}(A) \to Y$ is bounded if we equip $\mathfrak{D}(A)$ with the graph norm. Show that the completion X_A of $(\mathfrak{D}(A), \|.\|_A)$ can be regarded as a subset of X if and only if A is closable. Show that in this case the completion can be identified with $\mathfrak{D}(\overline{A})$ and that the closure of A in X coincides with the extension from Theorem 1.29 of A in X_A.

4.2. The Hahn–Banach theorem and its consequences

Let X be a Banach space. Recall that we have called the set of all bounded linear functionals the dual space X^* (which is again a Banach space by Theorem 1.30).

Example. Consider the Banach space $\ell^p(\mathbb{N})$, $1 \le p < \infty$. Then, by Hölder's inequality (1.25), every $y \in \ell^q(\mathbb{N})$ gives rise to a bounded linear functional

$$l_y(x) = \sum_{n\in\mathbb{N}} y_n x_n \tag{4.6}$$

whose norm is $\|l_y\| = \|y\|_q$ (Problem 4.8). But can every element of $\ell^p(\mathbb{N})^*$ be written in this form?

Suppose $p = 1$ and choose $l \in \ell^1(\mathbb{N})^*$. Now define

$$y_n = l(\delta^n), \tag{4.7}$$

where $\delta^n_n = 1$ and $\delta^n_m = 0$, $n \ne m$. Then

$$|y_n| = |l(\delta^n)| \le \|l\| \, \|\delta^n\|_1 = \|l\| \tag{4.8}$$

shows $\|y\|_\infty \le \|l\|$, that is, $y \in \ell^\infty(\mathbb{N})$. By construction $l(x) = l_y(x)$ for every $x \in \operatorname{span}\{\delta^n\}$. By continuity of ℓ it even holds for $x \in \overline{\operatorname{span}\{\delta^n\}} = \ell^1(\mathbb{N})$. Hence the map $y \mapsto l_y$ is an isomorphism, that is, $\ell^1(\mathbb{N})^* \cong \ell^\infty(\mathbb{N})$. A similar argument shows $\ell^p(\mathbb{N})^* \cong \ell^q(\mathbb{N})$, $1 \le p < \infty$ (Problem 4.9). One usually identifies $\ell^p(\mathbb{N})^*$ with $\ell^q(\mathbb{N})$ using this canonical isomorphism and simply writes $\ell^p(\mathbb{N})^* = \ell^q(\mathbb{N})$. In the case $p = \infty$ this is not true, as we will see soon. ⋄

It turns out that many questions are easier to handle after applying a linear functional $\ell \in X^*$. For example, suppose $x(t)$ is a function $\mathbb{R} \to X$ (or $\mathbb{C} \to X$), then $\ell(x(t))$ is a function $\mathbb{R} \to \mathbb{C}$ (respectively $\mathbb{C} \to \mathbb{C}$) for any $\ell \in X^*$. So to investigate $\ell(x(t))$ we have all tools from real/complex analysis at our disposal. But how do we translate this information back to $x(t)$? Suppose we have $\ell(x(t)) = \ell(y(t))$ for all $\ell \in X^*$. Can we conclude $x(t) = y(t)$? The answer is yes and will follow from the Hahn–Banach theorem.

We first prove the real version from which the complex one then follows easily.

Theorem 4.8 (Hahn–Banach, real version)**.** *Let X be a real vector space and $\varphi : X \to \mathbb{R}$ a convex function (i.e., $\varphi(\lambda x+(1-\lambda)y) \le \lambda\varphi(x)+(1-\lambda)\varphi(y)$ for $\lambda \in (0,1)$).*

If ℓ is a linear functional defined on some subspace $Y \subset X$ which satisfies $\ell(y) \le \varphi(y)$, $y \in Y$, then there is an extension $\overline{\ell}$ to all of X satisfying $\overline{\ell}(x) \le \varphi(x)$, $x \in X$.

Proof. Let us first try to extend ℓ in just one direction: Take $x \notin Y$ and set $\tilde{Y} = \operatorname{span}\{x, Y\}$. If there is an extension $\tilde{\ell}$ to $\tilde{Y}$ it must clearly satisfy

$$\tilde{\ell}(y + \alpha x) = \ell(y) + \alpha\tilde{\ell}(x).$$

So all we need to do is to choose $\tilde{\ell}(x)$ such that $\tilde{\ell}(y+\alpha x) \le \varphi(y+\alpha x)$. But this is equivalent to

$$\sup_{\alpha>0, y\in Y} \frac{\varphi(y-\alpha x)-\ell(y)}{-\alpha} \le \tilde{\ell}(x) \le \inf_{\alpha>0, y\in Y} \frac{\varphi(y+\alpha x)-\ell(y)}{\alpha}$$

and is hence only possible if

$$\frac{\varphi(y_1-\alpha_1 x)-\ell(y_1)}{-\alpha_1} \le \frac{\varphi(y_2+\alpha_2 x)-\ell(y_2)}{\alpha_2}$$

for every $\alpha_1, \alpha_2 > 0$ and $y_1, y_2 \in Y$. Rearranging this last equations we see that we need to show

$$\alpha_2 \ell(y_1) + \alpha_1 \ell(y_2) \le \alpha_2 \varphi(y_1 - \alpha_1 x) + \alpha_1 \varphi(y_2 + \alpha_2 x).$$

Starting with the left-hand side we have

$$\begin{aligned}
\alpha_2 \ell(y_1) + \alpha_1 \ell(y_2) &= (\alpha_1+\alpha_2)\ell\left(\lambda y_1 + (1-\lambda) y_2\right) \\
&\le (\alpha_1+\alpha_2)\varphi\left(\lambda y_1 + (1-\lambda) y_2\right) \\
&= (\alpha_1+\alpha_2)\varphi\left(\lambda(y_1-\alpha_1 x) + (1-\lambda)(y_2+\alpha_2 x)\right) \\
&\le \alpha_2 \varphi(y_1-\alpha_1 x) + \alpha_1 \varphi(y_2+\alpha_2 x),
\end{aligned}$$

where $\lambda = \frac{\alpha_2}{\alpha_1+\alpha_2}$. Hence one dimension works.

To finish the proof we appeal to Zorn's lemma (see Appendix A): Let E be the collection of all extensions $\tilde{\ell}$ satisfying $\tilde{\ell}(x) \le \varphi(x)$. Then E can be partially ordered by inclusion (with respect to the domain) and every linear chain has an upper bound (defined on the union of all domains). Hence there is a maximal element $\overline{\ell}$ by Zorn's lemma. This element is defined on X, since if it were not, we could extend it as before contradicting maximality. □

Theorem 4.9 (Hahn–Banach, complex version)**.** *Let X be a complex vector space and $\varphi : X \to \mathbb{R}$ a convex function satisfying $\varphi(\alpha x) \le \varphi(x)$ if $|\alpha| = 1$.*

If ℓ is a linear functional defined on some subspace $Y \subset X$ which satisfies $|\ell(y)| \le \varphi(y)$, $y \in Y$, then there is an extension $\overline{\ell}$ to all of X satisfying $|\overline{\ell}(x)| \le \varphi(x)$, $x \in X$.

Proof. Set $\ell_r = \mathrm{Re}(\ell)$ and observe

$$\ell(x) = \ell_r(x) - \mathrm{i}\ell_r(\mathrm{i}x).$$

By our previous theorem, there is a real linear extension $\overline{\ell}_r$ satisfying $\overline{\ell}_r(x) \le \varphi(x)$. Now set $\overline{\ell}(x) = \overline{\ell}_r(x) - \mathrm{i}\overline{\ell}_r(\mathrm{i}x)$. Then $\overline{\ell}(x)$ is real linear and by $\overline{\ell}(\mathrm{i}x) = \overline{\ell}_r(\mathrm{i}x) + \mathrm{i}\overline{\ell}_r(x) = \mathrm{i}\overline{\ell}(x)$ also complex linear. To show $|\overline{\ell}(x)| \le \varphi(x)$ we abbreviate $\alpha = \frac{\overline{\ell}(x)^*}{|\overline{\ell}(x)|}$ and use

$$|\overline{\ell}(x)| = \alpha\overline{\ell}(x) = \overline{\ell}(\alpha x) = \overline{\ell}_r(\alpha x) \le \varphi(\alpha x) \le \varphi(x),$$

which finishes the proof. □

Note that $\varphi(\alpha x) \le \varphi(x)$, $|\alpha| = 1$ is in fact equivalent to $\varphi(\alpha x) = \varphi(x)$, $|\alpha| = 1$.

If ℓ is a linear functional defined on some subspace, the choice $\varphi(x) = \|\ell\|\|x\|$ implies:

Corollary 4.10. *Let X be a Banach space and let ℓ be a bounded linear functional defined on some subspace $Y \subseteq X$. Then there is an extension $\overline{\ell} \in X^*$ preserving the norm.*

Moreover, we can now easily prove our anticipated result

Corollary 4.11. *Suppose $\ell(x) = 0$ for all ℓ in some total subset $Y \subseteq X^*$. Then $x = 0$.*

Proof. Clearly if $\ell(x) = 0$ holds for all ℓ in some total subset, this holds for all $\ell \in X^*$. If $x \neq 0$ we can construct a bounded linear functional on $\mathrm{span}\{x\}$ by setting $\ell(\alpha x) = \alpha$ and extending it to X^* using the previous corollary. But this contradicts our assumption. □

Example. Let us return to our example $\ell^\infty(\mathbb{N})$. Let $c(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$ be the subspace of convergent sequences. Set

$$l(x) = \lim_{n\to\infty} x_n, \qquad x \in c(\mathbb{N}), \tag{4.9}$$

then l is bounded since

$$|l(x)| = \lim_{n\to\infty} |x_n| \le \|x\|_\infty. \tag{4.10}$$

Hence we can extend it to $\ell^\infty(\mathbb{N})$ by Corollary 4.10. Then $l(x)$ cannot be written as $l(x) = l_y(x)$ for some $y \in \ell^1(\mathbb{N})$ (as in (4.6)) since $y_n = l(\delta^n) = 0$ shows $y = 0$ and hence $\ell_y = 0$. The problem is that $\overline{\mathrm{span}\{\delta^n\}} = c_0(\mathbb{N}) \neq \ell^\infty(\mathbb{N})$, where $c_0(\mathbb{N})$ is the subspace of sequences converging to 0.

Moreover, there is also no other way to identify $\ell^\infty(\mathbb{N})^*$ with $\ell^1(\mathbb{N})$, since $\ell^1(\mathbb{N})$ is separable whereas $\ell^\infty(\mathbb{N})$ is not. This will follow from Lemma 4.15 (iii) below. ◇

Another useful consequence is

Corollary 4.12. *Let $Y \subseteq X$ be a subspace of a normed vector space and let $x_0 \in X\backslash\overline{Y}$. Then there exists an $\ell \in X^*$ such that (i) $\ell(y) = 0$, $y \in Y$, (ii) $\ell(x_0) = \mathrm{dist}(x_0, Y)$, and (iii) $\|\ell\| = 1$.*

Proof. Replacing Y by $\overline{Y}$ we see that it is no restriction to assume that Y is closed. (Note that $x_0 \in X\backslash\overline{Y}$ if and only if $\mathrm{dist}(x_0, Y) > 0$.) Let $\tilde{Y} = \mathrm{span}\{x_0, Y\}$ and define

$$\ell(y + \alpha x_0) = \alpha\ \mathrm{dist}(x_0, Y).$$

By construction ℓ is linear on $\tilde{Y}$ and satisfies (i) and (ii). Moreover, by $\mathrm{dist}(x_0, Y) \le \|x_0 - \frac{-y}{\alpha}\|$ for every $y \in Y$ we have

$$|\ell(y + \alpha x_0)| = |\alpha| \ \mathrm{dist}(x_0, Y) \le \|y + \alpha x_0\|, \quad y \in Y.$$

Hence $\|\ell\| \le 1$ and there is an extension to X^* by Corollary 4.10. To see that the norm is in fact equal to one, take a sequence $y_n \in Y$ such that $\mathrm{dist}(x_0, Y) \ge (1 - \frac{1}{n})\|x_0 + y_n\|$. Then

$$|\ell(y_n + x_0)| = \mathrm{dist}(x_0, Y) \ge (1 - \frac{1}{n})\|y_n + x_0\|$$

establishing (iii). □

A straightforward consequence of the last corollary is also worthwhile noting:

Corollary 4.13. *Let $Y \subseteq X$ be a subspace of a normed vector space. Then $x \in \overline{Y}$ if and only if $\ell(x) = 0$ for every $\ell \in X^*$ which vanishes on Y.*

If we take the **bidual** (or **double dual**) X^{**}, then the Hahn–Banach theorem tells us, that X can be identified with a subspace of X^{**}. In fact, consider the linear map $J : X \to X^{**}$ defined by $J(x)(\ell) = \ell(x)$ (i.e., $J(x)$ is evaluation at x). Then

Theorem 4.14. *Let X be a Banach space. Then $J : X \to X^{**}$ is isometric (norm preserving).*

Proof. Fix $x_0 \in X$. By $|J(x_0)(\ell)| = |\ell(x_0)| \le \|\ell\|_* \|x_0\|$ we have at least $\|J(x_0)\|_{**} \le \|x_0\|$. Next, by Hahn–Banach there is a linear functional ℓ_0 with norm $\|\ell_0\|_* = 1$ such that $\ell_0(x_0) = \|x_0\|$. Hence $|J(x_0)(\ell_0)| = |\ell_0(x_0)| = \|x_0\|$ shows $\|J(x_0)\|_{**} = \|x_0\|$. □

Thus $J : X \to X^{**}$ is an isometric embedding. In many cases we even have $J(X) = X^{**}$ and X is called **reflexive** in this case.

Example. The Banach spaces $\ell^p(\mathbb{N})$ with $1 < p < \infty$ are reflexive: Identify $\ell^p(\mathbb{N})^*$ with $\ell^q(\mathbb{N})$ and choose $z \in \ell^p(\mathbb{N})^{**}$. Then there is some $x \in \ell^p(\mathbb{N})$ such that

$$z(y) = \sum_{j \in \mathbb{N}} y_j x_j, \qquad y \in \ell^q(\mathbb{N}) \cong \ell^p(\mathbb{N})^*.$$

But this implies $z(y) = y(x)$, that is, $z = J(x)$, and thus J is surjective. (Warning: It does not suffice to just argue $\ell^p(\mathbb{N})^{**} \cong \ell^q(\mathbb{N})^* \cong \ell^p(\mathbb{N})$.)

However, ℓ^1 is not reflexive since $\ell^1(\mathbb{N})^* \cong \ell^\infty(\mathbb{N})$ but $\ell^\infty(\mathbb{N})^* \not\cong \ell^1(\mathbb{N})$ as noted earlier. ⋄

Example. By the same argument (using the Riesz lemma), every Hilbert space is reflexive. ⋄

Lemma 4.15. *Let X be a Banach space.*

(i) *If X is reflexive, so is every closed subspace.*

(ii) *X is reflexive if and only if X^* is.*

(iii) *If X^* is separable, so is X.*

Proof. (i) Let Y be a closed subspace. Denote by $j : Y \hookrightarrow X$ the natural inclusion and define $j_{**} : Y^{**} \to X^{**}$ via $(j_{**}(y''))(\ell) = y''(\ell|_Y)$ for $y'' \in Y^{**}$ and $\ell \in X^*$. Note that j_{**} is isometric by Corollary 4.10. Then

$$\begin{array}{ccc} X & \xrightarrow{J_X} & X^{**} \\ j\uparrow & & \uparrow j_{**} \\ Y & \xrightarrow[J_Y]{} & Y^{**} \end{array}$$

commutes. In fact, we have $j_{**}(J_Y(y))(\ell) = J_Y(y)(\ell|_Y) = \ell(y) = J_X(y)(\ell)$. Moreover, since J_X is surjective, for every $y'' \in Y^{**}$ there is an $x \in X$ such that $j_{**}(y'') = J_X(x)$. Since $j_{**}(y'')(\ell) = y''(\ell|_Y)$ vanishes on all $\ell \in X^*$ which vanish on Y, so does $\ell(x) = J_X(x)(\ell) = j_{**}(y'')(\ell)$ and thus $x \in Y$ by Corollary 4.13. That is, $j_{**}(Y^{**}) = J_X(Y)$ and $J_Y = j \circ J_X \circ j_{**}^{-1}$ is surjective.

(ii) Suppose X is reflexive. Then the two maps

$$\begin{array}{cccc} (J_X)_* : & X^* & \to & X^{***} \\ & x' & \mapsto & x' \circ J_X^{-1} \end{array} \qquad \begin{array}{cccc} (J_X)^* : & X^{***} & \to & X^* \\ & x''' & \mapsto & x''' \circ J_X \end{array}$$

are inverse of each other. Moreover, fix $x'' \in X^{**}$ and let $x = J_X^{-1}(x'')$. Then $J_{X^*}(x')(x'') = x''(x') = J(x)(x') = x'(x) = x'(J_X^{-1}(x''))$, that is $J_{X^*} = (J_X)_*$ respectively $(J_{X^*})^{-1} = (J_X)^*$, which shows X^* reflexive if X reflexive. To see the converse, observe that X^* reflexive implies X^{**} reflexive and hence $J_X(X) \cong X$ is reflexive by (i).

(iii) Let $\{\ell_n\}_{n=1}^\infty$ be a dense set in X^*. Then we can choose $x_n \in X$ such that $\|x_n\| = 1$ and $\ell_n(x_n) \ge \|\ell_n\|/2$. We will show that $\{x_n\}_{n=1}^\infty$ is total in X. If it were not, we could find some $x \in X \setminus \overline{\operatorname{span}\{x_n\}_{n=1}^\infty}$ and hence there is a functional $\ell \in X^*$ as in Corollary 4.12. Choose a subsequence $\ell_{n_k} \to \ell$. Then

$$\|\ell - \ell_{n_k}\| \ge |(\ell - \ell_{n_k})(x_{n_k})| = |\ell_{n_k}(x_{n_k})| \ge \|\ell_{n_k}\|/2,$$

which implies $\ell_{n_k} \to 0$ and contradicts $\|\ell\| = 1$. □

If X is reflexive, then the converse of (iii) is also true (since $X \cong X^{**}$ separable implies X^* separable), but in general this fails as the example $\ell^1(\mathbb{N})^* = \ell^\infty(\mathbb{N})$ shows.

Problem 4.6. *Let X be some Banach space. Show that*

$$\|x\| = \sup_{\ell \in V,\, \|\ell\|=1} |\ell(x)|, \tag{4.11}$$

where $V \subset Y^*$ is some dense subspace. Show that equality is attained if $V = Y^*$.

Problem 4.7. *Let X, Y be some Banach spaces and $A : \mathfrak{D}(A) \subseteq X \to Y$. Show*

$$\|A\| = \sup_{x \in U,\, \|x\|=1;\, \ell \in V,\, \|\ell\|=1} |\ell(Ax)|, \tag{4.12}$$

where $U \subseteq \mathfrak{D}(A)$ and $V \subset Y^*$ are some dense subspaces.

Problem 4.8. *Show that $\|l_y\| = \|y\|_q$, where $l_y \in \ell^p(\mathbb{N})^*$ as defined in (4.6). (Hint: Choose $x \in \ell^p$ such that $x_n y_n = |y_n|^q$.)*

Problem 4.9. *Show that every $l \in \ell^p(\mathbb{N})^*$, $1 \le p < \infty$, can be written as*

$$l(x) = \sum_{n \in \mathbb{N}} y_n x_n$$

with some $y \in \ell^q(\mathbb{N})$. (Hint: To see $y \in \ell^q(\mathbb{N})$ consider x^N defined such that $x_n y_n = |y_n|^q$ for $n \le N$ and $x_n = 0$ for $n > N$. Now look at $|\ell(x^N)| \le \|\ell\| \|x^N\|_p$.)

Problem 4.10. *Let $c_0(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$ be the subspace of sequences which converge to 0, and $c(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$ the subspace of convergent sequences.*

(i) *Show that $c_0(\mathbb{N})$, $c(\mathbb{N})$ are both Banach spaces and that $c(\mathbb{N}) = \operatorname{span}\{c_0(\mathbb{N}), e\}$, where $e = (1, 1, 1, \dots) \in c(\mathbb{N})$.*

(ii) *Show that every $l \in c_0(\mathbb{N})^*$ can be written as*

$$l(x) = \sum_{n \in \mathbb{N}} y_n x_n$$

with some $y \in \ell^1(\mathbb{N})$ which satisfies $\|y\|_1 = \|\ell\|$.

(iii) *Show that every $l \in c(\mathbb{N})^*$ can be written as*

$$l(x) = \sum_{n \in \mathbb{N}} y_n x_n + y_0 \lim_{n \to \infty} x_n$$

with some $y \in \ell^1(\mathbb{N})$ which satisfies $|y_0| + \|y\|_1 = \|\ell\|$.

Problem 4.11. *Let $\{x_n\} \subset X$ be a total set of linearly independent vectors and suppose the complex numbers c_n satisfy $|c_n| \le c\|x_n\|$. Is there a bounded linear functional $\ell \in X^*$ with $\ell(x_n) = c_n$ and $\|\ell\| \le c$? (Hint: Consider e.g. $X = \ell^2(\mathbb{Z})$.)*

4.3. Weak convergence

In the last section we have seen that $\ell(x) = 0$ for all $\ell \in X^*$ implies $x = 0$. Now what about convergence? Does $\ell(x_n) \to \ell(x)$ for every $\ell \in X^*$ imply $x_n \to x$? Unfortunately the answer is no:

Example. Let u_n be an infinite orthonormal set in some Hilbert space. Then $\langle g, u_n\rangle \to 0$ for every g since these are just the expansion coefficients of g which are in ℓ^2 by Bessel's inequality. Since by the Riesz lemma (Theorem 2.10), every bounded linear functional is of this form, we have $\ell(u_n) \to 0$ for every bounded linear functional. (Clearly u_n does not converge to 0, since $\|u_n\| = 1$.) ⋄

If $\ell(x_n) \to \ell(x)$ for every $\ell \in X^*$ we say that x_n **converges weakly** to x and write

$$\operatorname*{w-lim}_{n\to\infty} x_n = x \quad \text{or} \quad x_n \rightharpoonup x. \tag{4.13}$$

Clearly $x_n \to x$ implies $x_n \rightharpoonup x$ and hence this notion of convergence is indeed weaker. Moreover, the weak limit is unique, since $\ell(x_n) \to \ell(x)$ and $\ell(x_n) \to \ell(\tilde{x})$ imply $\ell(x - \tilde{x}) = 0$. A sequence x_n is called a **weak Cauchy sequence** if $\ell(x_n)$ is Cauchy (i.e. converges) for every $\ell \in X^*$.

Lemma 4.16. *Let X be a Banach space.*

(i) *$x_n \rightharpoonup x$ implies $\|x\| \le \liminf \|x_n\|$.*

(ii) *Every weak Cauchy sequence x_n is bounded: $\|x_n\| \le C$.*

(iii) *If X is reflexive, then every weak Cauchy sequence converges weakly.*

(iv) *A sequence x_n is Cauchy if and only if $\ell(x_n)$ is Cauchy, uniformly for $\ell \in X^*$ with $\|\ell\| = 1$.*

Proof. (i) Choose $\ell \in X^*$ such that $\ell(x) = \|x\|$ (for the limit x) and $\|\ell\| = 1$. Then

$$\|x\| = \ell(x) = \liminf \ell(x_n) \le \liminf \|x_n\|.$$

(ii) For every ℓ we have that $|J(x_n)(\ell)| = |\ell(x_n)| \le C(\ell)$ is bounded. Hence by the uniform boundedness principle we have $\|x_n\| = \|J(x_n)\| \le C$.
(iii) If x_n is a weak Cauchy sequence, then $\ell(x_n)$ converges and we can define $j(\ell) = \lim \ell(x_n)$. By construction j is a linear functional on X^*. Moreover, by (ii) we have $|j(\ell)| \le \sup \|\ell(x_n)\| \le \|\ell\| \sup \|x_n\| \le C\|\ell\|$ which shows $j \in X^{**}$. Since X is reflexive, $j = J(x)$ for some $x \in X$ and by construction $\ell(x_n) \to J(x)(\ell) = \ell(x)$, that is, $x_n \rightharpoonup x$. (iv) This follows from

$$\|x_n - x_m\| = \sup_{\|\ell\|=1} |\ell(x_n - x_m)|$$

(cf. Problem 4.6). □

Remark: One can equip X with the weakest topology for which all $\ell \in X^*$ remain continuous. This topology is called the **weak topology** and it is given by taking all finite intersections of inverse images of open sets as a base. By construction, a sequence will converge in the weak topology if and only if it converges weakly. By Corollary 4.12 the weak topology is

Hausdorff, but it will not be metrizable in general. In particular, sequences do not suffice to describe this topology.

In a Hilbert space there is also a simple criterion for a weakly convergent sequence to converge in norm.

Lemma 4.17. *Let $\mathfrak{H}$ be a Hilbert space and let $f_n \rightharpoonup f$. Then $f_n \to f$ if and only if* $\limsup \|f_n\| \le \|f\|$.

Proof. By (i) of the previous lemma we have $\lim \|f_n\| = \|f\|$ and hence

$$\|f - f_n\|^2 = \|f\|^2 - 2\mathrm{Re}(\langle f, f_n\rangle) + \|f_n\|^2 \to 0.$$

The converse is straightforward. □

Now we come to the main reason why weakly convergent sequences are of interest: A typical approach for solving a given equation in a Banach space is as follows:

(i) Construct a (bounded) sequence x_n of approximating solutions (e.g. by solving the equation restricted to a finite dimensional subspace and increasing this subspace).

(ii) Use a compactness argument to extract a convergent subsequence.

(iii) Show that the limit solves the equation.

Our aim here is to provide some results for the step (ii). In a finite dimensional vector space the most important compactness criterion is boundedness (Heine-Borel theorem, Theorem 1.12). In infinite dimensions this breaks down:

Theorem 4.18. *The closed unit ball in X is compact if and only if X is finite dimensional.*

For the proof we will need

Lemma 4.19. *Let X be a normed vector space and $Y \subset X$ some subspace. If $\overline{Y} \ne X$, then for every $\varepsilon \in (0,1)$ there exists an x_ε with $\|x_\varepsilon\| = 1$ and*

$$\inf_{y\in Y} \|x_\varepsilon - y\| \ge 1 - \varepsilon. \tag{4.14}$$

Proof. Pick $x \in X \backslash \overline{Y}$ and abbreviate $d = \mathrm{dist}(x, Y) > 0$. Choose $y_\varepsilon \in Y$ such that $\|x - y_\varepsilon\| \le \frac{d}{1-\varepsilon}$. Set

$$x_\varepsilon = \frac{x - y_\varepsilon}{\|x - y_\varepsilon\|}.$$

Then x_ε is the vector we look for since

$$\begin{aligned}\|x_\varepsilon - y\| &= \frac{1}{\|x-y_\varepsilon\|}\|x-(y_\varepsilon + \|x-y_\varepsilon\|y)\| \\ &\geq \frac{d}{\|x-y_\varepsilon\|} \geq 1-\varepsilon\end{aligned}$$

as required. □

Proof. (of Theorem 4.18) If X is finite dimensional, then X is isomorphic to $\mathbb{C}^n$ and the closed unit ball is compact by the Heine-Borel theorem (Theorem 1.12).

Conversely, suppose X is infinite dimensional and abbreviate $S^1 = \{x \in X| \, \|x\| = 1\}$. Choose $x_1 \in S^1$ and set $Y_1 = \mathrm{span}\{x_1\}$. Then, by the lemma there is an $x_2 \in S^1$ such that $\|x_2 - x_1\| \geq \frac{1}{2}$. Setting $Y_2 = \mathrm{span}\{x_1, x_2\}$ and invoking again our lemma, there is an $x_3 \in S^1$ such that $\|x_3 - x_j\| \geq \frac{1}{2}$ for $j = 1, 2$. Proceeding by induction, we obtain a sequence $x_n \in S^1$ such that $\|x_n - x_m\| \geq \frac{1}{2}$ for $m \neq n$. In particular, this sequence cannot have any convergent subsequence. (Recall that in a metric space compactness and sequential compactness are equivalent — Lemma 1.11.) □

If we are willing to treat convergence for weak convergence, the situation looks much brighter!

Theorem 4.20. *Let X be a reflexive Banach space. Then every bounded sequence has a weakly convergent subsequence.*

Proof. Let x_n be some bounded sequence and consider $Y = \overline{\mathrm{span}\{x_n\}}$. Then Y is reflexive by Lemma 4.15 (i). Moreover, by construction Y is separable and so is Y^* by the remark after Lemma 4.15.

Let ℓ_k be a dense set in Y^*. Then by the usual diagonal sequence argument we can find a subsequence x_{n_m} such that $\ell_k(x_{n_m})$ converges for every k. Denote this subsequence again by x_n for notational simplicity. Then,

$$\begin{aligned}\|\ell(x_n) - \ell(x_m)\| \leq& \|\ell(x_n) - \ell_k(x_n)\| + \|\ell_k(x_n) - \ell_k(x_m)\| \\ &+ \|\ell_k(x_m) - \ell(x_m)\| \\ \leq& 2C\|\ell - \ell_k\| + \|\ell_k(x_n) - \ell_k(x_m)\|\end{aligned}$$

shows that $\ell(x_n)$ converges for every $\ell \in \overline{\mathrm{span}\{\ell_k\}} = Y^*$. Thus there is a limit by Lemma 4.16 (iii). □

Note that this theorem breaks down if X is not reflexive.

Example. Consider the sequence of vectors δ^n (with $\delta^n_n = 1$ and $\delta^n_m = 0$, $n \neq m$) in $\ell^p(\mathbb{N})$, $1 \leq p < \infty$. Then $\delta^n \rightharpoonup 0$ for $1 < p < \infty$. In fact, since every $l \in \ell^p(\mathbb{N})^*$ is of the form $l = l_y$ for some $y \in \ell^q(\mathbb{N})$ we have $l_y(\delta^n) = y_n \to 0$.

If we consider the same sequence in $\ell^1(\mathbb{N})$ there is no weakly convergent subsequence. In fact, since $l_y(\delta^n) \to 0$ for every sequence $y \in \ell^\infty(\mathbb{N})$ with finitely many nonzero entries, the only possible weak limit is zero. On the other hand choosing the constant sequence $y = (1)_{j=1}^\infty$ we see $l_y(\delta^n) = 1 \not\to 0$, a contradiction. ◇

Example. Let $X = L^1(\mathbb{R})$. Every bounded φ gives rise to a linear functional

$$\ell_\varphi(f) = \int f(x)\varphi(x)\,dx$$

in $L^1(\mathbb{R})^*$. Take some nonnegative u_1 with compact support, $\|u_1\|_1 = 1$, and set $u_k(x) = k u_1(k\,x)$. Then we have

$$\int u_k(x)\varphi(x)\,dx \to \varphi(0)$$

(see Problem 8.9) for every continuous φ. Furthermore, if $u_{k_j} \rightharpoonup u$ we conclude

$$\int u(x)\varphi(x)\,dx = \varphi(0).$$

In particular, choosing $\varphi_k(x) = \max(0, 1-k|x|)$ we infer from the dominated convergence theorem

$$1 = \int u(x)\varphi_k(x)\,dx \to \int u(x)\chi_{\{0\}}(x)\,dx = 0,$$

a contradiction.

In fact, u_k *converges* to the Dirac measure centered at 0, which is not in $L^1(\mathbb{R})$. ◇

Finally, let me remark that similar concepts can be introduced for operators. This is of particular importance for the case of unbounded operators, where convergence in the operator norm makes no sense at all.

A sequence of operators A_n is said to **converge strongly** to A,

$$\operatorname*{s-lim}_{n\to\infty} A_n = A \quad :\Leftrightarrow \quad A_n x \to Ax \quad \forall x \in \mathfrak{D}(A) \subseteq \mathfrak{D}(A_n). \tag{4.15}$$

It is said to **converge weakly** to A,

$$\operatorname*{w-lim}_{n\to\infty} A_n = A \quad :\Leftrightarrow \quad A_n x \rightharpoonup Ax \quad \forall x \in \mathfrak{D}(A) \subseteq \mathfrak{D}(A_n). \tag{4.16}$$

Clearly norm convergence implies strong convergence and strong convergence implies weak convergence.

Example. Consider the operator $S_n \in \mathfrak{L}(\ell^2(\mathbb{N}))$ which shifts a sequence n places to the left, that is,

$$S_n\,(x_1, x_2, \dots) = (x_{n+1}, x_{n+2}, \dots) \tag{4.17}$$

and the operator $S_n^* \in \mathfrak{L}(\ell^2(\mathbb{N}))$ which shifts a sequence n places to the right and fills up the first n places with zeros, that is,

$$S_n^*(x_1, x_2, \dots) = (\underbrace{0, \dots, 0}_{n \text{ places}}, x_1, x_2, \dots). \tag{4.18}$$

Then S_n converges to zero strongly but not in norm (since $\|S_n\| = 1$) and S_n^* converges weakly to zero (since $\langle x, S_n^* y\rangle = \langle S_n x, y\rangle$) but not strongly (since $\|S_n^* x\| = \|x\|$) . ◇

Lemma 4.21. *Suppose $A_n \in \mathfrak{L}(X)$ is a sequence of bounded operators.*

(i) $\text{s-lim}_{n\to\infty} A_n = A$ *implies* $\|A\| \le \liminf \|A_n\|$.

(ii) *Every strong Cauchy sequence A_n is bounded:* $\|A_n\| \le C$.

(iii) *If $A_n y \to Ay$ for y in a dense set and $\|A_n\| \le C$, then* $\text{s-lim}_{n\to\infty} A_n = A$.

The same result holds if strong convergence is replaced by weak convergence.

Proof. (i) and (ii) follow as in Lemma 4.16 (i).
(iii) Just use

$$\begin{aligned}\|A_n x - Ax\| &\le \|A_n x - A_n y\| + \|A_n y - Ay\| + \|Ay - Ax\|\\ &\le 2C\|x - y\| + \|A_n y - Ay\|\end{aligned}$$

and choose y in the dense subspace such that $\|x - y\| \le \frac{\varepsilon}{4C}$ and n large such that $\|A_n y - Ay\| \le \frac{\varepsilon}{2}$.

The case of weak convergence is left as an exercise. □

For an application of this lemma see Problem 8.11.

Lemma 4.22. *Suppose $A_n, B_n \in \mathfrak{L}(X)$ are two sequences of bounded operators.*

(i) $\text{s-lim}_{n\to\infty} A_n = A$ *and* $\text{s-lim}_{n\to\infty} B_n = B$ *implies* $\text{s-lim}_{n\to\infty} A_n B_n = AB$.

(ii) $\text{w-lim}_{n\to\infty} A_n = A$ *and* $\text{s-lim}_{n\to\infty} B_n = B$ *implies* $\text{w-lim}_{n\to\infty} A_n B_n = AB$.

(iii) $\lim_{n\to\infty} A_n = A$ *and* $\text{w-lim}_{n\to\infty} B_n = B$ *implies* $\text{w-lim}_{n\to\infty} A_n B_n = AB$.

Proof. For the first case just observe

$$\|(A_n B_n - AB)x\| \le \|(A_n - A)Bx\| + \|A_n\|\|(B_n - B)x\| \to 0.$$

The remaining cases are similar and again left as an exercise. □

Example. Consider again the last example. Then

$$S_n^* S_n\,(x_1, x_2, \dots) = (\underbrace{0, \dots, 0}_{n \text{ places}}, x_{n+1}, x_{n+2}, \dots)$$

converges to 0 weakly (in fact even strongly) but

$$S_n S_n^*(x_1, x_2, \dots) = (x_1, x_2, \dots)$$

does not! Hence the order in the second claim is important. ⋄

Remark: For a sequence of linear functionals ℓ_n, strong convergence is also called **weak-**$*$ convergence. That is, the weak-$*$ limit of ℓ_n is ℓ if

$$\ell_n(x) \to \ell(x) \qquad \forall x \in X. \tag{4.19}$$

Note that this is not the same as weak convergence on X^*, since ℓ is the weak limit of ℓ_n if

$$j(\ell_n) \to j(\ell) \qquad \forall j \in X^{**}, \tag{4.20}$$

whereas for the weak-$*$ limit this is only required for $j \in J(X) \subseteq X^{**}$ (recall $J(x)(\ell) = \ell(x)$). So the weak topology on X^* is the weakest topology for which all $j \in X^{**}$ remain continuous and the weak-$*$ topology on X^* is the weakest topology for which all $j \in J(X)$ remain continuous. In particular, the weak-$*$ topology is weaker than the weak topology and both are equal if X is reflexive.

With this notation it is also possible to slightly generalize Theorem 4.20 (Problem 4.15):

Theorem 4.23. *Suppose X is separable. Then every bounded sequence $\ell_n \in X^*$ has a weak-$*$ convergent subsequence.*

Example. Let us return to the example after Theorem 4.20. Consider the Banach space of bounded continuous functions $X = C(\mathbb{R})$. Using $\ell_f(\varphi) = \int \varphi f\, dx$ we can regard $L^1(\mathbb{R})$ as a subspace of X^*. Then the Dirac measure centered at 0 is also in X^* and it is the weak-$*$ limit of the sequence u_k. ⋄

Problem 4.12. *Suppose $\ell_n \to \ell$ in X^* and $x_n \rightharpoonup x$ in X. Then $\ell_n(x_n) \to \ell(x)$.*

Similarly, suppose s-lim $\ell_n \to \ell$ *and $x_n \to x$. Then $\ell_n(x_n) \to \ell(x)$.*

Problem 4.13. *Show that $x_n \rightharpoonup x$ implies $Ax_n \rightharpoonup Ax$ for $A \in \mathfrak{L}(X)$.*

Problem 4.14. *Show that if $\{\ell_j\} \subseteq X^*$ is some total set, then $x_n \rightharpoonup x$ if and only if x_n is bounded and $\ell_j(x_n) \to \ell_j(x)$ for all j. Show that this is wrong without the boundedness assumption (Hint: Take e.g. $X = \ell^2(\mathbb{N})$).*

Problem 4.15. *Prove Theorem 4.23*

More on compact operators

5.1. Canonical form of compact operators

Our first aim is to find a generalization of Corollary 3.9 for general compact operators. The key observation is that if K is compact, then K^*K is compact and symmetric and thus, by Corollary 3.9, there is a countable orthonormal set $\{u_j\}$ and nonzero numbers $s_j \neq 0$ such that

$$K^*Kf = \sum_{j=1}^{\infty} s_j^2 \langle u_j, f\rangle u_j. \tag{5.1}$$

Moreover, $\|Ku_j\|^2 = \langle u_j, K^*Ku_j\rangle = \langle u_j, s_j^2 u_j\rangle = s_j^2$ implies

$$s_j = \|Ku_j\| > 0. \tag{5.2}$$

The numbers $s_j = s_j(K)$ are called **singular values** of K. There are either finitely many singular values or they converge to zero.

Theorem 5.1 (Canonical form of compact operators)**.** *Let K be compact and let s_j be the singular values of K and $\{u_j\}$ corresponding orthonormal eigenvectors of K^*K. Then*

$$K = \sum_j s_j \langle u_j, .\rangle v_j, \tag{5.3}$$

where $v_j = s_j^{-1} K u_j$. The norm of K is given by the largest singular value

$$\|K\| = \max_j s_j(K). \tag{5.4}$$

Moreover, the vectors v_j *are again orthonormal and satisfy* $K^* v_j = s_j u_j$*. In particular,* v_j *are eigenvectors of* KK^* *corresponding to the eigenvalues* s_j^2*.*

Proof. For any $f \in \mathfrak{H}$ we can write

$$f = \sum_j \langle u_j, f\rangle u_j + f_\perp$$

with $f_\perp \in \mathrm{Ker}(K^*K) = \mathrm{Ker}(K)$ (Problem 5.1). Then

$$Kf = \sum_j \langle u_j, f\rangle K u_j = \sum_j s_j \langle u_j, f\rangle v_j$$

as required. Furthermore,

$$\langle v_j, v_k\rangle = (s_j s_k)^{-1}\langle K u_j, K u_k\rangle = (s_j s_k)^{-1}\langle K^* K u_j, u_k\rangle = s_j s_k^{-1}\langle u_j, u_k\rangle$$

shows that $\{v_j\}$ are orthonormal. Finally, (5.4) follows from

$$\|Kf\|^2 = \|\sum_j s_j\langle u_j, f\rangle v_j\|^2 = \sum_j s_j^2 |\langle u_j, f\rangle|^2 \le \Big(\max_j s_j(K)^2\Big)\|f\|^2,$$

where equality holds for $f = u_{j_0}$ if $s_{j_0} = \max_j s_j(K)$. □

If K is self-adjoint, then $u_j = \sigma_j v_j$, $\sigma_j^2 = 1$, are the eigenvectors of K and $\sigma_j s_j$ are the corresponding eigenvalues.

An operator $K \in \mathfrak{L}(\mathfrak{H})$ is called a **finite rank operator** if its range is finite dimensional. The dimension

$$\mathrm{rank}(K) = \dim \mathrm{Ran}(K)$$

is called the **rank** of K. Since for a compact operator

$$\mathrm{Ran}(K) = \mathrm{span}\{v_j\} \tag{5.5}$$

we see that a compact operator is finite rank if and only if the sum in (5.3) is finite. Note that the finite rank operators form an ideal in $\mathfrak{L}(\mathfrak{H})$ just as the compact operators do. Moreover, every finite rank operator is compact by the Heine–Borel theorem (Theorem 1.12).

Lemma 5.2. *The closure of the ideal of finite rank operators in* $\mathfrak{L}(\mathfrak{H})$ *is the ideal of compact operators.*

Proof. Since the limit of compact operators is compact, it remains to show that every compact operator K can be approximated by finite rank ones. To this end assume that K is not finite rank and note that

$$K_n = \sum_{j=1}^n s_j \langle u_j, .\rangle v_j$$

converges to K as $n \to \infty$ since

$$\|K - K_n\| = \max_{j \geq n} s_j(K)$$

by (5.4). □

Moreover, this also shows that the adjoint of a compact operator is again compact.

Corollary 5.3. *An operator K is compact (finite rank) if and only K^* is. In fact, $s_j(K) = s_j(K^*)$ and*

$$K^* = \sum_j s_j \langle v_j, .\rangle u_j. \tag{5.6}$$

Proof. First of all note that (5.6) follows from (5.3) since taking adjoints is continuous and $(\langle u_j, .\rangle v_j)^* = \langle v_j, .\rangle u_j$. The rest is straightforward. □

Finally, let me remark that there are a number of other equivalent definitions for compact operators.

Lemma 5.4. *For $K \in \mathfrak{L}(\mathfrak{H})$ the following statements are equivalent:*

(i) *K is compact.*

(ii) *$A_n \in \mathfrak{L}(\mathfrak{H})$ and $A_n \overset{s}{\to} A$ strongly implies $A_n K \to AK$.*

(iii) *$f_n \rightharpoonup f$ weakly implies $K f_n \to K f$ in norm.*

Proof. (i) ⇒ (ii). Translating $A_n \to A_n - A$, it is no restriction to assume $A = 0$. Since $\|A_n\| \leq M$, it suffices to consider the case where K is finite rank. Then using (5.3) and applying the triangle plus Cauchy–Schwarz inequalities

$$\|A_n K\|^2 \leq \sup_{\|f\|=1} \left(\sum_{j=1}^N s_j |\langle u_j, f\rangle| \|A_n v_j\| \right)^2 \leq \sum_{j=1}^N s_j^2 \|A_n v_j\|^2 \to 0.$$

(ii) ⇒ (iii). Again, replace $f_n \to f_n - f$ and assume $f = 0$. Choose $A_n = \langle f_n, .\rangle u$, $\|u\| = 1$. Then $\|K f_n\| = \|A_n K^*\| \to 0$.

(iii) ⇒ (i). If f_n is bounded, it has a weakly convergent subsequence by Theorem 4.20. Now apply (iii) to this subsequence. □

The last condition explains the name compact. Moreover, note that one cannot replace $A_n K \to AK$ by $K A_n \to KA$ in (ii) unless one additionally requires A_n to be normal (then this follows by taking adjoints — recall that only for normal operators is taking adjoints continuous with respect to strong convergence). Without the requirement that A_n be normal, the claim is wrong as the following example shows.

Example. Let $\mathfrak{H} = \ell^2(\mathbb{N})$ and let A_n be the operator which shifts each sequence n places to the left and let $K = \langle \delta_1, . \rangle \delta_1$, where $\delta_1 = (1, 0, \dots)$. Then s-lim $A_n = 0$ but $\|KA_n\| = 1$. ⋄

Problem 5.1. *Show that* $\mathrm{Ker}(A^* A) = \mathrm{Ker}(A)$ *for any* $A \in \mathfrak{L}(\mathfrak{H})$.

Problem 5.2. *Show* (5.4).

5.2. Hilbert–Schmidt and trace class operators

We can further subdivide the class of operators according to the decay of the singular values. We define

$$\|K\|_p = \Big(\sum_j s_j(K)^p \Big)^{1/p} \tag{5.7}$$

plus corresponding spaces

$$\mathcal{J}_p(\mathfrak{H}) = \{K \in \mathfrak{C}(\mathfrak{H}) | \|K\|_p < \infty\}, \tag{5.8}$$

which are known as **Schatten p-classes**. Note that by (5.4)

$$\|K\| \le \|K\|_p \tag{5.9}$$

and that by $s_j(K) = s_j(K^*)$ we have

$$\|K\|_p = \|K^*\|_p. \tag{5.10}$$

The two most important cases are $p = 1$ and $p = 2$: $\mathcal{J}_2(\mathfrak{H})$ is the space of **Hilbert–Schmidt operators** and $\mathcal{J}_1(\mathfrak{H})$ is the space of **trace class** operators.

We first prove an alternate definition for the Hilbert–Schmidt norm.

Lemma 5.5. *A bounded operator K is Hilbert–Schmidt if and only if*

$$\sum_{j \in J} \|K w_j\|^2 < \infty \tag{5.11}$$

for some orthonormal basis and

$$\|K\|_2 = \Big(\sum_{j \in J} \|K w_j\|^2 \Big)^{1/2}, \tag{5.12}$$

for every orthonormal basis in this case.

Proof. First of all note that (5.11) implies that K is compact. To see this let P_n be the projection onto the space spanned by the first n elements of the orthonormal basis $\{w_j\}$. Then $K_n = K P_n$ is finite rank and converges to K since

$$\|(K - K_n) f\| = \| \sum_{j > n} c_j K w_j \| \le \sum_{j > n} |c_j| \|K w_j\| \le \Big(\sum_{j > n} \|K w_j\|^2 \Big)^{1/2} \|f\|,$$

where $f = \sum_j c_j w_j$.

The rest follows from (5.3) and

$$\sum_j \|Kw_j\|^2 = \sum_{k,j} |\langle v_k, Kw_j\rangle|^2 = \sum_{k,j} |\langle K^* v_k, w_j\rangle|^2 = \sum_k \|K^* v_k\|^2$$
$$= \sum_k s_k(K)^2 = \|K\|_2^2.$$

Here we have used $\mathrm{span}\{v_k\} = \mathrm{Ker}(K^*)^\perp = \overline{\mathrm{Ran}(K)}$ in the first step. □

Now we can show

Lemma 5.6. *The set of Hilbert–Schmidt operators forms an ideal in* $\mathfrak{L}(\mathfrak{H})$ *and*

$$\|KA\|_2 \le \|A\|\|K\|_2, \quad \textit{respectively}, \quad \|AK\|_2 \le \|A\|\|K\|_2. \tag{5.13}$$

Proof. Let K be Hilbert–Schmidt and A bounded. Then AK is compact and

$$\|AK\|_2^2 = \sum_j \|AKw_j\|^2 \le \|A\|^2 \sum_j \|Kw_j\|^2 = \|A\|^2 \|K\|_2^2.$$

For KA just consider adjoints. □

Since Hilbert–Schmidt operators turn out easy to identify (cf. Section 8.4), it is important to relate $\mathcal{J}_1(\mathfrak{H})$ with $\mathcal{J}_2(\mathfrak{H})$:

Lemma 5.7. *An operator is trace class if and only if it can be written as the product of two Hilbert–Schmidt operators,* $K = K_1 K_2$, *and in this case we have*

$$\|K\|_1 \le \|K_1\|_2 \|K_2\|_2. \tag{5.14}$$

Proof. Using (5.3) (where we can extend u_n and v_n to orthonormal bases if necessary) and Cauchy–Schwarz we have

$$\|K\|_1 = \sum_n \langle v_n, K u_n\rangle = \sum_n |\langle K_1^* v_n, K_2 u_n\rangle|$$
$$\le \Big(\sum_n \|K_1^* v_n\|^2 \sum_n \|K_2 u_n\|^2\Big)^{1/2} = \|K_1\|_2 \|K_2\|_2$$

and hence $K = K_1 K_2$ is trace class if both K_1 and K_2 are Hilbert–Schmidt operators. To see the converse, let K be given by (5.3) and choose $K_1 = \sum_j \sqrt{s_j(K)} \langle u_j, .\rangle v_j$, respectively, $K_2 = \sum_j \sqrt{s_j(K)} \langle u_j, .\rangle u_j$. □

Now we can also explain the name trace class:

Lemma 5.8. *If K is trace class, then for every orthonormal basis $\{w_n\}$ the* **trace**

$$\mathrm{tr}(K) = \sum_n \langle w_n, K w_n\rangle \tag{5.15}$$

is finite and independent of the orthonormal basis.

Proof. Let $\{v_n\}$ be another ONB. If we write $K = K_1 K_2$ with K_1, K_2 Hilbert–Schmidt, we have

$$\begin{aligned}\sum_n \langle w_n, K_1 K_2 w_n\rangle &= \sum_n \langle K_1^* w_n, K_2 w_n\rangle = \sum_{n,m} \langle K_1^* w_n, v_m\rangle\langle v_m, K_2 w_n\rangle \\ &= \sum_{m,n} \langle K_2^* v_m, w_n\rangle\langle w_n, K_1 v_m\rangle = \sum_m \langle K_2^* v_m, K_1 v_m\rangle \\ &= \sum_m \langle v_m, K_2 K_1 v_m\rangle.\end{aligned}$$

Hence the trace is independent of the ONB and we even have $\mathrm{tr}(K_1 K_2) = \mathrm{tr}(K_2 K_1)$. □

Clearly for self-adjoint trace class operators, the trace is the sum over all eigenvalues (counted with their multiplicity). To see this, one just has to choose the orthonormal basis to consist of eigenfunctions. This is even true for all trace class operators and is known as Lidskij trace theorem (see [8] for an easy to read introduction).

Problem 5.3. *Let $\mathfrak{H} = \ell^2(\mathbb{N})$ and let A be multiplication by a sequence $a = (a_j)_{j=1}^\infty$. Show that A is Hilbert–Schmidt if and only if $a \in \ell^2(\mathbb{N})$. Furthermore, show that $\|A\|_2 = \|a\|$ in this case.*

Problem 5.4. *Show that $K : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$, $f_n \mapsto \sum_{j\in\mathbb{N}} k_{n+j} f_j$ is Hilbert–Schmidt with $\|K\|_2 \le \|c\|_1$ if $|k_j| \le c_j$, where c_j is decreasing and summable.*

5.3. Fredholm theory for compact operators

In this section we want to investigate solvability of the equation

$$f = Kf + g \tag{5.16}$$

for given g. Clearly there exists a solution if $g \in \mathrm{Ran}(1-K)$ and this solution is unique if $\mathrm{Ker}(1-K) = \{0\}$. Hence these subspaces play a crucial role. Moreover, if the underlying Hilbert space is finite dimensional it is well-known that $\mathrm{Ker}(1-K) = \{0\}$ automatically implies $\mathrm{Ran}(1-K) = \mathfrak{H}$ since

$$\dim \mathrm{Ker}(1-K) + \dim \mathrm{Ran}(1-K) = \dim \mathfrak{H}. \tag{5.17}$$

Unfortunately this formula is of no use if $\mathfrak{H}$ is infinite dimensional, but if we rewrite it as

$$\dim \operatorname{Ker}(1-K) = \dim \mathfrak{H} - \dim \operatorname{Ran}(1-K) = \dim \operatorname{Ran}(1-K)^{\perp} \quad (5.18)$$

there is some hope. In fact, we will show that this formula (makes sense and) holds if K is a compact operator.

Lemma 5.9. *Let $K \in \mathfrak{C}(\mathfrak{H})$ be compact. Then* $\operatorname{Ker}(1-K)$ *is finite dimensional and* $\operatorname{Ran}(1-K)$ *is closed.*

Proof. We first show $\dim \operatorname{Ker}(1-K) < \infty$. If not we could find an infinite orthonormal system $\{u_j\}_{j=1}^{\infty} \subset \operatorname{Ker}(1-K)$. By $K u_j = u_j$ compactness of K implies that there is a convergent subsequence u_{j_k}. But this is impossible by $\|u_j - u_k\|^2 = 2$ for $j \neq k$.

To see that $\operatorname{Ran}(1-K)$ is closed we first claim that there is a $\gamma > 0$ such that

$$\|(1-K)f\| \geq \gamma \|f\|, \qquad \forall f \in \operatorname{Ker}(1-K)^{\perp}. \quad (5.19)$$

In fact, if there were no such γ, we could find a normalized sequence $f_j \in \operatorname{Ker}(1-K)^{\perp}$ with $\|f_j - K f_j\| < \frac{1}{j}$, that is, $f_j - K f_j \to 0$. After passing to a subsequence we can assume $K f_j \to f$ by compactness of K. Combining this with $f_j - K f_j \to 0$ implies $f_j \to f$ and $f - K f = 0$, that is, $f \in \operatorname{Ker}(1-K)$. On the other hand, since $\operatorname{Ker}(1-K)^{\perp}$ is closed, we also have $f \in \operatorname{Ker}(1-K)^{\perp}$ which shows $f = 0$. This contradicts $\|f\| = \lim \|f_j\| = 1$ and thus (5.19) holds.

Now choose a sequence $g_j \in \operatorname{Ran}(1-K)$ converging to some g. By assumption there are f_k such that $(1-K)f_k = g_k$ and we can even assume $f_k \in \operatorname{Ker}(1-K)^{\perp}$ by removing the projection onto $\operatorname{Ker}(1-K)$. Hence (5.19) shows

$$\|f_j - f_k\| \leq \gamma^{-1} \|(1-K)(f_j - f_k)\| = \gamma^{-1} \|g_j - g_k\|$$

that f_j converges to some f and $(1-K)f = g$ implies $g \in \operatorname{Ran}(1-K)$. □

Since

$$\operatorname{Ran}(1-K)^{\perp} = \operatorname{Ker}(1-K^*) \quad (5.20)$$

by (2.27) we see that the left and right hand side of (5.18) are at least finite for compact K and we can try to verify equality.

Theorem 5.10. *Suppose K is compact. Then*

$$\dim \operatorname{Ker}(1-K) = \dim \operatorname{Ran}(1-K)^{\perp}, \quad (5.21)$$

where both quantities are finite.

Proof. It suffices to show

$$\dim \operatorname{Ker}(1-K) \ge \dim \operatorname{Ran}(1-K)^{\perp}, \tag{5.22}$$

since replacing K by K^* in this inequality and invoking (2.27) provides the reversed inequality.

We begin by showing that $\dim \operatorname{Ker}(1-K) = 0$ implies $\dim \operatorname{Ran}(1-K)^{\perp} = 0$, that is $\operatorname{Ran}(1-K) = \mathfrak{H}$. To see this suppose $\mathfrak{H}_1 = \operatorname{Ran}(1-K) = (1-K)\mathfrak{H}$ is not equal to $\mathfrak{H}$. Then $\mathfrak{H}_2 = (1-K)\mathfrak{H}_1$ can also not be equal to $\mathfrak{H}_1$. Otherwise for any given element in $\mathfrak{H}_1^{\perp}$ there would be an element in $\mathfrak{H}_2$ with the same image under $1-K$ contradicting our assumption that $1-K$ is injective. Proceeding inductively we obtain a sequence of subspaces $\mathfrak{H}_j = (1-K)^j\mathfrak{H}$ with $\mathfrak{H}_{j+1} \subset \mathfrak{H}_j$. Now choose a normalized sequence $f_j \in \mathfrak{H}_j \cap \mathfrak{H}_{j+1}^{\perp}$. Then for $k > j$ we have

$$\begin{aligned}\|Kf_j - Kf_k\|^2 &= \|f_j - f_k - (1-K)(f_j - f_k)\|^2\\ &= \|f_j\|^2 + \|f_k + (1-K)(f_j - f_k)\|^2 \ge 1\end{aligned}$$

since $f_j \in \mathfrak{H}_{j+1}^{\perp}$ and $f_k + (1-K)(f_j - f_k) \in \mathfrak{H}_{j+1}$. But this contradicts the fact that Kf_j must have a convergent subsequence.

To show (5.22) in the general case, suppose $\dim \operatorname{Ker}(1-K) < \dim \operatorname{Ran}(1-K)^{\perp}$ instead. Then we can find a bounded map $A : \operatorname{Ker}(1-K) \to \operatorname{Ran}(1-K)^{\perp}$ which is injective but not onto. Extend A to a map on $\mathfrak{H}$ by setting $Af = 0$ for $f \in \operatorname{Ker}(1-K)^{\perp}$. Since A is finite rank, the operator $\tilde{K} = K + A$ is again compact. We claim $\operatorname{Ker}(1-\tilde{K}) = \{0\}$. Indeed, if $f \in \operatorname{Ker}(1-\tilde{K})$, then $f - Kf = Af \in \operatorname{Ran}(1-K)^{\perp}$ implies $f \in \operatorname{Ker}(1-K) \cap \operatorname{Ker}(A)$. But A is injective on $\operatorname{Ker}(1-K)$ and thus $f = 0$ as claimed. Thus the first step applied to $\tilde{K}$ implies $\operatorname{Ran}(1-\tilde{K}) = \mathfrak{H}$. But this is impossible since the equation

$$f - \tilde{K}f = (1-K)f - Af = g$$

for $g \in \operatorname{Ran}(1-K)^{\perp}$ reduces to $(1-K)f = 0$ and $Af = -g$ which has no solution if we choose $g \notin \operatorname{Ran}(A)$. □

As a special case we obtain the famous

Theorem 5.11 (Fredholm alternative). *Suppose $K \in \mathfrak{C}(\mathfrak{H})$ is compact. Then either the inhomogeneous equation*

$$f = Kf + g \tag{5.23}$$

has a unique solution for every $g \in \mathfrak{H}$ or the corresponding homogenous equation

$$f = Kf \tag{5.24}$$

has a nontrivial solution.

Note that (5.20) implies that in any case the inhomogeneous equation $f = Kf + g$ has a solution if and only if $g \in \mathrm{Ker}(1-K^*)^\perp$. Moreover, combining (5.21) with (5.20) also shows

$$\dim \mathrm{Ker}(1-K) = \dim \mathrm{Ker}(1-K^*) \tag{5.25}$$

for compact K.

This theory can be generalized to the case of operators where both $\mathrm{Ker}(1-K)$ and $\mathrm{Ran}(1-K)^\perp$ are finite dimensional. Such operators are called **Fredholm operators** (also **Noether operators**) and the number

$$\mathrm{ind}(1-K) = \dim \mathrm{Ker}(1-K) - \dim \mathrm{Ran}(1-K)^\perp \tag{5.26}$$

is the called the **index** of K. Theorem 5.10 now says that a compact operator is Fredholm of index zero.

Problem 5.5. *Compute* $\mathrm{Ker}(1-K)$ *and* $\mathrm{Ran}(1-K)^\perp$ *for the operator* $K = \langle v, .\rangle u$, *where* $u, v \in \mathfrak{H}$ *satisfy* $\langle u, v\rangle = 1$.

Chapter 6

Bounded linear operators

6.1. Banach algebras

In this section we want to have a closer look at the set of bounded linear operators $\mathfrak{L}(X)$ from a Banach space X into itself. We already know from Section 1.5 that they form a Banach space which has a multiplication given by composition. In this section we want to further investigate this structure.

A Banach space X together with a multiplication satisfying

$$(x+y)z = xz + yz, \quad x(y+z) = xy + xz, \quad x,y,z \in X, \tag{6.1}$$

and

$$(xy)z = x(yz), \qquad \alpha\,(xy) = (\alpha x)y = x\,(\alpha y), \quad \alpha \in \mathbb{C}. \tag{6.2}$$

and

$$\|xy\| \le \|x\|\|y\|. \tag{6.3}$$

is called a **Banach algebra**. In particular, note that (6.3) ensures that multiplication is continuous (Problem 6.1). An element $e \in X$ satisfying

$$ex = xe = x, \qquad \forall x \in X \tag{6.4}$$

is called **identity** (show that e is unique) and we will assume $\|e\| = 1$ in this case.

Example. The continuous functions $C(I)$ over some compact interval form a commutative Banach algebra with identity 1. ⋄

Example. The bounded linear operators $\mathfrak{L}(X)$ form a Banach algebra with identity $\mathbb{I}$. ⋄

Example. The space $L^1(\mathbb{R}^n)$ together with the convolution

$$(g * f)(x) = \int_{\mathbb{R}^n} g(x-y)f(y)dy = \int_{\mathbb{R}^n} g(y)f(x-y)dy \tag{6.5}$$

is a commutative Banach algebra (Problem 6.4) without identity. ◇

Let X be a Banach algebra with identity e. Then $x \in X$ is called **invertible** if there is some $y \in X$ such that

$$xy = yx = e. \tag{6.6}$$

In this case y is called the inverse of x and is denoted by x^{-1}. It is straightforward to show that the inverse is unique (if one exists at all) and that

$$(xy)^{-1} = y^{-1}x^{-1}. \tag{6.7}$$

Example. Let $X = \mathfrak{L}(\ell^1(\mathbb{N}))$ and let $S^{\pm}$ be defined via

$$S^- x_n = \begin{cases} 0 & n = 1 \\ x_{n-1} & n > 1 \end{cases}, \qquad S^+ x_n = x_{n+1} \tag{6.8}$$

(i.e., S^- shifts each sequence one place right (filling up the first place with a 0) and S^+ shifts one place left (dropping the first place)). Then $S^+S^- = \mathbb{I}$ but $S^-S^+ \neq \mathbb{I}$. So you really need to check both $xy = e$ and $yx = e$ in general. ◇

Lemma 6.1. *Let X be a Banach algebra with identity e. Suppose $\|x\| < 1$. Then $e - x$ is invertible and*

$$(e-x)^{-1} = \sum_{n=0}^{\infty} x^n. \tag{6.9}$$

Proof. Since $\|x\| < 1$ the series converges and

$$(e-x)\sum_{n=0}^{\infty} x^n = \sum_{n=0}^{\infty} x^n - \sum_{n=1}^{\infty} x^n = e$$

respectively

$$\left(\sum_{n=0}^{\infty} x^n\right)(e-x) = \sum_{n=0}^{\infty} x^n - \sum_{n=1}^{\infty} x^n = e.$$

□

Corollary 6.2. *Suppose x is invertible and $\|x^{-1}y\| < 1$ or $\|yx^{-1}\| < 1$. Then $(x-y)$ is invertible as well and*

$$(x-y)^{-1} = \sum_{n=0}^{\infty} (x^{-1}y)^n x^{-1} \quad \text{or} \quad (x-y)^{-1} = \sum_{n=0}^{\infty} x^{-1}(yx^{-1})^n. \tag{6.10}$$

In particular, both conditions are satisfied if $\|y\| < \|x^{-1}\|^{-1}$ and the set of invertible elements is open.

Proof. Just observe $x - y = x(e - x^{-1}y) = (e - yx^{-1})x$. □

The **resolvent set** is defined as

$$\rho(x) = \{\alpha \in \mathbb{C} | \exists (x - \alpha)^{-1}\} \subseteq \mathbb{C}, \tag{6.11}$$

where we have used the shorthand notation $x - \alpha = x - \alpha e$. Its complement is called the **spectrum**

$$\sigma(x) = \mathbb{C} \backslash \rho(x). \tag{6.12}$$

It is important to observe that the fact that the inverse has to exist as an element of X. That is, if X are bounded linear operators, it does not suffice that $x - \alpha$ is bijective, the inverse must also be bounded!

Example. If $X = \mathfrak{L}(\mathbb{C}^n)$ is the space of n by n matrices, then the spectrum is just the set of eigenvalues. ⋄

Example. If $X = C(I)$, then the spectrum of a function $x \in C(I)$ is just its range, $\sigma(x) = x(I)$. ⋄

The map $\alpha \mapsto (x - \alpha)^{-1}$ is called the **resolvent** of $x \in X$. If $\alpha_0 \in \rho(x)$ we can choose $x \to x - \alpha_0$ and $y \to \alpha - \alpha_0$ in (6.10) which implies

$$(x-\alpha)^{-1} = \sum_{n=0}^{\infty} (\alpha - \alpha_0)^n (x - \alpha_0)^{-n-1}, \quad |\alpha - \alpha_0| < \|(x - \alpha_0)^{-1}\|^{-1}. \tag{6.13}$$

This shows that $(x - \alpha)^{-1}$ has a convergent power series with coefficients in X around every point $\alpha_0 \in \rho(x)$. As in the case of coefficients in $\mathbb{C}$, such functions will be called **analytic**. In particular, $\ell((x - \alpha)^{-1})$ is a complex-valued analytic function for every $\ell \in X^*$ and we can apply well-known results from complex analysis:

Theorem 6.3. *For every $x \in X$, the spectrum $\sigma(x)$ is compact, nonempty and satisfies*

$$\sigma(x) \subseteq \{\alpha | \, |\alpha| \le \|x\|\}. \tag{6.14}$$

Proof. Equation (6.13) already shows that $\rho(x)$ is open. Hence $\sigma(x)$ is closed. Moreover, $x - \alpha = -\alpha(e - \frac{1}{\alpha}x)$ together with Lemma 6.1 shows

$$(x - \alpha)^{-1} = -\frac{1}{\alpha} \sum_{n=0}^{\infty} \left(\frac{x}{\alpha}\right)^n, \quad |\alpha| > \|x\|,$$

which implies $\sigma(x) \subseteq \{\alpha | \, |\alpha| \le \|x\|\}$ is bounded and thus compact. Moreover, taking norms shows

$$\|(x - \alpha)^{-1}\| \le \frac{1}{|\alpha|} \sum_{n=0}^{\infty} \frac{\|x\|^n}{|\alpha|^n} = \frac{1}{|\alpha| - \|x\|}, \quad |\alpha| > \|x\|,$$

which implies $(x - \alpha)^{-1} \to 0$ as $\alpha \to \infty$. In particular, if $\sigma(x)$ is empty, then $\ell((x - \alpha)^{-1})$ is an entire analytic function which vanishes at infinity.

By Liouville's theorem we must have $\ell((x-\alpha)^{-1}) = 0$ in this case, and so $(x-\alpha)^{-1} = 0$, which is impossible. □

As another simple consequence we obtain:

Theorem 6.4. *Suppose X is a Banach algebra in which every element except 0 is invertible. Then X is isomorphic to $\mathbb{C}$.*

Proof. Pick $x \in X$ and $\alpha \in \sigma(x)$. Then $x - \alpha$ is not invertible and hence $x-\alpha = 0$, that is $x = \alpha$. Thus every element is a multiple of the identity. □

Theorem 6.5 (Spectral mapping)**.** *For every polynomial p and $x \in X$ we have*

$$\sigma(p(x)) = p(\sigma(x)). \tag{6.15}$$

Proof. Fix $\alpha_0 \in \mathbb{C}$ and observe

$$p(x) - p(\alpha_0) = (x - \alpha_0) q_0(x).$$

If $p(\alpha_0) \notin \sigma(p(x))$ we have

$$(x-\alpha_0)^{-1} = q_0(x)((x-\alpha_0)q_0(x))^{-1} = ((x-\alpha_0)q_0(x))^{-1} q_0(x)$$

(check this — since $q_0(x)$ commutes with $(x-\alpha_0)q_0(x)$ it also commutes with its inverse). Hence $\alpha_0 \notin \sigma(x)$.

Conversely, let $\alpha_0 \in \sigma(p(x))$. Then

$$p(x) - \alpha_0 = a(x-\lambda_1)\cdots(x-\lambda_n)$$

and at least one $\lambda_j \in \sigma(x)$ since otherwise the right-hand side would be invertible. But then $p(\lambda_j) = \alpha_0$, that is, $\alpha_0 \in p(\sigma(x))$. □

Next let us look at the convergence radius of the **Neumann series** for the resolvent

$$(x-\alpha)^{-1} = -\frac{1}{\alpha}\sum_{n=0}^{\infty}\left(\frac{x}{\alpha}\right)^n \tag{6.16}$$

encountered in the proof of Theorem 6.3 (which is just the Laurent expansion around infinity).

The number

$$r(x) = \sup_{\alpha\in\sigma(x)} |\alpha| \tag{6.17}$$

is called the **spectral radius** of x. Note that by (6.14) we have

$$r(x) \le \|x\|. \tag{6.18}$$

Theorem 6.6. *The spectral radius satisfies*

$$r(x) = \inf_{n\in\mathbb{N}} \|x^n\|^{1/n} = \lim_{n\to\infty} \|x^n\|^{1/n}. \tag{6.19}$$

Proof. By spectral mapping we have $r(x)^n = r(x^n) \le \|x^n\|$ and hence

$$r(x) \le \inf \|x^n\|^{1/n}.$$

Conversely, fix $\ell \in X^*$, and consider

$$\ell((x-\alpha)^{-1}) = -\frac{1}{\alpha}\sum_{n=0}^{\infty} \frac{1}{\alpha^n}\ell(x^n). \tag{6.20}$$

Then $\ell((x-\alpha)^{-1})$ is analytic in $|\alpha| > r(x)$ and hence (6.20) converges absolutely for $|\alpha| > r(x)$ by a well-known result from complex analysis. Hence for fixed α with $|\alpha| > r(x)$, $\ell(x^n/\alpha^n)$ converges to zero for every $\ell \in X^*$. Since every weakly convergent sequence is bounded we have

$$\frac{\|x^n\|}{|\alpha|^n} \le C(\alpha)$$

and thus

$$\limsup_{n\to\infty} \|x^n\|^{1/n} \le \limsup_{n\to\infty} C(\alpha)^{1/n}|\alpha| = |\alpha|.$$

Since this holds for every $|\alpha| > r(x)$ we have

$$r(x) \le \inf \|x^n\|^{1/n} \le \liminf_{n\to\infty} \|x^n\|^{1/n} \le \limsup_{n\to\infty} \|x^n\|^{1/n} \le r(x),$$

which finishes the proof. □

To end this section let us look at two examples illustrating these ideas.

Example. Let $X = \mathfrak{L}(\mathbb{C}^2)$ be the space of two by two matrices and consider

$$x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{6.21}$$

Then $x^2 = 0$ and consequently $r(x) = 0$. This is not surprising, since x has the only eigenvalue 0. The same is true for any nilpotent matrix. ⋄

Example. Consider the linear Volterra integral operator

$$K(x)(t) = \int_0^t k(t,s)x(s)ds, \qquad x \in C([0,1]), \tag{6.22}$$

then, using induction, it is not hard to verify (Problem 6.3)

$$|K^n(x)(t)| \le \frac{\|k\|_\infty^n t^n}{n!}\|x\|_\infty. \tag{6.23}$$

Consequently

$$\|K^n x\|_\infty \le \frac{\|k\|_\infty^n}{n!}\|x\|_\infty,$$

that is $\|K^n\| \le \frac{\|k\|_\infty^n}{n!}$, which shows

$$r(K) \le \lim_{n\to\infty} \frac{\|k\|_\infty}{(n!)^{1/n}} = 0.$$

Hence $r(K) = 0$ and for every $\lambda \in \mathbb{C}$ and every $y \in C(I)$ the equation

$$x - \lambda K\, x = y \tag{6.24}$$

has a unique solution given by

$$x = (\mathbb{I} - \lambda K)^{-1} y = \sum_{n=0}^{\infty} \lambda^n K^n\, y. \tag{6.25}$$

⋄

Problem 6.1. *Show that the multiplication in a Banach algebra X is continuous: $x_n \to x$ and $y_n \to y$ imply $x_n y_n \to xy$.*

Problem 6.2. *Suppose x has both a right inverse y (i.e., $xy = e$) and a left inverse z (i.e., $zx = e$). Show that $y = z = x^{-1}$.*

Problem 6.3. *Show* (6.23).

Problem 6.4. *Show that $L^1(\mathbb{R}^n)$ with convolution as multiplication is a commutative Banach algebra without identity (Hint: Problem 8.10).*

6.2. The C^* algebra of operators and the spectral theorem

We begin by recalling that if $\mathfrak{H}$ is some Hilbert space, then for every $A \in \mathfrak{L}(\mathfrak{H})$ we can define its adjoint $A^* \in \mathfrak{L}(\mathfrak{H})$. Hence the Banach algebra $\mathfrak{L}(\mathfrak{H})$ has an additional operation in this case. In general, a Banach algebra X together with an **involution**

$$(x+y)^* = x^* + y^*, \quad (\alpha x)^* = \alpha^* x^*, \quad x^{**} = x, \quad (xy)^* = y^* x^*, \tag{6.26}$$

satisfying

$$\|x\|^2 = \|x^* x\| \tag{6.27}$$

is called a C^* **algebra**. Any subalgebra which is also closed under involution, is called a $*$-subalgebra. Note that (6.27) implies $\|x\|^2 = \|x^* x\| \le \|x\| \|x^*\|$ and hence $\|x\| \le \|x^*\|$. By $x^{**} = x$ this also implies $\|x^*\| \le \|x^{**}\| = \|x\|$ and hence

$$\|x\| = \|x^*\|, \qquad \|x\|^2 = \|x^* x\| = \|x x^*\|. \tag{6.28}$$

Example. The continuous functions $C(I)$ together with complex conjugation form a commutative C^* algebra. ⋄

Example. The Banach algebra $\mathfrak{L}(\mathfrak{H})$ is a C^* algebra by Lemma 2.13. ⋄

If X has an identity e, we clearly have $e^* = e$ and $(x^{-1})^* = (x^*)^{-1}$ (show this). We will always assume that we have an identity. In particular,

$$\sigma(x^*) = \sigma(x)^*. \tag{6.29}$$

If X is a C^* algebra, then $x \in X$ is called **normal** if $x^*x = xx^*$, **self-adjoint** if $x^* = x$, and **unitary** if $x^* = x^{-1}$. Moreover, x is called **positive** if $x = y^2$ for some $y = y^* \in X$. Clearly both self-adjoint and unitary elements are normal and positive elements are self-adjoint.

Lemma 6.7. *If $x \in X$ is normal, then $\|x^2\| = \|x\|^2$ and $r(x) = \|x\|$.*

Proof. Using (6.27) three times we have

$$\|x^2\| = \|(x^2)^*(x^2)\|^{1/2} = \|(xx^*)^*(xx^*)\|^{1/2} = \|x^*x\| = \|x\|^2$$

and hence $r(x) = \lim_{k\to\infty} \|x^{2^k}\|^{1/2^k} = \|x\|$. □

Lemma 6.8. *If x is self-adjoint, then $\sigma(x) \subseteq \mathbb{R}$.*

Proof. Suppose $\alpha + \mathrm{i}\beta \in \sigma(x)$, $\lambda \in \mathbb{R}$. Then $\alpha + \mathrm{i}(\beta + \lambda) \in \sigma(x + \mathrm{i}\lambda)$ and

$$\alpha^2 + (\beta + \lambda)^2 \le \|x + \mathrm{i}\lambda\|^2 = \|(x + \mathrm{i}\lambda)(x - \mathrm{i}\lambda)\| = \|x^2 + \lambda^2\| \le \|x\|^2 + \lambda^2.$$

Hence $\alpha^2 + \beta^2 + 2\beta\lambda \le \|x\|^2$ which gives a contradiction if we let $|\lambda| \to \infty$ unless $\beta = 0$. □

Given $x \in X$ we can consider the C^* algebra $C^*(x)$ (with identity) generated by x (i.e., the smallest closed $*$-subalgebra containing x). If x is normal we explicitly have

$$C^*(x) = \overline{\{p(x, x^*) | p : \mathbb{C}^2 \to \mathbb{C} \text{ polynomial}\}}, \qquad xx^* = x^*x, \tag{6.30}$$

and in particular $C^*(x)$ is commutative (Problem 6.6). In the self-adjoint case this simplifies to

$$C^*(x) = \overline{\{p(x) | p : \mathbb{C} \to \mathbb{C} \text{ polynomial}\}}, \qquad x = x^*. \tag{6.31}$$

Moreover, in this case $C^*(x)$ is isomorphic to $C(\sigma(x))$ (the continuous functions on the spectrum).

Theorem 6.9 (Spectral theorem)**.** *If X is a C^* algebra and x is self-adjoint, then there is an isometric isomorphism $\Phi : C(\sigma(x)) \to C^*(x)$ such that $f(t) = t$ maps to $\Phi(t) = x$ and $f(t) = 1$ maps to $\Phi(1) = e$.*

Moreover, for every $f \in C(\sigma(x))$ we have

$$\sigma(f(x)) = f(\sigma(x)), \tag{6.32}$$

where $f(x) = \Phi(f(t))$.

Proof. First of all, Φ is well-defined for polynomials. Moreover, by spectral mapping we have

$$\|p(x)\| = r(p(x)) = \sup_{\alpha \in \sigma(p(x))} |\alpha| = \sup_{\alpha \in \sigma(x)} |p(\alpha)| = \|p\|_\infty$$

for every polynomial p. Hence Φ is isometric. Since the polynomials are dense by the Stone–Weierstraß theorem (see the next section) Φ uniquely extends to a map on all of $C(\sigma(x))$ by Theorem 1.29. By continuity of the norm this extension is again isometric. Similarly we have $\Phi(f\,g) = \Phi(f)\Phi(g)$ and $\Phi(f)^* = \Phi(f^*)$ since both relations hold for polynomials.

To show $\sigma(f(x)) = f(\sigma(x))$ fix some $\alpha \in \mathbb{C}$. If $\alpha \notin f(\sigma(x))$, then $g(t) = \frac{1}{f(t)-\alpha} \in C(\sigma(x))$ and $\Phi(g) = (f(x)-\alpha)^{-1} \in X$ shows $\alpha \notin \sigma(f(x))$. Conversely, if $\alpha \notin \sigma(f(x))$ then $g = \Phi^{-1}((f(x)-\alpha)^{-1}) = \frac{1}{f-\alpha}$ is continuous, which shows $\alpha \notin f(\sigma(x))$. □

In particular this last theorem tells us that we have a functional calculus for self-adjoint operators, that is, if $A \in \mathfrak{L}(\mathfrak{H})$ is self-adjoint, then $f(A)$ is well defined for every $f \in C(\sigma(A))$. If f is given by a power series, $f(A)$ defined via Φ coincides with $f(A)$ defined via its power series.

Problem 6.5. *Let X be a C^* algebra and Y a $*$-subalgebra. Show that if Y is commutative, then so is $\overline{Y}$.*

Problem 6.6. *Show that the map Φ from the spectral theorem is positivity preserving, that is, $f \geq 0$ if and only if $\Phi(f)$ is positive.*

Problem 6.7. *Show that $\sigma(x) \subset \{t \in \mathbb{R} | t \geq 0\}$ if and only if x is positive.*

Problem 6.8. *Let $A \in \mathfrak{L}(\mathfrak{H})$. Show that A is normal if and only if*

$$\|Au\| = \|A^* u\|, \quad \forall u \in \mathfrak{H}. \tag{6.33}$$

(Hint: Problem 1.20.)

6.3. Spectral measures

Note: This section requires familiarity with measure theory.

Using the Riesz representation theorem we get another formulation in terms of spectral measures:

Theorem 6.10. *Let $\mathfrak{H}$ be a Hilbert space, and let $A \in \mathfrak{L}(\mathfrak{H})$ be self-adjoint. For every $u, v \in \mathfrak{H}$ there is a corresponding complex Borel measure $\mu_{u,v}$ (the* **spectral measure***) such that*

$$\langle u, f(A)v\rangle = \int_{\sigma(A)} f(t) d\mu_{u,v}(t), \qquad f \in C(\sigma(A)). \tag{6.34}$$

We have

$$\mu_{u,v_1+v_2} = \mu_{u,v_1} + \mu_{u,v_2}, \quad \mu_{u,\alpha v} = \alpha\mu_{u,v}, \quad \mu_{v,u} = \mu_{u,v}^* \tag{6.35}$$

and $|\mu_{u,v}|(\sigma(A)) \leq \|u\|\|v\|$. Furthermore, $\mu_u = \mu_{u,u}$ is a positive Borel measure with $\mu_u(\sigma(A)) = \|u\|^2$.

Proof. Consider the continuous functions on $I = [-\|A\|, \|A\|]$ and note that every $f \in C(I)$ gives rise to some $f \in C(\sigma(A))$ by restricting its domain. Clearly $\ell_{u,v}(f) = \langle u, f(A)v\rangle$ is a bounded linear functional and the existence of a corresponding measure $\mu_{u,v}$ with $|\mu_{u,v}|(I) = \|\ell_{u,v}\| \le \|u\|\|v\|$ follows from Theorem 10.5. Since $\ell_{u,v}(f)$ depends only on the value of f on $\sigma(A) \subseteq I$, $\mu_{u,v}$ is supported on $\sigma(A)$.

Moreover, if $f \ge 0$ we have $\ell_u(f) = \langle u, f(A)u\rangle = \langle f(A)^{1/2}u, f(A)^{1/2}u\rangle = \|f(A)^{1/2}u\|^2 \ge 0$ and hence ℓ_u is positive and the corresponding measure μ_u is positive. The rest follows from the properties of the scalar product. □

It is often convenient to regard $\mu_{u,v}$ as a complex measure on $\mathbb{R}$ by using $\mu_{u,v}(\Omega) = \mu_{u,v}(\Omega \cap \sigma(A))$. If we do this, we can also consider f as a function on $\mathbb{R}$. However, note that $f(A)$ depends only on the values of f on $\sigma(A)$!

Note that the last theorem can be used to define $f(A)$ for every bounded measurable function $f \in B(\sigma(A))$ via Lemma 2.11 and extend the functional calculus from continuous to measurable functions:

Theorem 6.11 (Spectral theorem)**.** *If $\mathfrak{H}$ is a Hilbert space and $A \in \mathfrak{L}(\mathfrak{H})$ is self-adjoint, then there is an homomorphism $\Phi : B(\sigma(A)) \to \mathfrak{L}(\mathfrak{H})$ given by*

$$\langle u, f(A)v\rangle = \int_{\sigma(A)} f(t) d\mu_{u,v}(t), \qquad f \in B(\sigma(A)). \tag{6.36}$$

Moreover, if $f_n(t) \to f(t)$ pointwise and $\sup_n \|f_n\|_\infty$ is bounded, then $f_n(A)u \to f(A)u$ for every $u \in \mathfrak{H}$.

Proof. The map Φ is well-defined linear operator by Lemma 2.11 since we have

$$\Big| \int_{\sigma(A)} f(t) d\mu_{u,v}(t) \Big| \le \|f\|_\infty |\mu_{u,v}|(\sigma(A)) \le \|f\|_\infty \|u\|\|v\|$$

and (6.35). Next, observe that $\Phi(f)^* = \Phi(f^*)$ and $\Phi(fg) = \Phi(f)\Phi(g)$ holds at least for continuous functions. To obtain it for arbitrary bounded functions, choose a (bounded) sequence f_n converging to f in $L^2(\sigma(A), d\mu_u)$ and observe

$$\|(f_n(A) - f(A))u\|^2 = \int |f_n(t) - f(t)|^2 d\mu_u(t)$$

(use $\|h(A)u\|^2 = \langle h(A)u, h(A)u\rangle = \langle u, h(A)^* h(A)u\rangle$). Thus $f_n(A)u \to f(A)u$ and for bounded g we also have that $(gf_n)(A)u \to (gf)(A)u$ and $g(A)f_n(A)u \to g(A)f(A)u$. This establishes the case where f is bounded and g is continuous. Similarly, approximating g removes the continuity requirement from g.

The last claim follows since $f_n \to f$ in L^2 by dominated convergence in this case. □

In particular, given a self-adjoint operator A we can define the **spectral projections**

$$P_A(\Omega) = \chi_\Omega(A), \qquad \Omega \in \mathfrak{B}(\mathbb{R}). \tag{6.37}$$

They are **orthogonal projections**, that is $P^2 = P$ and $P^* = P$.

Lemma 6.12. *Suppose P is an orthogonal projection. Then $\mathfrak{H}$ decomposes in an orthogonal sum*

$$\mathfrak{H} = \operatorname{Ker}(P) \oplus \operatorname{Ran}(P) \tag{6.38}$$

and $\operatorname{Ker}(P) = (\mathbb{I} - P)\mathfrak{H}$, $\operatorname{Ran}(P) = P\mathfrak{H}$.

Proof. Clearly, every $u \in \mathfrak{H}$ can be written as $u = (\mathbb{I} - P)u + Pu$ and

$$\langle(\mathbb{I} - P)u, Pu\rangle = \langle P(\mathbb{I} - P)u, u\rangle = \langle(P - P^2)u, u\rangle = 0$$

shows $\mathfrak{H} = (\mathbb{I}-P)\mathfrak{H} \oplus P\mathfrak{H}$. Moreover, $P(\mathbb{I}-P)u = 0$ shows $(\mathbb{I}-P)\mathfrak{H} \subseteq \operatorname{Ker}(P)$ and if $u \in \operatorname{Ker}(P)$ then $u = (\mathbb{I}-P)u \in (\mathbb{I}-P)\mathfrak{H}$ shows $\operatorname{Ker}(P) \subseteq (\mathbb{I}-P)\mathfrak{H}$. □

In addition, the spectral projections satisfy

$$P_A(\mathbb{R}) = \mathbb{I}, \quad P_A(\bigcup_{n=1}^\infty \Omega_n)u = \sum_{n=1}^\infty P_A(\Omega_n)u, \quad \Omega_n \cap \Omega_m = \emptyset,\ m \neq n. \tag{6.39}$$

for every $u \in \mathfrak{H}$. Such a family of projections is called a **projection valued measure** and

$$P_A(t) = P_A((-\infty, t]) \tag{6.40}$$

is called a **resolution of the identity**. Note that we have

$$\mu_{u,v}(\Omega) = \langle u, P_A(\Omega)v\rangle. \tag{6.41}$$

Using them we can define an operator-valued integral as usual such that

$$A = \int t\, dP_A(t). \tag{6.42}$$

In particular, if $P_A(\{\alpha\}) \neq 0$, then α is an eigenvalue and $\operatorname{Ran}(P_A(\{\alpha\}))$ is the corresponding eigenspace since

$$AP_A(\{\alpha\}) = \alpha P_A(\{\alpha\}). \tag{6.43}$$

The fact that eigenspaces to different eigenvalues are orthogonal now generalizes to

Lemma 6.13. *Suppose $\Omega_1 \cap \Omega_2 = \emptyset$. Then*

$$\operatorname{Ran}(P_A(\Omega_1)) \perp \operatorname{Ran}(P_A(\Omega_2)). \tag{6.44}$$

Proof. Clearly $\chi_{\Omega_1}\chi_{\Omega_2} = \chi_{\Omega_1\cap\Omega_2}$ and hence

$$P_A(\Omega_1)P_A(\Omega_2) = P_A(\Omega_1 \cap \Omega_2).$$

Now if $\Omega_1 \cap \Omega_2 = \emptyset$, then

$$\langle P_A(\Omega_1)u, P_A(\Omega_2)v\rangle = \langle u, P_A(\Omega_1)P_A(\Omega_2)v\rangle = \langle u, P_A(\emptyset)v\rangle = 0,$$

which shows that the ranges are orthogonal to each other. □

Example. Let $A \in \mathfrak{L}(\mathbb{C}^n)$ be some symmetric matrix and let $\alpha_1, \dots, \alpha_m$ be its (distinct) eigenvalues. Then

$$A = \sum_{j=1}^{m} \alpha_j P_A(\{\alpha_j\}), \tag{6.45}$$

where $P_A(\{\alpha_j\})$ is the projection onto the eigenspace corresponding to the eigenvalue α_j. ⋄

Problem 6.9. *Show the following estimates for the total variation of spectral measures*

$$|\mu_{u,v}|(\Omega) \le \mu_u(\Omega)^{1/2}\mu_v(\Omega)^{1/2} \tag{6.46}$$

(Hint: Use $\mu_{u,v}(\Omega) = \langle u, P_A(\Omega)v\rangle = \langle P_A(\Omega)u, P_A(\Omega)v\rangle$ together with Cauchy–Schwarz to estimate (9.14).)

6.4. The Stone–Weierstraß theorem

In the last section we have seen that the C^* algebra of continuous functions $C(K)$ over some compact set $K \subseteq \mathbb{C}$ plays a crucial role and that it is important to be able to identify dense sets. We will be slightly more general and assume that K is some compact metric space. Then it is straightforward to check that the same proof as in the case $K = [a, b]$ (Section 1.2) shows that $C(K, \mathbb{R})$ and $C(K) = C(K, \mathbb{C})$ are Banach spaces when equipped with the maximum norm $\|f\|_\infty = \max_{x\in K} |f(x)|$.

Theorem 6.14 (Stone–Weierstraß, real version)**.** *Suppose K is a compact metric space and let $C(K, \mathbb{R})$ be the Banach algebra of continuous functions (with the maximum norm).*

If $F \subset C(K, \mathbb{R})$ contains the identity 1 and separates points (i.e., for every $x_1 \ne x_2$ there is some function $f \in F$ such that $f(x_1) \ne f(x_2)$), then the algebra generated by F is dense.

Proof. Denote by A the algebra generated by F. Note that if $f \in \overline{A}$, we have $|f| \in \overline{A}$: By the Weierstraß approximation theorem (Theorem 1.20) there is a polynomial $p_n(t)$ such that $\big||t| - p_n(t)\big| < \frac{1}{n}$ for $t \in f(K)$ and hence $p_n(f) \to |f|$.

In particular, if f, g are in $\overline{A}$, we also have

$$\max\{f,g\} = \frac{(f+g)+|f-g|}{2}, \quad \min\{f,g\} = \frac{(f+g)-|f-g|}{2}$$

in $\overline{A}$.

Now fix $f \in C(K,\mathbb{R})$. We need to find some $f^\varepsilon \in \overline{A}$ with $\|f - f^\varepsilon\|_\infty < \varepsilon$.

First of all, since A separates points, observe that for given $y, z \in K$ there is a function $f_{y,z} \in A$ such that $f_{y,z}(y) = f(y)$ and $f_{y,z}(z) = f(z)$ (show this). Next, for every $y \in K$ there is a neighborhood $U(y)$ such that

$$f_{y,z}(x) > f(x) - \varepsilon, \quad x \in U(y),$$

and since K is compact, finitely many, say $U(y_1), \dots, U(y_j)$, cover K. Then

$$f_z = \max\{f_{y_1,z}, \dots, f_{y_j,z}\} \in \overline{A}$$

and satisfies $f_z > f - \varepsilon$ by construction. Since $f_z(z) = f(z)$ for every $z \in K$, there is a neighborhood $V(z)$ such that

$$f_z(x) < f(x) + \varepsilon, \quad x \in V(z),$$

and a corresponding finite cover $V(z_1), \dots, V(z_k)$. Now

$$f^\varepsilon = \min\{f_{z_1}, \dots, f_{z_k}\} \in \overline{A}$$

satisfies $f^\varepsilon < f + \varepsilon$. Since $f - \varepsilon < f_{z_l}$ we also have $f - \varepsilon < f^\varepsilon$ and we have found a required function. □

Theorem 6.15 (Stone–Weierstraß). *Suppose K is a compact metric space and let $C(K)$ be the C^* algebra of continuous functions (with the maximum norm).*

If $F \subset C(K)$ contains the identity 1 and separates points, then the $*$-subalgebra generated by F is dense.

Proof. Just observe that $\tilde{F} = \{\mathrm{Re}(f), \mathrm{Im}(f) | f \in F\}$ satisfies the assumption of the real version. Hence every real-valued continuous functions can be approximated by elements from the subalgebra generated by $\tilde{F}$, in particular this holds for the real and imaginary parts for every given complex-valued function. Finally, note that the subalgebra spanned by $\tilde{F}$ contains the $*$-subalgebra spanned by F. □

Note that the additional requirement of being closed under complex conjugation is crucial: The functions holomorphic on the unit ball and continuous on the boundary separate points, but they are not dense (since the uniform limit of holomorphic functions is again holomorphic).

Corollary 6.16. *Suppose K is a compact metric space and let $C(K)$ be the C^* algebra of continuous functions (with the maximum norm).*

If $F \subset C(K)$ separates points, then the closure of the $*$-subalgebra generated by F is either $C(K)$ or $\{f \in C(K) | f(t_0) = 0\}$ for some $t_0 \in K$.

Proof. There are two possibilities: either all $f \in F$ vanish at one point $t_0 \in K$ (there can be at most one such point since F separates points) or there is no such point. If there is no such point, we can proceed as in the proof of the Stone–Weierstraß theorem to show that the identity can be approximated by elements in $\overline{A}$ (note that to show $|f| \in \overline{A}$ if $f \in \overline{A}$, we do not need the identity, since p_n can be chosen to contain no constant term). If there is such a t_0, the identity is clearly missing from $\overline{A}$. However, adding the identity to $\overline{A}$, we get $\overline{A} + \mathbb{C} = C(K)$ and it is easy to see that $\overline{A} = \{f \in C(K) | f(t_0) = 0\}$. □

Problem 6.10. *Show that the functions $\varphi_n(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\mathrm{i}nx}$, $n \in \mathbb{Z}$, form an orthonormal basis for $\mathfrak{H} = L^2(0, 2\pi)$.*

Problem 6.11. *Let $k \in \mathbb{N}$ and $I \subseteq \mathbb{R}$. Show that the $*$-subalgebra generated by $f_{z_0}(t) = \frac{1}{(t-z_0)^k}$ for one $z_0 \in \mathbb{C}$ is dense in the C^* algebra $C_\infty(I)$ of continuous functions vanishing at infinity*

- *for $I = \mathbb{R}$ if $z_0 \in \mathbb{C} \backslash \mathbb{R}$ and $k = 1, 2$,*
- *for $I = [a, \infty)$ if $z_0 \in (-\infty, a)$ and every k,*
- *for $I = (-\infty, a] \cup [b, \infty)$ if $z_0 \in (a, b)$ and k odd.*

(Hint: Add ∞ to $\mathbb{R}$ to make it compact.)

Problem 6.12. *Let $K \subseteq \mathbb{C}$ be a compact set. Show that the set of all functions $f(z) = p(x, y)$, where $p : \mathbb{R}^2 \to \mathbb{C}$ is polynomial and $z = x + \mathrm{i}y$, is dense in $C(K)$.*

Part 2

Real Analysis

Chapter 7

Almost everything about Lebesgue integration

7.1. Borel measures in a nut shell

The first step in defining the Lebesgue integral is extending the notion of size from intervals to arbitrary sets. Unfortunately, this turns out to be too much, since a classical paradox by Banach and Tarski shows that one can break the unit ball in $\mathbb{R}^3$ into a finite number of (wild – choosing the pieces uses the Axiom of Choice and cannot be done with a jigsaw;-) pieces, rotate and translate them, and reassemble them to get two copies of the unit ball (compare Problem 7.4). Hence any reasonable notion of size (i.e., one which is translation and rotation invariant) cannot be defined for all sets!

A collection of subsets $\mathcal{A}$ of a given set X such that

- $X \in \mathcal{A}$,
- $\mathcal{A}$ is closed under finite unions,
- $\mathcal{A}$ is closed under complements

is called an **algebra**. Note that $\emptyset \in \mathcal{A}$ and that $\mathcal{A}$ is also closed under finite intersections and relative complements: $\emptyset = X'$, $A \cap B = (A' \cup B')'$ (de Morgan), and $A \backslash B = A \cap B'$, where $A' = X \backslash A$ denotes the complement. If an algebra is closed under countable unions (and hence also countable intersections), it is called a **σ-algebra**.

Example. Let $X = \{1, 2, 3\}$, then $\mathcal{A} = \{\emptyset, \{1\}, \{2, 3\}, X\}$ is an algebra. ◇

Moreover, the intersection of any family of (σ-)algebras $\{\mathcal{A}_\alpha\}$ is again a (σ-)algebra and for any collection S of subsets there is a unique smallest (σ-)algebra $\Sigma(S)$ containing S (namely the intersection of all (σ-)algebras containing S). It is called the (σ-)algebra generated by S.

Example. For a given set X the power set $\mathfrak{P}(X)$ is clearly the largest σ-algebra and $\{\emptyset, X\}$ is the smallest. ⋄

Example. Let X be some set with a σ-algebra Σ. Then every subset $Y \subseteq X$ has a natural σ-algebra $\Sigma \cap Y = \{A \cap Y | A \in \Sigma\}$ (show that this is indeed a σ-algebra) known as the **relative σ-algebra**.

Note that if S generates Σ, then $S \cap Y$ generates $\Sigma \cap Y$: $\Sigma(S) \cap Y = \Sigma(S \cap Y)$. Indeed, since $\Sigma \cap Y$ is a σ-algebra containing $S \cap Y$, we have $\Sigma(S \cap Y) \subseteq \Sigma(S) \cap Y = \Sigma \cap Y$. Conversely, consider $\{A \in \Sigma | A \cap Y \in \Sigma(S \cap Y)\}$ which is a σ-algebra (check this). Since this last σ-algebra contains S it must be equal to $\Sigma = \Sigma(S)$ and thus $\Sigma \cap Y \subseteq \Sigma(S \cap Y)$. ⋄

If X is a topological space, the **Borel σ-algebra** $\mathfrak{B}(X)$ of X is defined to be the σ-algebra generated by all open (respectively, all closed) sets. In fact, if X is second countable, any countable base will suffice to generate the Borel σ-algebra (recall Lemma 1.1).

Sets in the Borel σ-algebra are called **Borel sets**.

Example. In the case $X = \mathbb{R}^n$ the Borel σ-algebra will be denoted by $\mathfrak{B}^n$ and we will abbreviate $\mathfrak{B} = \mathfrak{B}^1$. Note that in order to generate $\mathfrak{B}$ open (or closed) intervals with rational boundary points suffice. ⋄

Example. If X is a topological space then any Borel set $Y \subseteq X$ is also a topological space equipped with the relative topology and its Borel σ-algebra is given by $\mathfrak{B}(Y) = \mathfrak{B}(X) \cap Y = \{A | A \in \mathfrak{B}(X), A \subseteq Y\}$ (show this). ⋄

Now let us turn to the definition of a measure: A set X together with a σ-algebra Σ is called a **measurable space**. A **measure** μ is a map $\mu : \Sigma \to [0, \infty]$ on a σ-algebra Σ such that

- $\mu(\emptyset) = 0$,
- $\mu(\bigcup_{j=1}^\infty A_j) = \sum_{j=1}^\infty \mu(A_j)$ if $A_j \cap A_k = \emptyset$ for all $j \neq k$ (σ-additivity).

Here the sum is set equal to ∞ if one of the summands is ∞ or if it diverges.

The measure μ is called **σ-finite** if there is a countable cover $\{X_j\}_{j=1}^\infty$ of X such that $X_j \in \Sigma$ and $\mu(X_j) < \infty$ for all j. (Note that it is no restriction to assume $X_j \subseteq X_{j+1}$.) It is called **finite** if $\mu(X) < \infty$. The sets in Σ are called **measurable sets** and the triple (X, Σ, μ) is referred to as a **measure space**.

Example. Take a set X and $\Sigma = \mathfrak{P}(X)$ and set $\mu(A)$ to be the number of elements of A (respectively, ∞ if A is infinite). This is the so-called **counting measure**. It will be finite if and only if X is finite and σ-finite if and only if X is countable. $\diamond$

Example. Take a set X and $\Sigma = \mathfrak{P}(X)$. Fix a point $x \in X$ and set $\mu(A) = 1$ if $x \in A$ and $\mu(A) = 0$ else. This is the **Dirac measure** centered at x. $\diamond$

Example. Let μ_1, μ_2 be two measures on (X, Σ) and $\alpha_1, \alpha_2 \geq 0$. Then $\mu = \alpha_1\mu_1 + \alpha_2\mu_2$ defined via

$$\mu(A) = \alpha_1\mu_1(A) + \alpha_2\mu_2(A)$$

is again a measure. Furthermore, given a countable number of measures μ_n and numbers $\alpha_n \geq 0$, then $\mu = \sum_n \alpha_n\mu_n$ is again a measure (show this). $\diamond$

Example. Let μ be a measure on (X, Σ) and $Y \subseteq X$ a measurable subset. Then

$$\nu(A) = \mu(A \cap Y)$$

is again a measure on (X, Σ) (show this). $\diamond$

If $Y \in \Sigma$ we can restrict the σ-algebra $\Sigma|_Y = \{A \in \Sigma | A \subseteq Y\}$ such that $(Y, \Sigma|_Y, \mu|_Y)$ is again a measurable space. It will be σ-finite if (X, Σ, μ) is.

If we replace the σ-algebra by an algebra $\mathcal{A}$, then μ is called a **premeasure**. In this case σ-additivity clearly only needs to hold for disjoint sets A_n for which $\bigcup_n A_n \in \mathcal{A}$.

We will write $A_n \nearrow A$ if $A_n \subseteq A_{n+1}$ (note $A = \bigcup_n A_n$) and $A_n \searrow A$ if $A_{n+1} \subseteq A_n$ (note $A = \bigcap_n A_n$).

Theorem 7.1. *Any measure μ satisfies the following properties:*

(i) *$A \subseteq B$ implies $\mu(A) \leq \mu(B)$ (monotonicity).*

(ii) *$\mu(A_n) \to \mu(A)$ if $A_n \nearrow A$ (continuity from below).*

(iii) *$\mu(A_n) \to \mu(A)$ if $A_n \searrow A$ and $\mu(A_1) < \infty$ (continuity from above).*

Proof. The first claim is obvious from $\mu(B) = \mu(A) + \mu(B \backslash A)$. To see the second define $\tilde{A}_1 = A_1$, $\tilde{A}_n = A_n \backslash A_{n-1}$ and note that these sets are disjoint and satisfy $A_n = \bigcup_{j=1}^n \tilde{A}_j$. Hence $\mu(A_n) = \sum_{j=1}^n \mu(\tilde{A}_j) \to \sum_{j=1}^\infty \mu(\tilde{A}_j) = \mu(\bigcup_{j=1}^\infty \tilde{A}_j) = \mu(A)$ by σ-additivity. The third follows from the second using $\tilde{A}_n = A_1 \backslash A_n \nearrow A_1 \backslash A$ implying $\mu(\tilde{A}_n) = \mu(A_1) - \mu(A_n) \to \mu(A_1 \backslash A) = \mu(A_1) - \mu(A)$. $\square$

Example. Consider the counting measure on $X = \mathbb{N}$ and let $A_n = \{j \in \mathbb{N} | j \geq n\}$, then $\mu(A_n) = \infty$, but $\mu(\bigcap_n A_n) = \mu(\emptyset) = 0$ which shows that the requirement $\mu(A_1) < \infty$ in the last claim of Theorem 7.1 is not superfluous. ⋄

A measure on the Borel σ-algebra is called a **Borel measure** if $\mu(K) < \infty$ for every compact set K. Note that some authors do not require this last condition.

Example. Let $X = \mathbb{R}$ and $\Sigma = \mathfrak{B}$. The Dirac measure is a Borel measure. The counting measure is no Borel measure since $\mu([a,b]) = \infty$ for $a < b$. ⋄

A measure on the Borel σ-algebra is called **outer regular** if

$$\mu(A) = \inf_{O \supseteq A, O \text{ open}} \mu(O) \tag{7.1}$$

and **inner regular** if

$$\mu(A) = \sup_{K \subseteq A, K \text{ compact}} \mu(K). \tag{7.2}$$

It is called **regular** if it is both outer and inner regular.

Example. Let $X = \mathbb{R}$ and $\Sigma = \mathfrak{B}$. The counting measure is inner regular but not outer regular (every nonempty open set has infinite measure). The Dirac measure is a regular Borel measure. ⋄

But how can we obtain some more interesting Borel measures? We will restrict ourselves to the case of $X = \mathbb{R}$ for simplicity, in which case Borel measures are also known as **Lebesgue–Stieltjes measures**. Then the strategy is as follows: Start with the algebra of finite unions of disjoint intervals and define μ for those sets (as the sum over the intervals). This yields a premeasure. Extend this to an *outer measure* for all subsets of $\mathbb{R}$. Show that the restriction to the Borel sets is a measure.

Let us first show how we should define μ for intervals: To every Borel measure on $\mathfrak{B}$ we can assign its **distribution function**

$$\mu(x) = \begin{cases} -\mu((x,0]), & x < 0, \\ 0, & x = 0, \\ \mu((0,x]), & x > 0, \end{cases} \tag{7.3}$$

which is right continuous and nondecreasing.

Conversely, to obtain a measure from a nondecreasing function $m : \mathbb{R} \to \mathbb{R}$ we proceed as follows: Recall that an interval is a subset of the real line of the form

$$I = (a,b], \quad I = [a,b], \quad I = (a,b), \quad \text{or} \quad I = [a,b), \tag{7.4}$$

with $a \le b$, $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$. Note that (a,a), $[a,a)$, and $(a,a]$ denote the empty set, whereas $[a,a]$ denotes the singleton $\{a\}$. For any proper interval with different endpoints we can define its measure to be

$$\mu(I) = \begin{cases} m(b+) - m(a+), & I = (a,b], \\ m(b+) - m(a-), & I = [a,b], \\ m(b-) - m(a+), & I = (a,b), \\ m(b-) - m(a-), & I = [a,b), \end{cases} \tag{7.5}$$

where $m(a\pm) = \lim_{\varepsilon\downarrow 0} m(a \pm \varepsilon)$ (which exist by monotonicity). If one of the endpoints is infinite we agree to use $m(\pm\infty) = \lim_{x\to\pm\infty} m(x)$. For the empty set we of course set $\mu(\emptyset) = 0$ and for the singletons we set

$$\mu(\{a\}) = m(a+) - m(a-) \tag{7.6}$$

(which agrees with (7.5) except for the case $I = (a, a)$ which would give a negative value for the empty set if μ jumps at a). Note that $\mu(\{a\}) = 0$ if and only if $m(x)$ is continuous at a and that there can be only countably many points with $\mu(\{a\}) > 0$ since a nondecreasing function can have at most countably many jumps. Moreover, observe that the definition of μ does not involve the actual value of m at a jump. Hence any function $\tilde{m}$ with $m(x-) \leq \tilde{m}(x) \leq m(x+)$ gives rise to the same μ. We will frequently assume that m is right continuous such that it coincides with the distribution function up to a constant, $\mu(x) = m(x+) - m(0+)$. In particular, μ determines m up to a constant and the value at the jumps.

Now we can consider the algebra of finite unions of intervals (check that this is indeed an algebra) and extend (7.5) to finite unions of disjoint intervals by summing over all intervals. It is straightforward to verify that μ is well-defined (one set can be represented by different unions of intervals) and by construction additive. In fact, it is even a premeasure.

Lemma 7.2. *The interval function μ defined in (7.5) gives rise to a unique σ-finite premeasure on the algebra $\mathcal{A}$ of finite unions of disjoint intervals.*

Proof. It remains to verify σ-additivity. We need to show that for any disjoint union

$$\mu(\bigcup_k I_k) = \sum_k \mu(I_k)$$

whenever $I_k \in \mathcal{A}$ and $I = \bigcup_k I_k \in \mathcal{A}$. Since each I_k is a finite union of intervals, we can as well assume each I_k is just one interval (just split I_k into its subintervals and note that the sum does not change by additivity). Similarly, we can assume that I is just one interval (just treat each subinterval separately).

By additivity μ is monotone and hence

$$\sum_{k=1}^{n} \mu(I_k) = \mu(\bigcup_{k=1}^{n} I_k) \leq \mu(I)$$

which shows
$$\sum_{k=1}^{\infty} \mu(I_k) \le \mu(I).$$
To get the converse inequality, we need to work harder:

We can cover each I_k by some slightly larger open interval J_k such that $\mu(J_k) \le \mu(I_k) + \frac{\varepsilon}{2^k}$. First suppose I is compact. Then finitely many of the J_k, say the first n, cover I and we have
$$\mu(I) \le \mu(\bigcup_{k=1}^{n} J_k) \le \sum_{k=1}^{n} \mu(J_k) \le \sum_{k=1}^{\infty} \mu(I_k) + \varepsilon.$$
Since $\varepsilon > 0$ is arbitrary, this shows σ-additivity for compact intervals. By additivity we can always add/subtract the endpoints of I and hence σ-additivity holds for any bounded interval. If I is unbounded we can again assume that it is closed by adding an endpoint if necessary. Then for any $x > 0$ we can find an n such that $\{J_k\}_{k=1}^{n}$ cover at least $I \cap [-x, x]$ and hence
$$\sum_{k=1}^{\infty} \mu(I_k) \ge \sum_{k=1}^{n} \mu(I_k) \ge \sum_{k=1}^{n} \mu(J_k) - \varepsilon \ge \mu([-x, x] \cap I) - \varepsilon.$$
Since $x > 0$ and $\varepsilon > 0$ are arbitrary, we are done. □

In particular, this is a premeasure on the algebra of finite unions of intervals which can be extended to a measure:

Theorem 7.3. *For every nondecreasing function $m : \mathbb{R} \to \mathbb{R}$ there exists a unique Borel measure μ which extends (7.5). Two different functions generate the same measure if and only if the difference is a constant away from the discontinuities.*

Since the proof of this theorem is rather involved, we defer it to the next section and look at some examples first.

Example. Suppose $\Theta(x) = 0$ for $x < 0$ and $\Theta(x) = 1$ for $x \ge 0$. Then we obtain the so-called **Dirac measure** at 0, which is given by $\Theta(A) = 1$ if $0 \in A$ and $\Theta(A) = 0$ if $0 \notin A$. ⋄

Example. Suppose $\lambda(x) = x$. Then the associated measure is the ordinary **Lebesgue measure** on $\mathbb{R}$. We will abbreviate the Lebesgue measure of a Borel set A by $\lambda(A) = |A|$. ⋄

A set $A \in \Sigma$ is called a **support** for μ if $\mu(X \backslash A) = 0$. Since a support is not unique (see the examples below) one also defines **the support** of μ via
$$\text{supp}(\mu) = \{x \in X | \mu(O) > 0 \text{ for every open neighborhood of } x\}. \qquad (7.7)$$

Equivalently one obtains $\operatorname{supp}(\mu)$ by removing all points which have an open neighborhood of measure zero. In particular, this shows that $\operatorname{supp}(\mu)$ is closed. If X is second countable, then $\operatorname{supp}(\mu)$ is indeed a support for μ: For every point $x \notin \operatorname{supp}(\mu)$ let O_x be an open neighborhood of measure zero. These sets cover $X \setminus \operatorname{supp}(\mu)$ and by the Lindelöf theorem there is a countable subcover, which shows that $X \setminus \operatorname{supp}(\mu)$ has measure zero.

Example. Let $X = \mathbb{R}$, $\Sigma = \mathfrak{B}$. The support of the Lebesgue measure λ is all of $\mathbb{R}$. However, every single point has Lebesgue measure zero and so has every countable union of points (by σ-additivity). Hence any set whose complement is countable is a support. There are even uncountable sets of Lebesgue measure zero (see the Cantor set below) and hence a support might even lack an uncountable number of points.

The support of the Dirac measure centered at 0 is the single point 0. Any set containing 0 is a support of the Dirac measure.

In general, the support of a Borel measure on $\mathbb{R}$ is given by

$$\operatorname{supp}(d\mu) = \{x \in \mathbb{R} | \mu(x - \varepsilon) < \mu(x + \varepsilon), \ \forall \varepsilon > 0\}.$$

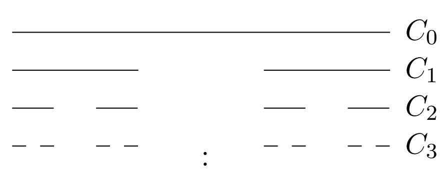
Here we have used $d\mu$ to emphasize that we are interested in the support of the measure $d\mu$ which is different from the support of its distribution function $\mu(x)$. ⋄

A property is said to hold **μ-almost everywhere** (a.e.) if it holds on a support for μ or, equivalently, if the set where it does not hold is contained in a set of measure zero.

Example. The set of rational numbers is countable and hence has Lebesgue measure zero, $\lambda(\mathbb{Q}) = 0$. So, for example, the characteristic function of the rationals $\mathbb{Q}$ is zero almost everywhere with respect to Lebesgue measure.

Any function which vanishes at 0 is zero almost everywhere with respect to the Dirac measure centered at 0. ⋄

Example. The **Cantor set** is an example of a closed uncountable set of Lebesgue measure zero. It is constructed as follows: Start with $C_0 = [0, 1]$ and remove the middle third to obtain $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Next, again remove the middle third's of the remaining sets to obtain $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$:



Proceeding like this, we obtain a sequence of nesting sets C_n and the limit $C = \bigcap_n C_n$ is the Cantor set. Since C_n is compact, so is C. Moreover,

C_n consists of 2^n intervals of length 3^{-n}, and thus its Lebesgue measure is $\lambda(C_n) = (2/3)^n$. In particular, $\lambda(C) = \lim_{n\to\infty} \lambda(C_n) = 0$. Using the ternary expansion, it is extremely simple to describe: C is the set of all $x \in [0,1]$ whose ternary expansion contains no one's, which shows that C is uncountable (why?). It has some further interesting properties: it is totally disconnected (i.e., it contains no subintervals) and perfect (it has no isolated points). $\diamond$

Problem 7.1. *Find all algebras over* $X = \{1, 2, 3\}$.

Problem 7.2. *Take some set* X. *Show that* $\mathcal{A} = \{A \subseteq X | A \text{ or } X\backslash A \text{ is finite}\}$ *is an algebra. Show that* $\Sigma = \{A \subseteq X | A \text{ or } X\backslash A \text{ is countable}\}$ *is a* σ-*algebra. (Hint: To verify closedness under unions consider the cases were all sets are finite and where one set has finite complement.)*

Problem 7.3. *Show that if* X *is finite, then every algebra is a* σ-*algebra. Show that this is not true in general if* X *is countable.*

Problem 7.4 (Vitali set)**.** *Call two numbers* $x, y \in [0, 1)$ *equivalent if* $x - y$ *is rational. Construct the set* V *by choosing one representative from each equivalence class. Show that* V *cannot be measurable with respect to any nontrivial finite translation invariant measure on* $[0, 1)$. *(Hint: How can you build up* $[0, 1)$ *from translations of* V*?)*

7.2. Extending a premeasure to a measure

The purpose of this section is to prove Theorem 7.3. It is rather technical and can be skipped on first reading.

In order to prove Theorem 7.3, we need to show how a premeasure can be extended to a measure. To show that the extension is unique we need a better criterion to check when a given system of sets is in fact a σ-algebra. In many situations it is easy to show that a given set is closed under complements and under countable unions of disjoint sets. Hence we call a collection of sets $\mathcal{D}$ with these properties a **Dynkin system** (also **λ-system**) if it also contains X.

Note that a Dynkin system is closed under proper relative complements since $A, B \in \mathcal{D}$ implies $B\backslash A = (B' \cup A)' \in \mathcal{D}$ provided $A \subseteq B$. Moreover, if it is also closed under finite intersections (or arbitrary finite unions) then it is an algebra and hence also a σ-algebra. To see the last claim note that if $A = \bigcup_j A_j$ then also $A = \bigcup_j B_j$ where the sets $B_j = A_j \backslash \bigcup_{k<j} A_k$ are disjoint.

As with σ-algebras the intersection of Dynkin systems is a Dynkin system and every collection of sets S generates a smallest Dynkin system $\mathcal{D}(S)$. The important observation is that if S is closed under finite intersections (in which case it is sometimes called a π-**system**), then so is $\mathcal{D}(S)$ and hence will be a σ-algebra.

Lemma 7.4 (Dynkin's π-λ theorem)**.** *Let S be a collection of subsets of X which is closed under finite intersections (or unions). Then $\mathcal{D}(S) = \Sigma(S)$.*

Proof. It suffices to show that $\mathcal{D} = \mathcal{D}(S)$ is closed under finite intersections. To this end consider the set $D(A) = \{B \in \mathcal{D} | A \cap B \in \mathcal{D}\}$ for $A \in \mathcal{D}$. I claim that $D(A)$ is a Dynkin system.

First of all $X \in D(A)$ since $A \cap X = A \in \mathcal{D}$. Next, if $B \in D(A)$ then $A \cap B' = A \backslash (B \cap A) \in \mathcal{D}$ (since $\mathcal{D}$ is closed under proper relative complements) implying $B' \in D(A)$. Finally if $B = \bigcup_j B_j$ with $B_j \in D(A)$ disjoint, then $A \cap B = \bigcup_j (A \cap B_j) \in \mathcal{D}$ with $B_j \in \mathcal{D}$ disjoint, implying $B \in D(A)$.

Now if $A \in S$ we have $S \subseteq D(A)$ implying $D(A) = \mathcal{D}$. Consequently $A \cap B \in \mathcal{D}$ if at least one of the sets is in S. But this shows $S \subseteq D(A)$ and hence $D(A) = \mathcal{D}$ for every $A \in \mathcal{D}$. So $\mathcal{D}$ is closed under finite intersections and thus a σ-algebra. The case of unions is analogous. $\square$

The typical use of this lemma is as follows: First verify some property for sets in a set S which is closed under finite intersections and generates the σ-algebra. In order to show that it holds for every set in $\Sigma(S)$, it suffices to show that the collection of sets for which it holds is a Dynkin system.

As an application we show that a premeasure determines the corresponding measure μ uniquely (if there is one at all):

Theorem 7.5 (Uniqueness of measures)**.** *Let $S \subseteq \Sigma$ be a collection of sets which generate Σ and which is closed under finite intersections and contains a sequence of increasing sets $X_n \nearrow X$ of finite measure $\mu(X_n) < \infty$. Then μ is uniquely determined by the values on S.*

Proof. Let $\tilde{\mu}$ be a second measure and note $\mu(X) = \lim_{n\to\infty} \mu(X_n) = \lim_{n\to\infty} \tilde{\mu}(X_n) = \tilde{\mu}(X)$. We first suppose $\mu(X) < \infty$.

Then

$$\mathcal{D} = \{A \in \Sigma | \mu(A) = \tilde{\mu}(A)\}$$

is a Dynkin system. In fact, by $\mu(A') = \mu(X) - \mu(A) = \tilde{\mu}(X) - \tilde{\mu}(A) = \tilde{\mu}(A')$ for $A \in \mathcal{D}$ we see that $\mathcal{D}$ is closed under complements. Finally, by continuity of measures from below it is also closed under countable disjoint unions. Since $\mathcal{D}$ contains S by assumption we conclude $\mathcal{D} = \Sigma(S) = \Sigma$ from Lemma 7.4. This finishes the finite case.

To extend our result to the general case observe that the finite case implies $\mu(A \cap X_j) = \tilde{\mu}(A \cap X_j)$ (just restrict μ, $\tilde{\mu}$ to X_j). Hence

$$\mu(A) = \lim_{j\to\infty} \mu(A \cap X_j) = \lim_{j\to\infty} \tilde{\mu}(A \cap X_j) = \tilde{\mu}(A)$$

and we are done. □

Corollary 7.6. *Let μ be a σ-finite premeasure on an algebra $\mathcal{A}$. Then there is at most one extension to $\Sigma(\mathcal{A})$.*

So it remains to ensure that there is an extension at all. For any premeasure μ we define

$$\mu^*(A) = \inf\Big\{ \sum_{n=1}^{\infty} \mu(A_n) \Big| A \subseteq \bigcup_{n=1}^{\infty} A_n,\ A_n \in \mathcal{A} \Big\} \tag{7.8}$$

where the infimum extends over all countable covers from $\mathcal{A}$. Then the function $\mu^* : \mathfrak{P}(X) \to [0,\infty]$ is an **outer measure**; that is, it has the properties (Problem 7.5)

- $\mu^*(\emptyset) = 0$,
- $A_1 \subseteq A_2 \Rightarrow \mu^*(A_1) \le \mu^*(A_2)$, and
- $\mu^*(\bigcup_{n=1}^{\infty} A_n) \le \sum_{n=1}^{\infty} \mu^*(A_n)$ (subadditivity).

Note that $\mu^*(A) = \mu(A)$ for $A \in \mathcal{A}$ (Problem 7.6).

Theorem 7.7 (Extensions via outer measures)**.** *Let μ^* be an outer measure. Then the set Σ of all sets A satisfying the Carathéodory condition*

$$\mu^*(E) = \mu^*(A \cap E) + \mu^*(A' \cap E), \quad \forall E \subseteq X \tag{7.9}$$

(where $A' = X \backslash A$ is the complement of A) forms a σ-algebra and μ^* restricted to this σ-algebra is a measure.

Proof. We first show that Σ is an algebra. It clearly contains X and is closed under complements. Concerning unions let $A, B \in \Sigma$. Applying Carathéodory's condition twice shows

$$\begin{aligned}\mu^*(E) =& \mu^*(A \cap B \cap E) + \mu^*(A' \cap B \cap E) + \mu^*(A \cap B' \cap E)\\ &+ \mu^*(A' \cap B' \cap E)\\ \ge& \mu^*((A \cup B) \cap E) + \mu^*((A \cup B)' \cap E),\end{aligned}$$

where we have used de Morgan and

$$\mu^*(A \cap B \cap E) + \mu^*(A' \cap B \cap E) + \mu^*(A \cap B' \cap E) \ge \mu^*((A \cup B) \cap E)$$

which follows from subadditivity and $(A \cup B) \cap E = (A \cap B \cap E) \cup (A' \cap B \cap E) \cup (A \cap B' \cap E)$. Since the reverse inequality is just subadditivity, we conclude that Σ is an algebra.

Next, let A_n be a sequence of sets from Σ. Without restriction we can assume that they are disjoint (compare the argument for item (ii) in the proof of Theorem 7.1). Abbreviate $\tilde{A}_n = \bigcup_{k\le n} A_k$, $A = \bigcup_n A_n$. Then for every set E we have

$$\begin{aligned}\mu^*(\tilde{A}_n \cap E) &= \mu^*(A_n \cap \tilde{A}_n \cap E) + \mu^*(A_n' \cap \tilde{A}_n \cap E)\\ &= \mu^*(A_n \cap E) + \mu^*(\tilde{A}_{n-1} \cap E)\\ &= \ldots = \sum_{k=1}^{n} \mu^*(A_k \cap E).\end{aligned}$$

Using $\tilde{A}_n \in \Sigma$ and monotonicity of μ^*, we infer

$$\begin{aligned}\mu^*(E) &= \mu^*(\tilde{A}_n \cap E) + \mu^*(\tilde{A}_n' \cap E)\\ &\ge \sum_{k=1}^{n} \mu^*(A_k \cap E) + \mu^*(A' \cap E).\end{aligned}$$

Letting $n \to \infty$ and using subadditivity finally gives

$$\begin{aligned}\mu^*(E) &\ge \sum_{k=1}^{\infty} \mu^*(A_k \cap E) + \mu^*(A' \cap E)\\ &\ge \mu^*(A \cap E) + \mu^*(A' \cap E) \ge \mu^*(E)\end{aligned} \tag{7.10}$$

and we infer that Σ is a σ-algebra.

Finally, setting $E = A$ in (7.10), we have

$$\mu^*(A) = \sum_{k=1}^{\infty} \mu^*(A_k \cap A) + \mu^*(A' \cap A) = \sum_{k=1}^{\infty} \mu^*(A_k)$$

and we are done. □

Remark: The constructed measure μ is **complete**; that is, for every measurable set A of measure zero, every subset of A is again measurable (Problem 7.7).

The only remaining question is whether there are any nontrivial sets satisfying the Carathéodory condition.

Lemma 7.8. *Let μ be a premeasure on $\mathcal{A}$ and let μ^* be the associated outer measure. Then every set in $\mathcal{A}$ satisfies the Carathéodory condition.*

Proof. Let $A_n \in \mathcal{A}$ be a countable cover for E. Then for every $A \in \mathcal{A}$ we have

$$\sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n \cap A) + \sum_{n=1}^{\infty} \mu(A_n \cap A') \ge \mu^*(E \cap A) + \mu^*(E \cap A')$$

since $A_n \cap A \in \mathcal{A}$ is a cover for $E \cap A$ and $A_n \cap A' \in \mathcal{A}$ is a cover for $E \cap A'$. Taking the infimum, we have $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A')$, which finishes the proof. □

Concerning regularity we note:

Lemma 7.9. *Suppose outer regularity (7.1) holds for every set in the algebra, then μ is outer regular.*

Proof. By assumption we can replace each set A_n in (7.8) by a possibly slightly larger open set and hence the infimum in (7.8) can be realized with open sets. □

Thus, as a consequence we obtain Theorem 7.3 except for regularity. Outer regularity is easy to see for a finite union of intervals since we can replace each interval by a possibly slightly larger open interval with only slightly larger measure. Inner regularity will be postponed until Lemma 7.14.

Problem 7.5. *Show that μ^* defined in (7.8) is an outer measure. (Hint for the last property: Take a cover $\{B_{nk}\}_{k=1}^\infty$ for A_n such that $\mu^*(A_n) = \frac{\varepsilon}{2^n} + \sum_{k=1}^\infty \mu(B_{nk})$ and note that $\{B_{nk}\}_{n,k=1}^\infty$ is a cover for $\bigcup_n A_n$.)*

Problem 7.6. *Show that μ^* defined in (7.8) extends μ. (Hint: For the cover A_n it is no restriction to assume $A_n \cap A_m = \emptyset$ and $A_n \subseteq A$.)*

Problem 7.7. *Show that the measure constructed in Theorem 7.7 is complete.*

Problem 7.8. *Let μ be a finite measure. Show that*

$$d(A, B) = \mu(A \Delta B), \qquad A \Delta B = (A \cup B) \backslash (A \cap B) \tag{7.11}$$

is a metric on Σ if we identify sets of measure zero. Show that if $\mathcal{A}$ is an algebra, then it is dense in $\Sigma(\mathcal{A})$. (Hint: Show that the sets which can be approximated by sets in $\mathcal{A}$ form a Dynkin system.)

7.3. Measurable functions

The Riemann integral works by splitting the x coordinate into small intervals and approximating $f(x)$ on each interval by its minimum and maximum. The problem with this approach is that the difference between maximum and minimum will only tend to zero (as the intervals get smaller) if $f(x)$ is sufficiently nice. To avoid this problem, we can force the difference to go to zero by considering, instead of an interval, the set of x for which $f(x)$ lies between two given numbers $a < b$. Now we need the size of the set of these x, that is, the size of the preimage $f^{-1}((a, b))$. For this to work, preimages of intervals must be measurable.

Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces. A function $f : X \to Y$ is called **measurable** if $f^{-1}(A) \in \Sigma_X$ for every $A \in \Sigma_Y$. Clearly it suffices to check this condition for every set A in a collection of sets which generate Σ_Y, since the collection of sets for which it holds forms a σ-algebra by $f^{-1}(Y \backslash A) = X \backslash f^{-1}(A)$ and $f^{-1}(\bigcup_j A_j) = \bigcup_j f^{-1}(A_j)$.

We will be mainly interested in the case where $(Y, \Sigma_Y) = (\mathbb{R}^n, \mathfrak{B}^n)$.

Lemma 7.10. *A function $f : X \to \mathbb{R}^n$ is measurable if and only if*

$$f^{-1}(I) \in \Sigma \qquad \forall\, I = \prod_{j=1}^{n} (a_j, \infty). \tag{7.12}$$

In particular, a function $f : X \to \mathbb{R}^n$ is measurable if and only if every component is measurable and a complex-valued function $f : X \to \mathbb{C}^n$ is measurable if and only if both its real and imaginary parts are.

Proof. We need to show that $\mathfrak{B}^n$ is generated by rectangles of the above form. The σ-algebra generated by these rectangles also contains all open rectangles of the form $I = \prod_{j=1}^{n} (a_j, b_j)$, which form a base for the topology. □

Clearly the intervals (a_j, ∞) can also be replaced by $[a_j, \infty)$, $(-\infty, a_j)$, or $(-\infty, a_j]$.

If X is a topological space and Σ the corresponding Borel σ-algebra, we will also call a measurable function **Borel function**. Note that, in particular,

Lemma 7.11. *Let (X, Σ_X), (Y, Σ_Y), (Z, Σ_Z) be topological spaces with their corresponding Borel σ-algebras. Any continuous function $f : X \to Y$ is measurable. Moreover, if $f : X \to Y$ and $g : Y \to Z$ are measurable functions, then the composition $g \circ f$ is again measurable.*

The set of all measurable functions forms an algebra.

Lemma 7.12. *Let X be a topological space and Σ its Borel σ-algebra. Suppose $f, g : X \to \mathbb{R}$ are measurable functions. Then the sum $f + g$ and the product fg are measurable.*

Proof. Note that addition and multiplication are continuous functions from $\mathbb{R}^2 \to \mathbb{R}$ and hence the claim follows from the previous lemma. □

Sometimes it is also convenient to allow $\pm\infty$ as possible values for f, that is, functions $f : X \to \overline{\mathbb{R}}$, $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. In this case $A \subseteq \overline{\mathbb{R}}$ is called Borel if $A \cap \mathbb{R}$ is. This implies that $f : X \to \overline{\mathbb{R}}$ will be Borel if and only if $f^{-1}(\pm\infty)$ are Borel and $f: X\backslash f^{-1}(\{-\infty,\infty\}) \to \mathbb{R}$ is Borel. Since

$$\{+\infty\} = \bigcap_n (n, +\infty], \qquad \{-\infty\} = \overline{\mathbb{R}}\backslash \bigcup_n (-n, +\infty], \tag{7.13}$$

we see that $f: X \to \overline{\mathbb{R}}$ is measurable if and only if

$$f^{-1}((a,\infty]) \in \Sigma \qquad \forall\, a \in \mathbb{R}. \tag{7.14}$$

Again the intervals $(a,\infty]$ can also be replaced by $[a,\infty]$, $[-\infty,a)$, or $[-\infty,a]$.

Hence it is not hard to check that the previous lemma still holds if one either avoids undefined expressions of the type $\infty-\infty$ and $\pm\infty\cdot 0$ or makes a definite choice, e.g., $\infty-\infty=0$ and $\pm\infty\cdot 0 = 0$.

Moreover, the set of all measurable functions is closed under all important limiting operations.

Lemma 7.13. *Suppose $f_n: X \to \overline{\mathbb{R}}$ is a sequence of measurable functions. Then*

$$\inf_{n\in\mathbb{N}} f_n, \quad \sup_{n\in\mathbb{N}} f_n, \quad \liminf_{n\to\infty} f_n, \quad \limsup_{n\to\infty} f_n \tag{7.15}$$

are measurable as well.

Proof. It suffices to prove that $\sup f_n$ is measurable since the rest follows from $\inf f_n = -\sup(-f_n)$, $\liminf f_n = \sup_k \inf_{n\ge k} f_n$, and $\limsup f_n = \inf_k \sup_{n\ge k} f_n$. But $(\sup f_n)^{-1}((a,\infty]) = \bigcup_n f_n^{-1}((a,\infty])$ are Borel and we are done. □

A few immediate consequences are worthwhile noting: It follows that if f and g are measurable functions, so are $\min(f,g)$, $\max(f,g)$, $|f| = \max(f,-f)$, and $f^\pm = \max(\pm f, 0)$. Furthermore, the pointwise limit of measurable functions is again measurable.

Sometimes the case of arbitrary suprema and infima is also of interest. In this respect the following observation is useful: Recall that a function $f: X \to \overline{\mathbb{R}}$ is **lower semicontinuous** if the set $f^{-1}((a,\infty])$ is open for every $a \in \mathbb{R}$. Then it follows from the definition that the sup over an arbitrary collection of lower semicontinuous functions

$$\overline{f}(x) = \sup_\alpha f_\alpha(x) \tag{7.16}$$

is again lower semicontinuous. Similarly, f is is **upper semicontinuous** if the set $f^{-1}([-\infty,a))$ is open for every $a \in \mathbb{R}$. In this case the infimum

$$\underline{f}(x) = \inf_\alpha f_\alpha(x) \tag{7.17}$$

is again upper semicontinuous.

Problem 7.9. *Show that the supremum over lower semicontinuous functions is again lower semicontinuous.*

7.4. How wild are measurable objects

In this section we want to investigate how far measurable objects are away from well understood ones. As our first task we want to show that measurable sets can be well approximated by using closed sets from the inside and open sets from the outside in nice spaces like $\mathbb{R}^n$.

Lemma 7.14. *Let X be a metric space and μ a Borel measure which is finite on finite balls. Then μ is σ-finite and for every $A \in \mathfrak{B}(X)$ and any given $\varepsilon > 0$ there exists an open set O and a closed set F such that*

$$F \subseteq A \subseteq O \quad \textit{and} \quad \mu(O\backslash F) \leq \varepsilon. \tag{7.18}$$

Proof. That μ is σ-finite is immediate from the definitions since for any fixed $x_0 \in X$ the open balls $X_n = B_n(x_0)$ have finite measure and satisfy $X_n \nearrow X$.

To see that (7.18) holds we begin with the case when μ is finite. Denote by $\mathcal{A}$ the set of all Borel sets satisfying (7.18). Then $\mathcal{A}$ contains every closed set F: Given F define $O_n = \{x \in X | d(x, F) < 1/n\}$ and note that O_n are open sets which satisfy $O_n \searrow F$. Thus by Theorem 7.1 (iii) $\mu(O_n\backslash F) \to 0$ and hence $F \in \mathcal{A}$.

Moreover, $\mathcal{A}$ is even a σ-algebra. That it is closed under complements is easy to see (note that $\tilde{O} = X\backslash F$ and $\tilde{F} = X\backslash O$ are the required sets for $\tilde{A} = X\backslash A$). To see that it is closed under countable unions consider $A = \bigcup_{n=1}^\infty A_n$ with $A_n \in \mathcal{A}$. Then there are F_n, O_n such that $\mu(O_n\backslash F_n) \leq \varepsilon 2^{-n-1}$. Now $O = \bigcup_{n=1}^\infty O_n$ is open and $F = \bigcup_{n=1}^N F_n$ is closed for any finite N. Since $\mu(A)$ is finite we can choose N sufficiently large such that $\mu(\bigcup_{N+1}^\infty F_n) \leq \varepsilon/2$. Then we have found two sets of the required type: $\mu(O\backslash F) \leq \sum_{n=1}^\infty \mu(O_n\backslash F_n) + \mu(\bigcup_{N+1}^\infty F_n) \leq \varepsilon$. Thus $\mathcal{A}$ is a σ-algebra containing the open sets, hence it is the entire Borel algebra.

Now suppose μ is not finite. Pick some $x_0 \in X$ and set $X_0 = B_{2/3}(x_0)$ and $X_n = B_{n+2/3}(x_0)\backslash\overline{B_{n-2/3}(x_0)}$, $n \in \mathbb{N}$. Let $A_n = A \cap X_n$ and note that $A = \bigcup_{n=0}^\infty A_n$. By the finite case we can choose $F_n \subseteq A_n \subseteq O_n \subseteq X_n$ such that $\mu(O_n\backslash F_n) \leq \varepsilon 2^{-n-1}$. Now set $F = \bigcup_n F_n$ and $O = \bigcup_n O_n$ and observe that F is closed. Indeed, let $x \in \overline{F}$ and let x_j be some sequence from F converging to x. Since $d(x_0, x_j) \to d(x_0, x)$ this sequence must eventually lie in $F_n \cup F_{n+1}$ for some fixed n implying $x \in \overline{F_n \cup F_{n+1}} = F_n \cup F_{n+1} \subseteq F$. Finally, $\mu(O\backslash F) \leq \sum_{n=0}^\infty \mu(O_n\backslash F_n) \leq \varepsilon$ as required. □

This result immediately gives us outer regularity and, if we strengthen our assumption, also inner regularity.

Corollary 7.15. *Under the assumptions of the previous lemma*

$$\mu(A) = \inf_{O\supseteq A, O \text{ open}} \mu(O) = \sup_{F\subseteq A, F \text{ closed}} \mu(F) \tag{7.19}$$

and μ is outer regular. If every finite ball in X is compact, then μ is also inner regular.

$$\mu(A) = \sup_{K\subseteq A, K \text{ compact}} \mu(K). \tag{7.20}$$

Proof. Finally, (7.19) follows from $\mu(A) = \mu(O) - \mu(O\backslash A) = \mu(F) + \mu(A\backslash F)$ and if every finite ball is compact, for every sequence of closed sets F_n with $\mu(F_n) \to \mu(A)$ we also have compact sets $K_n = F_n \cap \overline{B_n(x_0)}$ with $\mu(K_n) \to \mu(A)$. □

By the Heine–Borel theorem every bounded closed ball in $\mathbb{R}^n$ (or $\mathbb{C}^n$) is compact and thus has finite measure by the very definition of a Borel measure. Hence every Borel measure on $\mathbb{R}^n$ (or $\mathbb{C}^n$) satisfies the assumptions of Lemma 7.14.

An inner regular measure on a Hausdorff space which is locally finite (every point has a neighborhood of finite measure) is called a **Radon measure**. Accordingly every Borel measure on $\mathbb{R}^n$ (or $\mathbb{C}^n$) is automatically a Radon measure.

Example. Since Lebesgue measure on $\mathbb{R}$ is regular, we can cover the rational numbers by an open set of arbitrary small measure (it is also not hard to find such an set directly) but we cannot cover it by an open set of measure zero (since any open set contains an interval and hence has positive measure). However, if we slightly extend the family of admissible sets, this will be possible. ⋄

Looking at the Borel σ-algebra the next general sets after open sets are countable intersections of open sets, known as G_δ sets (here G and δ stand for the german words *Gebiet* and *Durchschnitt*, respectively). The next general sets after closed sets are countable unions of closed sets, known as F_σ sets (here F and σ stand for the french words *fermé* and *somme*, respectively).

Example. The irrational numbers are a G_δ set in $\mathbb{R}$. To see this let x_n be an enumeration of the rational numbers and consider the intersection of the open sets $O_n = \mathbb{R}\backslash\{x_n\}$. The rational numbers are hence an F_σ set. ⋄

Corollary 7.16. *A Borel set in $\mathbb{R}^n$ is measurable if and only if it differs from a G_δ set by a Borel set of measure zero. Similarly, a Borel set in $\mathbb{R}^n$ is measurable if and only if it differs from an F_δ set by a Borel set of measure zero.*

Proof. Since G_δ sets are Borel, only the converse direction is nontrivial. By Lemma 7.14 we can find open sets O_n such that $\mu(O_n \backslash A) \le 1/n$. Now let $G = \bigcap_n O_n$. Then $\mu(G\backslash A) \le \mu(O_n\backslash A) \le 1/n$ for any n and thus $\mu(G\backslash A) = 0$ The second claim is analogous. □

Problem 7.10. *Show directly (without using regularity) that for every $\varepsilon > 0$ there is an open set O of Lebesgue measure $|O| < \varepsilon$ which covers the rational numbers.*

Problem 7.11. *Show that a Borel set $A \subseteq \mathbb{R}$ has Lebesgue measure zero if and only if for every ε there exists a countable set of Intervals I_j which cover A and satisfy $\sum_j |I_j| < \varepsilon$.*

7.5. Integration — Sum me up, Henri

Throughout this section (X, Σ, μ) will be a measure space. A measurable function $s : X \to \mathbb{R}$ is called **simple** if its range is finite; that is, if

$$s = \sum_{j=1}^{p} \alpha_j \, \chi_{A_j}, \qquad \mathrm{Ran}(s) = \{\alpha_j\}_{j=1}^p, \quad A_j = s^{-1}(\alpha_j) \in \Sigma. \tag{7.21}$$

Here χ_A is the **characteristic function** of A; that is, $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ otherwise. Note that the set of simple functions is a vector space and while there are different ways of writing a simple function as a linear combination of characteristic functions, the representation (7.21) is unique.

For a nonnegative simple function s as in (7.21) we define its **integral** as

$$\int_A s \, d\mu = \sum_{j=1}^{p} \alpha_j \, \mu(A_j \cap A). \tag{7.22}$$

Here we use the convention $0 \cdot \infty = 0$.

Lemma 7.17. *The integral has the following properties:*

(i) $\int_A s \, d\mu = \int_X \chi_A \, s \, d\mu$.
(ii) $\int_{\bigcup_{j=1}^\infty A_j} s \, d\mu = \sum_{j=1}^\infty \int_{A_j} s \, d\mu, \qquad A_j \cap A_k = \emptyset$ *for* $j \ne k$.
(iii) $\int_A \alpha \, s \, d\mu = \alpha \int_A s \, d\mu$, $\alpha \ge 0$.
(iv) $\int_A (s + t) d\mu = \int_A s \, d\mu + \int_A t \, d\mu$.
(v) $A \subseteq B \;\Rightarrow\; \int_A s \, d\mu \le \int_B s \, d\mu$.
(vi) $s \le t \;\Rightarrow\; \int_A s \, d\mu \le \int_A t \, d\mu$.

Proof. (i) is clear from the definition. (ii) follows from σ-additivity of μ. (iii) is obvious. (iv) Let $s = \sum_j \alpha_j \, \chi_{A_j}$, $t = \sum_j \beta_j \, \chi_{B_j}$ and abbreviate

$C_{jk} = (A_j \cap B_k) \cap A$. Then, by (ii),

$$\int_A (s+t) d\mu = \sum_{j,k} \int_{C_{jk}} (s+t) d\mu = \sum_{j,k} (\alpha_j + \beta_k) \mu(C_{jk})$$
$$= \sum_{j,k} \left(\int_{C_{jk}} s\, d\mu + \int_{C_{jk}} t\, d\mu \right) = \int_A s\, d\mu + \int_A t\, d\mu.$$

(v) follows from monotonicity of μ. (vi) follows since by (iv) we can write $s = \sum_j \alpha_j \chi_{C_j}$, $t = \sum_j \beta_j \chi_{C_j}$ where, by assumption, $\alpha_j \le \beta_j$. □

Our next task is to extend this definition to nonnegative measurable functions by

$$\int_A f\, d\mu = \sup_{s \le f} \int_A s\, d\mu, \tag{7.23}$$

where the supremum is taken over all simple functions $s \le f$. Note that, except for possibly (ii) and (iv), Lemma 7.17 still holds for arbitrary nonnegative functions s, t.

Theorem 7.18 (Monotone convergence, Beppo Levi's theorem)**.** *Let f_n be a monotone nondecreasing sequence of nonnegative measurable functions, $f_n \nearrow f$. Then*

$$\int_A f_n\, d\mu \to \int_A f\, d\mu. \tag{7.24}$$

Proof. By property (vi), $\int_A f_n\, d\mu$ is monotone and converges to some number α. By $f_n \le f$ and again (vi) we have

$$\alpha \le \int_A f\, d\mu.$$

To show the converse, let s be simple such that $s \le f$ and let $\theta \in (0,1)$. Put $A_n = \{x \in A | f_n(x) \ge \theta s(x)\}$ and note $A_n \nearrow A$ (show this). Then

$$\int_A f_n\, d\mu \ge \int_{A_n} f_n\, d\mu \ge \theta \int_{A_n} s\, d\mu.$$

Letting $n \to \infty$, we see

$$\alpha \ge \theta \int_A s\, d\mu.$$

Since this is valid for every $\theta < 1$, it still holds for $\theta = 1$. Finally, since $s \le f$ is arbitrary, the claim follows. □

In particular

$$\int_A f\, d\mu = \lim_{n\to\infty} \int_A s_n\, d\mu, \tag{7.25}$$

for every monotone sequence $s_n \nearrow f$ of simple functions. Note that there is always such a sequence, for example,

$$s_n(x) = \sum_{k=0}^{n2^n} \frac{k}{2^n} \chi_{f^{-1}(A_k)}(x), \quad A_k = [\frac{k}{2^n}, \frac{k+1}{2^n}),\ A_{n2^n} = [n, \infty). \tag{7.26}$$

By construction s_n converges uniformly if f is bounded, since $s_n(x) = n$ if $f(x) \geq n$ and $f(x) - s_n(x) < \frac{1}{2^n}$ if $f(x) \leq n$.

Now what about the missing items (ii) and (iv) from Lemma 7.17? Since limits can be spread over sums, the extension is linear (i.e., item (iv) holds) and (ii) also follows directly from the monotone convergence theorem. We even have the following result:

Lemma 7.19. *If $f \geq 0$ is measurable, then $d\nu = f\, d\mu$ defined via*

$$\nu(A) = \int_A f\, d\mu \tag{7.27}$$

is a measure such that

$$\int g\, d\nu = \int g f\, d\mu \tag{7.28}$$

for every measurable function g.

Proof. As already mentioned, additivity of μ is equivalent to linearity of the integral and σ-additivity follows from Lemma 7.17 (ii):

$$\nu(\bigcup_{n=1}^{\infty} A_n) = \int (\sum_{n=1}^{\infty} \chi_{A_n}) f\, d\mu = \sum_{n=1}^{\infty} \int \chi_{A_n} f\, d\mu = \sum_{n=1}^{\infty} \nu(A_n).$$

The second claim holds for simple functions and hence for all functions by construction of the integral. □

If f_n is not necessarily monotone, we have at least

Theorem 7.20 (Fatou's lemma)**.** *If f_n is a sequence of nonnegative measurable function, then*

$$\int_A \liminf_{n\to\infty} f_n\, d\mu \leq \liminf_{n\to\infty} \int_A f_n\, d\mu. \tag{7.29}$$

Proof. Set $g_n = \inf_{k\geq n} f_k$. Then $g_n \leq f_n$ implying

$$\int_A g_n\, d\mu \leq \int_A f_n\, d\mu.$$

Now take the lim inf on both sides and note that by the monotone convergence theorem

$$\liminf_{n\to\infty} \int_A g_n\, d\mu = \lim_{n\to\infty} \int_A g_n\, d\mu = \int_A \lim_{n\to\infty} g_n\, d\mu = \int_A \liminf_{n\to\infty} f_n\, d\mu,$$

proving the claim. □

If the integral is finite for both the positive and negative part $f^\pm = \max(\pm f, 0)$ of an arbitrary measurable function f, we call f **integrable** and set

$$\int_A f\, d\mu = \int_A f^+ d\mu - \int_A f^- d\mu. \tag{7.30}$$

The set of all integrable functions is denoted by $\mathcal{L}^1(X, d\mu)$.

Lemma 7.21. *Lemma 7.17 holds for integrable functions s, t.*

Similarly, we handle the case where f is complex-valued by calling f integrable if both the real and imaginary part are and setting

$$\int_A f\, d\mu = \int_A \mathrm{Re}(f) d\mu + \mathrm{i} \int_A \mathrm{Im}(f) d\mu. \tag{7.31}$$

Clearly f is integrable if and only if $|f|$ is.

Lemma 7.22. *For all integrable functions f, g we have*

$$|\int_A f\, d\mu| \le \int_A |f|\, d\mu \tag{7.32}$$

and (triangle inequality)

$$\int_A |f + g|\, d\mu \le \int_A |f|\, d\mu + \int_A |g|\, d\mu. \tag{7.33}$$

Proof. Put $\alpha = \frac{z^*}{|z|}$, where $z = \int_A f\, d\mu$ (without restriction $z \neq 0$). Then

$$|\int_A f\, d\mu| = \alpha \int_A f\, d\mu = \int_A \alpha\, f\, d\mu = \int_A \mathrm{Re}(\alpha\, f)\, d\mu \le \int_A |f|\, d\mu,$$

proving the first claim. The second follows from $|f + g| \le |f| + |g|$. □

Lemma 7.23. *Let f be measurable. Then*

$$\int_X |f|\, d\mu = 0 \qquad \Leftrightarrow \qquad f(x) = 0 \qquad \mu - a.e. \tag{7.34}$$

Moreover, suppose f is nonnegative or integrable. Then

$$\mu(A) = 0 \qquad \Rightarrow \qquad \int_A f\, d\mu = 0. \tag{7.35}$$

Proof. Observe that we have $A = \{x | f(x) \neq 0\} = \bigcup_n A_n$, where $A_n = \{x |\, |f(x)| > \frac{1}{n}\}$. If $\int_X |f| d\mu = 0$ we must have $\mu(A_n) = 0$ for every n and hence $\mu(A) = \lim_{n\to\infty} \mu(A_n) = 0$.

The converse will follow from (7.35) since $\mu(A) = 0$ (with A as before) implies $\int_X |f| d\mu = \int_A |f| d\mu = 0$.

Finally, to see (7.35) note that by our convention $0 \cdot \infty = 0$ it holds for any simple function and hence for any f by definition of the integral (7.23).

Since any function can be written as a linear combination of four nonnegative functions this also implies the last part. □

Note that the proof also shows that if f is not 0 almost everywhere, there is an $\varepsilon > 0$ such that $\mu(\{x| \, |f(x)| \geq \varepsilon\}) > 0$.

In particular, the integral does not change if we restrict the domain of integration to a support of μ or if we change f on a set of measure zero. In particular, functions which are equal a.e. have the same integral.

Finally, our integral is well behaved with respect to limiting operations.

Theorem 7.24 (Dominated convergence). *Let f_n be a convergent sequence of measurable functions and set $f = \lim_{n\to\infty} f_n$. Suppose there is an integrable function g such that $|f_n| \leq g$. Then f is integrable and*

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu. \tag{7.36}$$

Proof. The real and imaginary parts satisfy the same assumptions and so do the positive and negative parts. Hence it suffices to prove the case where f_n and f are nonnegative.

By Fatou's lemma

$$\liminf_{n\to\infty} \int_A f_n d\mu \geq \int_A f \, d\mu$$

and

$$\liminf_{n\to\infty} \int_A (g - f_n) d\mu \geq \int_A (g - f) d\mu.$$

Subtracting $\int_A g \, d\mu$ on both sides of the last inequality finishes the proof since $\liminf(-f_n) = -\limsup f_n$. □

Remark: Since sets of measure zero do not contribute to the value of the integral, it clearly suffices if the requirements of the dominated convergence theorem are satisfied almost everywhere (with respect to μ).

Example. Note that the existence of g is crucial: The functions $f_n(x) = \frac{1}{2n}\chi_{[-n,n]}(x)$ on $\mathbb{R}$ converge uniformly to 0 but $\int_{\mathbb{R}} f_n(x) dx = 1$. ◇

Example. If $\mu(x) = \Theta(x)$ is the Dirac measure at 0, then

$$\int_{\mathbb{R}} f(x) d\mu(x) = f(0).$$

In fact, the integral can be restricted to any support and hence to $\{0\}$.

If $\mu(x) = \sum_n \alpha_n \Theta(x - x_n)$ is a sum of Dirac measures, $\Theta(x)$ centered at $x = 0$, then (Problem 7.12)

$$\int_{\mathbb{R}} f(x) d\mu(x) = \sum_n \alpha_n f(x_n).$$

Hence our integral contains sums as special cases. ◇

Problem 7.12. *Consider a countable set of measures* μ_n *and numbers* $\alpha_n \geq 0$*. Let* $\mu = \sum_n \alpha_n \mu_n$ *and show*

$$\int_A f\, d\mu = \sum_n \alpha_n \int_A f\, d\mu_n \tag{7.37}$$

for any measurable function which is either nonnegative or integrable.

Problem 7.13. *Show that the set* $B(X)$ *of bounded measurable functions with the sup norm is a Banach space. Show that the set* $S(X)$ *of simple functions is dense in* $B(X)$*. Show that the integral is a bounded linear functional on* $B(X)$ *if* $\mu(X) < \infty$*. (Hence Theorem 1.29 could be used to extend the integral from simple to bounded measurable functions.)*

Problem 7.14. *Show that the dominated convergence theorem implies (under the same assumptions)*

$$\lim_{n\to\infty} \int |f_n - f| d\mu = 0.$$

Problem 7.15. *Let* $X \subseteq \mathbb{R}$*,* Y *be some measure space, and* $f : X \times Y \to \mathbb{R}$*. Suppose* $y \mapsto f(x, y)$ *is measurable for every* x *and* $x \mapsto f(x, y)$ *is continuous for every* y*. Show that*

$$F(x) = \int_A f(x, y)\, d\mu(y) \tag{7.38}$$

is continuous if there is an integrable function $g(y)$ *such that* $|f(x, y)| \leq g(y)$*.*

Problem 7.16. *Let* $X \subseteq \mathbb{R}$*,* Y *be some measure space, and* $f : X \times Y \to \mathbb{R}$*. Suppose* $y \mapsto f(x, y)$ *is integrable for all* x *and* $x \mapsto f(x, y)$ *is differentiable for a.e.* y*. Show that*

$$F(x) = \int_A f(x, y)\, d\mu(y) \tag{7.39}$$

is differentiable if there is an integrable function $g(y)$ *such that* $|\frac{\partial}{\partial x} f(x, y)| \leq g(y)$*. Moreover,* $y \mapsto \frac{\partial}{\partial x} f(x, y)$ *is measurable and*

$$F'(x) = \int_A \frac{\partial}{\partial x} f(x, y)\, d\mu(y) \tag{7.40}$$

in this case. (See Problem 9.24 for an extension.)

7.6. Product measures

Let μ_1 and μ_2 be two measures on Σ_1 and Σ_2, respectively. Let $\Sigma_1 \otimes \Sigma_2$ be the σ-algebra generated by **rectangles** of the form $A_1 \times A_2$.

Example. Let $\mathfrak{B}$ be the Borel sets in $\mathbb{R}$. Then $\mathfrak{B}^2 = \mathfrak{B} \otimes \mathfrak{B}$ are the Borel sets in $\mathbb{R}^2$ (since the rectangles are a basis for the product topology). ◇

Any set in $\Sigma_1 \otimes \Sigma_2$ has the **section property**; that is,

Lemma 7.25. *Suppose $A \in \Sigma_1 \otimes \Sigma_2$. Then its sections*

$$A_1(x_2) = \{x_1 | (x_1, x_2) \in A\} \quad \text{and} \quad A_2(x_1) = \{x_2 | (x_1, x_2) \in A\} \tag{7.41}$$

are measurable.

Proof. Denote all sets $A \in \Sigma_1 \otimes \Sigma_2$ with the property that $A_1(x_2) \in \Sigma_1$ by S. Clearly all rectangles are in S and it suffices to show that S is a σ-algebra. Now, if $A \in S$, then $(A')_1(x_2) = (A_1(x_2))' \in \Sigma_1$ and thus S is closed under complements. Similarly, if $A_n \in S$, then $(\bigcup_n A_n)_1(x_2) = \bigcup_n (A_n)_1(x_2)$ shows that S is closed under countable unions. □

This implies that if f is a measurable function on $X_1 \times X_2$, then $f(., x_2)$ is measurable on X_1 for every x_2 and $f(x_1, .)$ is measurable on X_2 for every x_1 (observe $A_1(x_2) = \{x_1 | f(x_1, x_2) \in B\}$, where $A = \{(x_1, x_2) | f(x_1, x_2) \in B\}$).

Given two measures μ_1 on Σ_1 and μ_2 on Σ_2, we now want to construct the **product measure** $\mu_1 \otimes \mu_2$ on $\Sigma_1 \otimes \Sigma_2$ such that

$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2), \quad A_j \in \Sigma_j, \; j = 1, 2. \tag{7.42}$$

Since the rectangles are closed under intersection Theorem 7.5 implies that there is at most one measure on $\Sigma_1 \otimes \Sigma_2$ provided μ_1 and μ_2 are σ-finite.

Theorem 7.26. *Let μ_1 and μ_2 be two σ-finite measures on Σ_1 and Σ_2, respectively. Let $A \in \Sigma_1 \otimes \Sigma_2$. Then $\mu_2(A_2(x_1))$ and $\mu_1(A_1(x_2))$ are measurable and*

$$\int_{X_1} \mu_2(A_2(x_1)) d\mu_1(x_1) = \int_{X_2} \mu_1(A_1(x_2)) d\mu_2(x_2). \tag{7.43}$$

Proof. As usual, we begin with the case where μ_1 and μ_2 are finite. Let $\mathcal{D}$ be the set of all subsets for which our claim holds. Note that $\mathcal{D}$ contains at least all rectangles. Thus it suffices to show that $\mathcal{D}$ is a Dynkin system by Lemma 7.4. To see this note that measurability and equality of both integrals follow from $A_1(x_2)' = A_1'(x_2)$ (implying $\mu_1(A_1'(x_2)) = \mu_1(X_1) - \mu_1(A_1(x_2))$) for complements and from the monotone convergence theorem for disjoint unions of sets.

If μ_1 and μ_2 are σ-finite, let $X_{i,j} \nearrow X_i$ with $\mu_i(X_{i,j}) < \infty$ for $i = 1, 2$. Now $\mu_2((A \cap X_{1,j} \times X_{2,j})_2(x_1)) = \mu_2(A_2(x_1) \cap X_{2,j})\chi_{X_{1,j}}(x_1)$ and similarly with 1 and 2 exchanged. Hence by the finite case

$$\int_{X_1} \mu_2(A_2 \cap X_{2,j})\chi_{X_{1,j}} d\mu_1 = \int_{X_2} \mu_1(A_1 \cap X_{1,j})\chi_{X_{2,j}} d\mu_2 \tag{7.44}$$

and the σ-finite case follows from the monotone convergence theorem. □

Hence for given $A \in \Sigma_1 \otimes \Sigma_2$ we can define

$$\mu_1 \otimes \mu_2(A) = \int_{X_1} \mu_2(A_2(x_1)) d\mu_1(x_1) = \int_{X_2} \mu_1(A_1(x_2)) d\mu_2(x_2) \qquad (7.45)$$

or equivalently, since $\chi_{A_1(x_2)}(x_1) = \chi_{A_2(x_1)}(x_2) = \chi_A(x_1, x_2)$,

$$\begin{aligned} \mu_1 \otimes \mu_2(A) &= \int_{X_1} \left(\int_{X_2} \chi_A(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1) \\ &= \int_{X_2} \left(\int_{X_1} \chi_A(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2). \end{aligned} \qquad (7.46)$$

Then $\mu_1 \otimes \mu_2$ gives rise to a unique measure on $A \in \Sigma_1 \otimes \Sigma_2$ since σ-additivity follows from the monotone convergence theorem.

Finally we have

Theorem 7.27 (Fubini). *Let f be a measurable function on $X_1 \times X_2$ and let μ_1, μ_2 be σ-finite measures on X_1, X_2, respectively.*

(i) *If $f \geq 0$, then $\int f(., x_2) d\mu_2(x_2)$ and $\int f(x_1, .) d\mu_1(x_1)$ are both measurable and*

$$\begin{aligned} \iint f(x_1, x_2) d\mu_1 \otimes \mu_2(x_1, x_2) &= \int \left(\int f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2) \\ &= \int \left(\int f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1). \end{aligned} \qquad (7.47)$$

(ii) *If f is complex, then*

$$\int |f(x_1, x_2)| d\mu_1(x_1) \in \mathcal{L}^1(X_2, d\mu_2), \qquad (7.48)$$

respectively,

$$\int |f(x_1, x_2)| d\mu_2(x_2) \in \mathcal{L}^1(X_1, d\mu_1), \qquad (7.49)$$

if and only if $f \in \mathcal{L}^1(X_1 \times X_2, d\mu_1 \otimes d\mu_2)$. In this case (7.47) holds.

Proof. By Theorem 7.26 and linearity the claim holds for simple functions. To see (i), let $s_n \nearrow f$ be a sequence of nonnegative simple functions. Then it follows by applying the monotone convergence theorem (twice for the double integrals).

For (ii) we can assume that f is real-valued by considering its real and imaginary parts separately. Moreover, splitting $f = f^+ - f^-$ into its positive and negative parts, the claim reduces to (i). □

In particular, if $f(x_1, x_2)$ is either nonnegative or integrable, then the order of integration can be interchanged.

Lemma 7.28. *If μ_1 and μ_2 are outer regular measures, then so is $\mu_1 \otimes \mu_2$.*

Proof. Outer regularity holds for every rectangle and hence also for the algebra of finite disjoint unions of rectangles (Problem 7.17). Thus the claim follows from Lemma 7.9. □

In connection with Theorem 7.5 the following observation is of interest:

Lemma 7.29. *If S_1 generates Σ_1 and S_2 generates Σ_2, then $S_1 \times S_2 = \{A_1 \times A_2 | A_j \in S_j,\ j = 1, 2\}$ generates $\Sigma_1 \otimes \Sigma_2$.*

Proof. Denote the σ-algebra generated by $S_1 \times S_2$ by Σ. Consider the set $\{A_1 \in \Sigma_1 | A_1 \times X_2 \in \Sigma\}$ which is clearly a σ-algebra containing S_1 and thus equal to Σ_1. In particular, $\Sigma_1 \times X_2 \subset \Sigma$ and similarly $X_1 \times \Sigma_2 \subset \Sigma$. Hence also $(\Sigma_1 \times X_2) \cap (X_1 \times \Sigma_2) = \Sigma_1 \times \Sigma_2 \subset \Sigma$. □

Finally, note that we can iterate this procedure.

Lemma 7.30. *Suppose (X_j, Σ_j, μ_j), $j = 1, 2, 3$, are σ-finite measure spaces. Then $(\Sigma_1 \otimes \Sigma_2) \otimes \Sigma_3 = \Sigma_1 \otimes (\Sigma_2 \otimes \Sigma_3)$ and*

$$(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3). \tag{7.50}$$

Proof. First of all note that $(\Sigma_1 \otimes \Sigma_2) \otimes \Sigma_3 = \Sigma_1 \otimes (\Sigma_2 \otimes \Sigma_3)$ is the sigma algebra generated by the rectangles $A_1 \times A_2 \times A_3$ in $X_1 \times X_2 \times X_3$. Moreover, since

$$\begin{aligned} ((\mu_1 \otimes \mu_2) \otimes \mu_3)(A_1 \times A_2 \times A_3) &= \mu_1(A_1)\mu_2(A_2)\mu_3(A_3) \\ &= (\mu_1 \otimes (\mu_2 \otimes \mu_3))(A_1 \times A_2 \times A_3), \end{aligned}$$

the two measures coincide on rectangles and hence everywhere by Theorem 7.5. □

Example. If λ is Lebesgue measure on $\mathbb{R}$, then $\lambda^n = \lambda \otimes \cdots \otimes \lambda$ is Lebesgue measure on $\mathbb{R}^n$. Since λ is outer regular, so is λ^n. Of course regularity also follows from Corollary 7.15.

Moreover, Lebesgue measure is translation invariant and up to normalization the only measure with this property. To see this let μ be a second translation invariant measure. Denote by Q_r a cube with side length $r > 0$. Without loss we can assume $\mu(Q_1) = 1$. Since we can split Q_1 cube into m^n cubes of side length $1/m$ we see that $\mu(Q_{1/m}) = m^{-n}$ by translation invariance and additivity. Hence we obtain $\mu(Q_r) = r^n$ for every rational r and thus for every r by continuity from below. Proceeding like this we see that λ^n and μ coincide on all rectangles which are products of bounded open intervals. Since this set is closed under intersections and generates the Borel algebra $\mathfrak{B}^n$ by Lemma 7.29 the claim follows again from Theorem 7.5. ⋄

Problem 7.17. *Show that the set of all finite union of rectangles $A_1 \times A_2$ forms an algebra. Moreover, every set in this algebra can be written a finite union of disjoint rectangles.*

Problem 7.18. *Let $U \subseteq \mathbb{C}$ be a domain, Y be some measure space, and $f : U \times Y \to \mathbb{R}$. Suppose $y \mapsto f(z,y)$ is measurable for every z and $z \mapsto f(z,y)$ is holomorphic for every y. Show that*

$$F(z) = \int_A f(z,y)\, d\mu(y) \tag{7.51}$$

is holomorphic if for every compact subset $V \subset U$ there is an integrable function $g(y)$ such that $|f(z,y)| \le g(y)$, $z \in V$. (Hint: Use Fubini and Morera.)

7.7. Transformation of measures and integrals

Finally we want to transform measures. Let $f : X \to Y$ be a measurable function. Given a measure μ on X we can introduce a measure $f_\star\mu$ on Y via

$$(f_\star\mu)(A) = \mu(f^{-1}(A)). \tag{7.52}$$

It is straightforward to check that $f_\star\mu$ is indeed a measure. Moreover, note that $f_\star\mu$ is supported on the range of f.

Theorem 7.31. *Let $f : X \to Y$ be measurable and let $g : Y \to \mathbb{C}$ be a Borel function. Then the Borel function $g \circ f : X \to \mathbb{C}$ is a.e. nonnegative or integrable if and only if g is and in both cases*

$$\int_Y g\, d(f_\star\mu) = \int_X g \circ f\, d\mu. \tag{7.53}$$

Proof. In fact, it suffices to check this formula for simple functions g, which follows since $\chi_A \circ f = \chi_{f^{-1}(A)}$. □

Example. Let $f(x) = Mx + a$ be an affine transformation, where $M : \mathbb{R}^n \to \mathbb{R}^n$ is some invertible matrix. Then Lebesgue measure transforms according to

$$f_\star\lambda^n = \frac{1}{|\det(M)|}\lambda^n.$$

In fact, it suffices to check $f_\star\lambda^n(R) = |\det(M)|^{-1}\lambda^n(R)$ for finite rectangles R by Theorem 7.5, which follows from the corresponding result in linear algebra.

As a consequence we obtain

$$\int_A g(Mx + a) d^n x = \frac{1}{|\det(M)|} \int_{MA+a} g(y) d^n y,$$

which applies for example to shifts $f(x) = x + a$ or scaling transforms $f(x) = \alpha x$. $\diamond$

This result can be generalized to diffeomorphisms:

Lemma 7.32. *Let $U, V \subseteq \mathbb{R}^n$ and suppose $f \in C^1(U, V)$ is bijective. Then*

$$(f^{-1})_\star d^n x = |J_f(x)| d^n x, \tag{7.54}$$

where $J_f = \det(\frac{\partial f}{\partial x})$ is the Jacobi determinant of f. In particular,

$$\int_U g(f(x)) |J_f(x)| d^n x = \int_V g(y) d^n y. \tag{7.55}$$

Proof. It suffices to show

$$\int_{f(R)} d^n y = \int_R |J_f(x)| d^n x$$

for every bounded rectangle $R \subseteq U$. Since the partial derivatives of f are uniformly continuous on every compact subset $K \subset U$ we can find a δ for every ε such that

$$\left| f(x+y) - f(x) - \frac{\partial f}{\partial x}(x_0) y \right| \leq \varepsilon |y|, \qquad |J_f(x+y) - J_f(x)| \leq \varepsilon$$

whenever $x \in K$ and $|y| \leq \delta$. Hence if $R \subset \overline{B_r(x_0)}$ with $x_0 \in K$ and $r \leq \delta$, then the image $f(R) - f(x_0)$ is contained in the rectangle whose sides are transformed with the Jacobian $\frac{\partial f}{\partial x}(x_0)$ and extended by ε on both ends. Thus there is some constant C_n with

$$\left| \int_{f(R)} d^n y - \int_R |J_f(x_0)| d^n x \right| \leq \varepsilon C_n r^{n-1}.$$

Moreover,

$$\left| \int_R |J_f(x_0)| d^n x - \int_R |J_f(x)| d^n x \right| \leq 2\varepsilon V_n r^n,$$

where V_n is the volume of the unit ball (cf. below). Now if $R \subseteq K$ is an arbitrary rectangle fitting into a ball of radius r we can divide it into m^n smaller rectangles which will fit into balls of radius r/m. Choosing m such that $r/m \leq \delta$ and using additivity we can remove the restriction $r \leq \delta$. Since $\varepsilon > 0$ was arbitrary the claim follows. $\square$

Example. For example we can consider **polar coordinates** $T_2 : [0, \infty) \times [0, 2\pi) \to \mathbb{R}^2$ defined by

$$T_2(\rho, \varphi) = (\rho \cos(\varphi), \rho \sin(\varphi)).$$

Then

$$\det \frac{\partial T_2}{\partial(\rho, \varphi)} = \rho$$

and one has

$$\int_U f(\rho\cos(\varphi), \rho\sin(\varphi))\rho\, d(\rho,\varphi) = \int_{T_2(U)} f(x)dx.$$

Note that T_2 is only bijective when restricted to $(0,\infty)\times[0,2\pi)$. However, since the set $\{0\}\times[0,2\pi)$ is of measure zero it does not contribute to the integral on the left. Similarly its image $T_2(\{0\}\times[0,2\pi)) = \{0\}$ does not contribute to the integral on the right. ⋄

Example. We can use the previous example to obtain the transformation formula for **spherical coordinates** in $\mathbb{R}^n$ by induction. We illustrate the process for $n = 3$. To this end let $x = (x_1, x_2, x_3)$ and start with spherical coordinates in $\mathbb{R}^2$ (which are just polar coordinates) for the first two components:

$$x = (\rho\cos(\varphi), \rho\sin(\varphi), x_3), \qquad \rho\in[0,\infty),\ \varphi\in[0,2\pi).$$

Next use polar coordinates for (ρ, x_3):

$$(\rho, x_3) = (r\sin(\theta), r\cos(\theta)), \qquad r\in[0,\infty),\ \theta\in[0,\pi].$$

Note that the range for θ follows since $\rho\ge 0$. Moreover, observe that $r^2 = \rho^2 + x_3^2 = x_1^2 + x_2^2 + x_3^2 = |x|^2$ as already anticipated by our notation. In summary,

$$x = T_3(r,\varphi,\theta) = (r\sin(\theta)\cos(\varphi), r\sin(\theta)\sin(\varphi), r\cos(\theta)).$$

Furthermore, since T_3 is the composition with T_2 acting on the first two coordinates with the last unchanged and polar coordinates P acting on the first and last coordinate, the chain rule implies

$$\det\frac{\partial T_3}{\partial(r,\varphi,\theta)} = \det\frac{\partial T_2}{\partial(\rho,\varphi,x_3)}\Big|_{\substack{\rho=r\sin(\theta)\\ x_3=r\cos(\theta)}} \det\frac{\partial P}{\partial(r,\varphi,\theta)} = r^2\sin(\theta).$$

Hence one has

$$\int_U f(T_3(r,\varphi,\theta))r^2\sin(\theta)d(r,\varphi,\theta) = \int_{T_3(U)} f(x)dx.$$

Again T_3 is only bijective on $(0,\infty)\times[0,2\pi)\times(0,\pi)$.

It is left as an exercise to check that the extension $T_n : [0,\infty)\times[0,2\pi)\times[0,\pi]^{n-2}\to\mathbb{R}^n$ is given by

$$x = T_n(r,\varphi,\theta_1,\dots,\theta_{n-1})$$

with

$$\begin{aligned} x_1 &= r\cos(\varphi)\sin(\theta_1)\sin(\theta_2)\sin(\theta_3)\cdots\sin(\theta_{n-2}),\\ x_2 &= r\sin(\varphi)\sin(\theta_1)\sin(\theta_2)\sin(\theta_3)\cdots\sin(\theta_{n-2}),\\ x_3 &= r\cos(\theta_1)\sin(\theta_2)\sin(\theta_3)\cdots\sin(\theta_{n-2}),\\ x_4 &= r\cos(\theta_2)\sin(\theta_3)\cdots\sin(\theta_{n-2}),\\ &\vdots\\ x_{n-1} &= r\cos(\theta_{n-3})\sin(\theta_{n-2}),\\ x_n &= r\cos(\theta_{n-2}). \end{aligned}$$

The Jacobi determinant is given by

$$\det\frac{\partial T_n}{\partial(r,\varphi,\theta_1,\dots,\theta_{n-2})} = r^{n-1}\sin(\theta_1)\sin(\theta_2)^2\cdots\sin(\theta_{n-2})^{n-2}.$$

◇

Another useful consequence of Theorem 7.31 is the following rule for integrating radial functions.

Lemma 7.33. *Let $g:[0,\infty)\to\mathbb{R}$ be measurable. Then*

$$\int_{\mathbb{R}^n} g(|x|)d^n x = nV_n\int_0^\infty g(r)r^{n-1}dr, \tag{7.56}$$

where V_n is the volume of the unit ball in $\mathbb{R}^n$.

Proof. Consider the transformation $f(x)=|x|$. Then the distribution function of $f_\star\lambda^n$ is given by $(f_\star\lambda^n)(r) = (f_\star\lambda^n)([0,r)) = \lambda^n(f^{-1}([0,r))) = \lambda^n(B_r(0)) = V_n r^n$ and the claim follows from Theorem 7.31. □

Example. Let us compute the volume of a ball in $\mathbb{R}^n$:

$$V_n(r) = \int_{\mathbb{R}^n}\chi_{B_r(0)}d^n x.$$

By the simple scaling transform $f(x)=rx$ we obtain $V_n(r)=V_n(1)r^n$ and hence it suffices to compute $V_n=V_n(1)$. Moreover, $V_1=2$ and $V_2=\pi$ are left as an exercise and we will use induction:

$$\begin{aligned} V_n &= \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}^{n-2}}\chi_{B_{(1-x_1^2+x_2^2)^{1/2}}(0)}(x_3,\cdots,x_n)dx_n\cdots dx_3dx_2dx_1\\ &= V_{n-2}\int_{\mathbb{R}}\int_{\mathbb{R}}\max(0,1-x_1^2+x_2^2)^{(n-2)/2}dx_1dx_2\\ &= 2V_{n-2}V_2\int_0^1(1-r^2)^{(n-2)/2}r\,dr = 2\pi V_{n-2}/n. \end{aligned}$$

Here we have used (7.56) in the last step. In summary

$$V_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}, \tag{7.57}$$

where Γ is the Gamma function and we recall $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. $\diamond$

Problem 7.19. *Let λ be Lebesgue measure on $\mathbb{R}$. Show that if $f \in C^1(\mathbb{R})$ with $f' > 0$, then*

$$d(f_\star\lambda)(x) = \frac{1}{f'(f^{-1}(x))} dx.$$

Problem 7.20. *Compute V_n using spherical coordinates. (Hint: $\int \sin(x)^n dx = -\frac{1}{n}\sin(x)^{n-1}\cos(x) + \frac{n-1}{n}\int \sin(x)^{n-2} dx$.)*

7.8. Appendix: Transformation of Lebesgue–Stieltjes integrals

In this section we will look at Borel measures on $\mathbb{R}$. In particular, we want to derive a generalized substitution rule.

As a preparation we will need a generalization of the usual inverse which works for arbitrary nondecreasing functions. Such a generalized inverse arises for example as quantile functions in probability theory.

So we look at nondecreasing functions $f : \mathbb{R} \to \mathbb{R}$. By monotonicity the limits from left and right exist at every point and we will denote them by

$$f(x\pm) = \lim_{\varepsilon \downarrow 0} f(x \pm \varepsilon). \tag{7.58}$$

Clearly we have $f(x-) \le f(x+)$ and a strict inequality can occur only at a countable number of points. By monotonicity the value of f has to lie between these two values $f(x-) \le f(x) \le f(x+)$. It will also be convenient to extend f to a function on the extended reals $\mathbb{R} \cup \{-\infty, +\infty\}$. Again by monotonicity the limits $f(\pm\infty\mp) = \lim_{x\to\pm\infty} f(x)$ exist and we will set $f(\pm\infty\pm) = \pm\infty$.

If we want to define an inverse, problems will occur at points where f jumps and on intervals where f is constant. Formally speaking, if f jumps, then the corresponding jump will be missing in the domain of the inverse and if f is constant, the inverse will be multivalued. For the first case there is a natural fix by choosing the inverse to be constant along the missing interval. In particular, observe that this natural choice is independent of the actual value of f at the jump and hence the inverse *loses* this information. The second case will result in a jump for the inverse function and here there is no natural choice for the value at the jump (except that it must be between the left and right limits such that the inverse is again a nondecreasing function).

To give a precise definition it will be convenient to look at relations instead of functions. Recall that a (binary) relation R on $\mathbb{R}$ is a subset of $\mathbb{R}^2$.

To every nondecreasing function f associate the relation

$$\Gamma(f) = \{(x, y)|y \in [f(x-), f(x+)]\}. \tag{7.59}$$

Note that $\Gamma(f)$ does not depend on the values of f at a discontinuity and f can be partially recovered from $\Gamma(f)$ using $f(x-) = \inf \Gamma(f)(x)$ and $f(x+) = \sup \Gamma(f)(x)$, where $\Gamma(f)(x) = \{y|(x, y) \in \Gamma(f)\} = [f(x-), f(x+)]$. Moreover, the relation is nondecreasing in the sense that $x_1 < x_2$ implies $y_1 \le y_2$ for $(x_1, y_1), (x_2, y_2) \in \Gamma(f)$. It is uniquely defined as the largest relation containing the graph of f with this property.

The graph of any reasonable inverse should be a subset of the inverse relation

$$\Gamma(f)^{-1} = \{(y, x)|(x, y) \in \Gamma(f)\} \tag{7.60}$$

and we will call any function f^{-1} whose graph is a subset of $\Gamma(f)^{-1}$ a **generalized inverse** of f. Note that any generalized inverse is again nondecreasing since a pair of points $(y_1, x_1), (y_2, x_2) \in \Gamma(f)^{-1}$ with $y_1 < y_2$ and $x_1 > x_2$ would contradict the fact that $\Gamma(f)$ is nondecreasing. Moreover, since $\Gamma(f)^{-1}$ and $\Gamma(f^{-1})$ are two nondecreasing relations containing the graph of f^{-1}, we conclude

$$\Gamma(f^{-1}) = \Gamma(f)^{-1} \tag{7.61}$$

since both are maximal. In particular, it follows that if f^{-1} a generalized inverse of f then f is a generalized inverse of f^{-1}.

There are two particular choices, namely the left continuous version $f_-^{-1}(y) = \inf \Gamma(f)^{-1}(y)$ and the right continuous version $f_+^{-1}(y) = \sup \Gamma(f)^{-1}(y)$. It is straightforward to verify that they can be equivalently defined via

$$\begin{aligned} f_-^{-1}(y) &= \inf f^{-1}([y, \infty)) = \sup f^{-1}((-\infty, y)), \\ f_+^{-1}(y) &= \inf f^{-1}((y, \infty)) = \sup f^{-1}((-\infty, y]). \end{aligned} \tag{7.62}$$

For example, $\inf f^{-1}([y, \infty)) = \inf\{x|(x, \tilde{y}) \in \Gamma(f),\ \tilde{y} \ge y\} = \inf \Gamma(f)^{-1}(y)$. The first one is typically used in probability theory, where it corresponds to the quantile function of a distribution.

If f is strictly increasing the generalized inverse coincides with the usual inverse and we have $f(f^{-1}(y)) = y$ for y in the range of f. The purpose of the next lemma is to investigate to what extend this remains valid for a generalized inverse.

Note that for every y there is some x with $y \in [f(x-), f(x+)]$. Moreover, if we can find two values, say x_1 and x_2, with this property, then $f(x) = y$ is constant for $x \in (x_1, x_2)$. Hence, the set of all such x is an interval which is closed since at the left, right boundary point the left, right limit equals y, respectively.

We collect a few simple facts for later use.

Lemma 7.34. *Let m be nondecreasing.*

(i) $f_-^{-1}(y) \le x$ *if and only if* $y \le f(x+)$.

(i') $f_+^{-1}(y) \ge x$ *if and only if* $y \ge f(x-)$.

(ii) $f(f_-^{-1}(y)) \ge y$ *if f is right continuous at* $f_-^{-1}(y)$ *with equality if* $y \in \mathrm{Ran}(f)$.

(ii') $f(f_+^{-1}(y)) \le y$ *if f is left continuous at* $f_+^{-1}(y)$ *with equality if* $y \in \mathrm{Ran}(f)$.

Proof. Item (i) follows since both claims are equivalent to $y \le f(\tilde{x})$ for all $\tilde{x} > x$. Next, let f be right-continuous. If y is in the range of f, then $y = f(f_-^{-1}(y)+) = f(f_-^{-1}(y))$ if $f^{-1}(\{y\})$ contains more than one point and $y = f(f_-^{-1}(y))$ else. If y is not in the range of m we must have $y \in [f(x-), f(x+)]$ for $x = f_-^{-1}(y)$ which establishes item (ii). Similarly for (i') and (ii'). □

In particular, $f(f^{-1}(y)) = y$ if f is continuous. We will also need the set

$$L(f) = \{y | f^{-1}((y,\infty)) = (f_+^{-1}(y), \infty)\}. \tag{7.63}$$

Note that $y \notin L(f)$ if and only if there is some x such that $y \in [f(x-), f(x))$.

Lemma 7.35. *Let $m : \mathbb{R} \to \mathbb{R}$ be a nondecreasing function on $\mathbb{R}$ and μ its associated measure via (7.5). Let $f(x)$ be a nondecreasing function on $\mathbb{R}$ such that $\mu((0,\infty)) < \infty$ if f is bounded above and $\mu((-\infty, 0)) < \infty$ if f is bounded below.*

Then $f_\star\mu$ is a Borel measure whose distribution function coincides up to a constant with $m_+ \circ f_+^{-1}$ at every point y which is in $L(f)$ or satisfies $\mu(\{f_+^{-1}(y)\}) = 0$. If $y \in [f(x-), f(x))$ and $\mu(\{f_+^{-1}(y)\}) > 0$, then $m_+ \circ f_+^{-1}$ jumps at $f(x-)$ and $(f_\star\mu)(y)$ jumps at $f(x)$.

Proof. First of all note that the assumptions in case f is bounded from above or below ensure that $(f_\star\mu)(K) < \infty$ for any compact interval. Moreover, we can assume $m = m_+$ without loss of generality. Now note that we have $f^{-1}((y,\infty)) = (f^{-1}(y), \infty)$ for $y \in L(f)$ and $f^{-1}((y,\infty)) = [f^{-1}(y), \infty)$ else. Hence

$$\begin{aligned}(f_\star\mu)((y_0, y_1]) &= \mu(f^{-1}((y_0, y_1])) = \mu((f^{-1}(y_0), f^{-1}(y_1)]) \\ &= m(f_+^{-1}(y_1)) - m(f_+^{-1}(y_0)) = (m \circ f_+^{-1})(y_1) - (m \circ f_+^{-1})(y_0)\end{aligned}$$

if y_j is either in $L(f)$ or satisfies $\mu(\{f_+^{-1}(y_j)\}) = 0$. For the last claim observe that $f^{-1}((y,\infty))$ will jump from $(f_+^{-1}(y), \infty)$ to $[f_+^{-1}(y), \infty)$ at $y = f(x)$. □

Example. For example, consider $f(x) = \chi_{[0,\infty)}(x)$ and $\mu = \Theta$, the Dirac measure centered at 0 (note that $\Theta(x) = f(x)$). Then

$$f_+^{-1}(y) = \begin{cases} +\infty, & 1 \le y, \\ 0, & 0 \le y < 1, \\ -\infty, & y < 0, \end{cases}$$

and $L(f) = (-\infty, 0) \cup [1, \infty)$. Moreover, $\mu(f_+^{-1}(y)) = \chi_{[0,\infty)}(y)$ and $(f_\star\mu)(y) = \chi_{[1,\infty)}(y)$. If we choose $g(x) = \chi_{(0,\infty)}(x)$, then $g_+^{-1}(y) = f_+^{-1}(y)$ and $L(g) = \mathbb{R}$. Hence $\mu(g_+^{-1}(y)) = \chi_{[0,\infty)}(y) = (g_\star\mu)(y)$. ⋄

For later use it is worth while to single out the following consequence:

Corollary 7.36. *Let m, f be as in the previous lemma and denote by μ, $\nu_\pm$ the measures associated with m, $m_\pm \circ f^{-1}$, respectively. Then, $(f_\mp)_\star\mu = \nu_\pm$ and hence*

$$\int g\, d(m_\pm \circ f^{-1}) = \int (g \circ f_\mp)\, dm. \tag{7.64}$$

In the special case where μ is Lebesgue measure this reduces to a way of expressing the Lebesgue–Stieltjes integral as a Lebesgue integral via

$$\int g\, dh = \int g(h^{-1}(y)) dy. \tag{7.65}$$

If we choose f to be the distribution function of μ we get the following generalization of the **integration by substitution** rule. To formulate it we introduce

$$i_m(y) = m(m_-^{-1}(y)). \tag{7.66}$$

Note that $i_m(y) = y$ if m is continuous. By hull(Ran(m)) we denote the convex hull of the range of m.

Corollary 7.37. *Suppose m, m are two nondecreasing functions on $\mathbb{R}$ with n right continuous. Then we have*

$$\int_{\mathbb{R}} (g \circ m)\, d(n \circ m) = \int_{\mathrm{hull}(\mathrm{Ran}(m))} (g \circ i_m) dn \tag{7.67}$$

for any Borel function g which is either nonnegative or for which one of the two integrals is finite. Similarly, if n is left continuous and i_m is replaced by $m(m_+^{-1}(y))$.

Hence the usual $\int_{\mathbb{R}} (g \circ m)\, d(n \circ m) = \int_{\mathrm{Ran}(m)} g\, dn$ only holds if m is continuous. In fact, the right-hand side looses all point masses of μ. The above formula fixes this problem by rendering g constant along a gap in the range of m and includes the gap in the range of integration such that it makes up for the lost point mass. It should be compared with the previous example!

If one does not want to bother with i_m one can at least get inequalities for monotone g.

Corollary 7.38. *Suppose m, n are nondecreasing functions on $\mathbb{R}$ and g is monotone. Then we have*

$$\int_{\mathbb{R}} (g \circ m)\, d(n \circ m) \le \int_{\mathrm{hull}(\mathrm{Ran}(m))} g\, dn \tag{7.68}$$

if m, n are right continuous and g nonincreasing or m, n left continuous and g nondecreasing. If m, n are right continuous and g nondecreasing or m, n left continuous and g nonincreasing the inequality has to be reversed.

Proof. Immediate from the previous corollary together with $i_m(y) \le y$ if $y = f(x) = f(x+)$ and $i_m(y) \ge y$ if $y = f(x) = f(x-)$ according to Lemma 7.34. □

Problem 7.21. *Show (7.62).*

Problem 7.22. *Show that $\Gamma(f) \circ \Gamma(f^{-1}) = \{(y,z) | y, z \in [f(x-), f(x+)]$ for some $x\}$.*

Problem 7.23. *Let $d\mu(\lambda) = \chi_{[0,1]}(\lambda) d\lambda$ and $f(\lambda) = \chi_{(t,\infty)}(\lambda)$, $t \in \mathbb{R}$. Compute $f_\star \mu$.*

7.9. Appendix: The connection with the Riemann integral

In this section we want to investigate the connection with the Riemann integral. We restrict our attention to compact intervals $[a,b]$ and bounded real-valued functions f. A **partition** of $[a,b]$ is a finite set $P = \{x_0, \dots, x_n\}$ with

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b. \tag{7.69}$$

The number

$$\|P\| = \max_{1 \le j \le n} x_j - x_{j-1} \tag{7.70}$$

is called the norm of P. Given a partition P and a bounded real-valued function f we can define

$$s_{P,f,-}(x) = \sum_{j=1}^{n} m_j \chi_{[x_{j-1},x_j)}(x), \qquad m_j = \inf_{x \in [x_{j-1},x_j)} f(x), \tag{7.71}$$

$$s_{P,f,+}(x) = \sum_{j=1}^{n} M_j \chi_{[x_{j-1},x_j)}(x), \qquad M_j = \sup_{x \in [x_{j-1},x_j)} f(x), \tag{7.72}$$

Hence $s_{P,f,-}(x)$ is a step function approximating f from below and $s_{P,f,+}(x)$ is a step function approximating f from above. In particular,

$$m \le s_{P,f,-}(x) \le f(x) \le s_{P,f,-}(x) \le M, \quad m = \inf_{x \in [a,b]} f(x),\ M = \sup_{x \in [a,b]} f(x). \tag{7.73}$$

Moreover, we can define the upper and lower Riemann sum associated with P as

$$L(P,f) = \sum_{j=1}^{n} m_j(x_j - x_{j-1}), \qquad U(P,f) = \sum_{j=1}^{n} M_j(x_j - x_{j-1}). \tag{7.74}$$

Of course, $L(f,P)$ is just the Lebesgue integral of $s_{P,f,-}$ and $U(f,P)$ is the Lebesgue integral of $s_{P,f,+}$. In particular, $L(P,f)$ approximates the area under the graph of f from below and $U(P,f)$ approximates this area from above.

By the above inequality

$$m\,(b-a) \le L(P,f) \le U(P,f) \le M\,(b-a). \tag{7.75}$$

We say that the partition P_2 is a refinement of P_1 if $P_1 \subseteq P_2$ and it is not hard to check, that in this case

$$s_{P_1,f,-}(x) \le s_{P_2,f,-}(x) \le f(x) \le s_{P_2,f,+}(x) \le s_{P_1,f,+}(x) \tag{7.76}$$

as well as

$$L(P_1,f) \le L(P_2,f) \le U(P_2,f) \le U(P_1,f). \tag{7.77}$$

Hence we define the **lower, upper Riemann integral** of f as

$$\underline{\int} f(x)dx = \sup_P L(P,f), \qquad \overline{\int} f(x)dx = \inf_P U(P,f), \tag{7.78}$$

respectively. Clearly

$$m\,(b-a) \le \underline{\int} f(x)dx \le \overline{\int} f(x)dx \le M\,(b-a). \tag{7.79}$$

We will call f **Riemann integrable** if both values coincide and the common value will be called the **Riemann integral** of f.

Example. Let $[a,b] = [0,1]$ and $f(x) = \chi_{\mathbb{Q}}(x)$. Then $s_{P,f,-}(x) = 0$ and $s_{P,f,+}(x) = 1$. Hence $\underline{\int} f(x)dx = 0$ and $\overline{\int} f(x)dx = 1$ and f is not Riemann integrable.

On the other hand, every continuous function $f \in C[a,b]$ is Riemann integrable (Problem 7.24). ◇

Lemma 7.39. *A function f is Riemann integrable if and only if there exists a sequence of partitions P_j such that*

$$\lim_{j\to\infty} L(P_j,f) = \lim_{j\to\infty} U(P_j,f). \tag{7.80}$$

In this case the above limit equals the Riemann integral of f and P_j can be chosen such that $P_j \subseteq P_{j+1}$ and $\|P_j\| \to 0$.

Proof. If there is such a sequence of partitions then f is integrable by $\lim_j L(P_j, f) \le \sup_P L(P, f) \le \inf_P U(P, f) \le \lim_j U(P_j, f)$.

Conversely, given an integrable f, there is a sequence of partitions $P_{L,j}$ such that $\underline{\int} f(x)dx = \lim_j L(P_{L,j}, f)$ and a sequence $P_{U,j}$ such that $\overline{\int} f(x)dx = \lim_j U(P_{U,j}, f)$. By (7.77) the common refinement $P_j = P_{L,j} \cup P_{U,j}$ is the partition we are looking for. Since, again by (7.77), any refinement will also work, the last claim follows. □

With the help of this lemma we can give a characterization of Riemann integrable functions and show that the Riemann integral coincides with the Lebesgue integral.

Theorem 7.40 (Lebesgue)**.** *A bounded measurable function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if the set of its discontinuities is of Lebesgue measure zero. Moreover, in this case the Riemann and the Lebesgue integral of f coincide.*

Proof. Suppose f is Riemann integrable and let P_j be a sequence of partitions as in Lemma 7.39. Then $s_{f,P_j,-}(x)$ will be monotone and hence converge to some function $s_{f,-}(x) \le f(x)$. Similarly, $s_{f,P_j,+}(x)$ will converge to some function $s_{f,+}(x) \ge f(x)$. Moreover, by dominated convergence

$$0 = \lim_j \int \big(s_{f,P_j,+}(x) - s_{f,P_j,-}(x)\big)dx = \int \big(s_{f,+}(x) - s_{f,-}(x)\big)dx$$

and thus by Lemma 7.23 $s_{f,+}(x) = s_{f,-}(x)$ almost everywhere. Moreover, f is continuous at every x at which equality holds and which is not in any of the partitions. Since the first as well as the second set have Lebesgue measure zero, f is continuous almost everywhere and

$$\lim_j L(P_j, f) = \lim_j U(P_j, f) = \int s_{f,\pm}(x)dx = \int f(x)dx.$$

Conversely, let f be continuous almost everywhere and choose some sequence of partitions P_j with $\|P_j\| \to 0$. Then at every x where f is continuous we have $\lim_j s_{f,P_j,\pm}(x) = f(x)$ implying

$$\lim_j L(P_j, f) = \int s_{f,-}(x)dx = \int f(x)dx = \int s_{f,+}(x)dx = \lim_j U(P_j, f)$$

by the dominated convergence theorem. □

Note that if f is not assumed to be measurable, the above proof still shows that f satisfies $s_{f,-} \le f \le s_{f,+}$ for two measurable functions $s_{f,\pm}$ which are equal almost everywhere. Hence if we replace the Lebesgue measure by its completion, we can drop this assumption.

Problem 7.24. *Show that for any function $f \in C[a,b]$ we have*

$$\lim_{\|P\|\to 0} L(P,f) = \lim_{\|P\|\to 0} U(P,f).$$

In particular, f is Riemann integrable.

Chapter 8

The Lebesgue spaces L^p

8.1. Functions almost everywhere

We fix some measure space (X, Σ, μ) and define the L^p norm by

$$\|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{1/p}, \qquad 1 \le p, \tag{8.1}$$

and denote by $\mathcal{L}^p(X, d\mu)$ the set of all complex-valued measurable functions for which $\|f\|_p$ is finite. First of all note that $\mathcal{L}^p(X, d\mu)$ is a vector space, since $|f + g|^p \le 2^p \max(|f|, |g|)^p = 2^p \max(|f|^p, |g|^p) \le 2^p(|f|^p + |g|^p)$. Of course our hope is that $\mathcal{L}^p(X, d\mu)$ is a Banach space. However, Lemma 7.23 implies that there is a small technical problem (recall that a property is said to hold almost everywhere if the set where it fails to hold is contained in a set of measure zero):

Lemma 8.1. *Let f be measurable. Then*

$$\int_X |f|^p \, d\mu = 0 \tag{8.2}$$

if and only if $f(x) = 0$ almost everywhere with respect to μ.

Thus $\|f\|_p = 0$ only implies $f(x) = 0$ for almost every x, but not for all! Hence $\|.\|_p$ is not a norm on $\mathcal{L}^p(X, d\mu)$. The way out of this misery is to identify functions which are equal almost everywhere: Let

$$\mathcal{N}(X, d\mu) = \{f | f(x) = 0 \ \mu\text{-almost everywhere}\}. \tag{8.3}$$

Then $\mathcal{N}(X, d\mu)$ is a linear subspace of $\mathcal{L}^p(X, d\mu)$ and we can consider the quotient space

$$L^p(X, d\mu) = \mathcal{L}^p(X, d\mu)/\mathcal{N}(X, d\mu). \tag{8.4}$$

If $d\mu$ is the Lebesgue measure on $X \subseteq \mathbb{R}^n$, we simply write $L^p(X)$. Observe that $\|f\|_p$ is well-defined on $L^p(X, d\mu)$.

Even though the elements of $L^p(X, d\mu)$ are, strictly speaking, equivalence classes of functions, we will still call them functions for notational convenience. However, note that for $f \in L^p(X, d\mu)$ the value $f(x)$ is not well defined (unless there is a continuous representative and different continuous functions are in different equivalence classes, e.g., in the case of Lebesgue measure).

With this modification we are back in business since $L^p(X, d\mu)$ turns out to be a Banach space. We will show this in the following sections. Moreover, note that $L^2(X, d\mu)$ is a Hilbert space with scalar product given by

$$\langle f, g\rangle = \int_X f(x)^* g(x) d\mu(x). \tag{8.5}$$

But before that let us also define $L^\infty(X, d\mu)$. It should be the set of bounded measurable functions $B(X)$ together with the sup norm. The only problem is that if we want to identify functions equal almost everywhere, the supremum is no longer independent of the representative in the equivalence class. The solution is the **essential supremum**

$$\|f\|_\infty = \inf\{C \,|\, \mu(\{x|\, |f(x)| > C\}) = 0\}. \tag{8.6}$$

That is, C is an essential bound if $|f(x)| \le C$ almost everywhere and the essential supremum is the infimum over all essential bounds.

Example. If λ is the Lebesgue measure, then the essential sup of $\chi_{\mathbb{Q}}$ with respect to λ is 0. If Θ is the Dirac measure centered at 0, then the essential sup of $\chi_{\mathbb{Q}}$ with respect to Θ is 1 (since $\chi_{\mathbb{Q}}(0) = 1$, and $x = 0$ is the only point which counts for Θ). ⋄

As before we set

$$L^\infty(X, d\mu) = B(X)/\mathcal{N}(X, d\mu) \tag{8.7}$$

and observe that $\|f\|_\infty$ is independent of the equivalence class.

If you wonder where the ∞ comes from, have a look at Problem 8.2.

Problem 8.1. *Let $\|.\|$ be a seminorm on a vector space X. Show that $N = \{x \in X|\, \|x\| = 0\}$ is a vector space. Show that the quotient space X/N is a normed space with norm $\|x + N\| = \|x\|$.*

Problem 8.2. *Suppose $\mu(X) < \infty$. Show that $L^\infty(X, d\mu) \subseteq L^p(X, d\mu)$ and*

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty, \qquad f \in L^\infty(X, d\mu).$$

Problem 8.3. *Construct a function $f \in L^p(0,1)$ which has a singularity at every rational number in $[0,1]$ (such that the essential supremum is infinite on every open subinterval). (Hint: Start with the function $f_0(x) = |x|^{-\alpha}$ which has a single singularity at 0, then $f_j(x) = f_0(x - x_j)$ has a singularity at x_j.)*

8.2. Jensen ≤ Hölder ≤ Minkowski

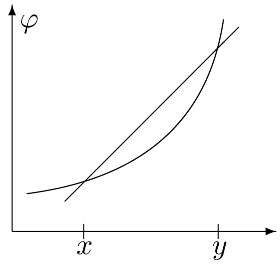
As a preparation for proving that L^p is a Banach space, we will need Hölder's inequality, which plays a central role in the theory of L^p spaces. In particular, it will imply Minkowski's inequality, which is just the triangle inequality for L^p. Our proof is based on Jensen's inequality and emphasizes the connection with convexity. In fact, the triangle inequality just states that a norm is convex:

$$\|\lambda f + (1-\lambda)g\| \le \lambda\|f\| + (1-\lambda)\|g\|, \quad \lambda \in (0,1). \tag{8.8}$$

Recall that a real function φ defined on an open interval (a,b) is called **convex** if

$$\varphi((1-\lambda)x + \lambda y) \le (1-\lambda)\varphi(x) + \lambda\varphi(y), \quad \lambda \in (0,1) \tag{8.9}$$

that is, on (x,y) the graph of $\varphi(x)$ lies below or on the line connecting $(x,\varphi(x))$ and $(y,\varphi(y))$:



If the inequality is strict, then φ is called **strictly convex**. It is not hard to see (use $z = (1-\lambda)x + \lambda y$) that the definition implies

$$\frac{\varphi(z)-\varphi(x)}{z-x} \le \frac{\varphi(y)-\varphi(x)}{y-x} \le \frac{\varphi(y)-\varphi(z)}{y-z}, \quad x < z < y, \tag{8.10}$$

where the inequalities are strict if φ is strictly convex.

Lemma 8.2. *Let $\varphi : (a,b) \to \mathbb{R}$ be convex. Then*

(i) *φ is locally Lipschitz continuous.*

(ii) *The left/right derivatives $\varphi'_\pm(x) = \lim_{\varepsilon\downarrow 0} \frac{\varphi(x\pm\varepsilon)-\varphi(x)}{\pm\varepsilon}$ exist and are monotone nondecreasing. Moreover, φ' exists except at a countable number of points.*

(iii) *For fixed x we have $\varphi(y) \ge \varphi(x) + \alpha(y - x)$ for every α with $\varphi'_-(x) \le \alpha \le \varphi'_+(x)$. The inequality is strict for $y \ne x$ if φ is strictly convex.*

Proof. Abbreviate $D(x,y) = D(y,x) = \frac{\varphi(y)-\varphi(x)}{y-x}$ and observe that (8.10) implies

$$D(x,z) \le D(y,z) \qquad \text{for } x < z < y.$$

Hence $\varphi'_\pm(x)$ exist and we have $\varphi'_-(x) \le \varphi'_+(x) \le \varphi'_-(y) \le \varphi'_+(y)$ for $x < y$. So (ii) follows after observing that a monotone function can have at most a countable number of jumps. Next

$$\varphi'_+(x) \le D(y,x) \le \varphi'_-(y)$$

shows $\varphi(y) \ge \varphi(x) + \varphi'_\pm(x)(y-x)$ if $\pm(y-x) > 0$ and proves (iii). Moreover, $\varphi'_+(z) \le |D(y,x)| \le \varphi'_-(\tilde{z})$ for $z < x, y < \tilde{z}$ proves (i). □

Remark: It is not hard to see that $\varphi \in C^1$ is convex if and only if $\varphi'(x)$ is monotone nondecreasing (e.g., $\varphi'' \ge 0$ if $\varphi \in C^2$).

With these preparations out of the way we can show

Theorem 8.3 (Jensen's inequality)**.** *Let $\varphi : (a,b) \to \mathbb{R}$ be convex ($a = -\infty$ or $b = \infty$ being allowed). Suppose μ is a finite measure satisfying $\mu(X) = 1$ and $f \in \mathcal{L}^1(X, d\mu)$ with $a < f(x) < b$. Then the negative part of $\varphi \circ f$ is integrable and*

$$\varphi(\int_X f\, d\mu) \le \int_X (\varphi \circ f)\, d\mu. \tag{8.11}$$

For $\varphi \ge 0$ nondecreasing and $f \ge 0$ the requirement that f is integrable can be dropped if $\varphi(b)$ is understood as $\lim_{x\to b} \varphi(x)$.

Proof. By (iii) of the previous lemma we have

$$\varphi(f(x)) \ge \varphi(I) + \alpha(f(x) - I), \qquad I = \int_X f\, d\mu \in (a,b).$$

This shows that the negative part of $\varphi \circ f$ is integrable and integrating over X finishes the proof in the case $f \in \mathcal{L}^1$. If $f \ge 0$ we note that for $X_n = \{x \in X | f(x) \le n\}$ the first part implies

$$\varphi(\frac{1}{\mu(X_n)} \int_{X_n} f\, d\mu) \le \frac{1}{\mu(X_n)} \int_{X_n} \varphi(f)\, d\mu.$$

Taking $n \to \infty$ the claim follows from $X_n \nearrow X$ and the monotone convergence theorem. □

Observe that if φ is strictly convex, then equality can only occur if f is constant.

Now we are ready to prove

Theorem 8.4 (Hölder's inequality)**.** *Let p and q be* **dual** *indices; that is,*

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{8.12}$$

with $1 \le p \le \infty$. If $f \in L^p(X, d\mu)$ and $g \in L^q(X, d\mu)$, then $fg \in L^1(X, d\mu)$ and

$$\|f\, g\|_1 \le \|f\|_p \|g\|_q. \tag{8.13}$$

Proof. The case $p = 1$, $q = \infty$ (respectively $p = \infty$, $q = 1$) follows directly from the properties of the integral and hence it remains to consider $1 < p, q < \infty$.

First of all it is no restriction to assume $\|g\|_q = 1$. Let $A = \{x \,|\, |g(x)| > 0\}$, then (note $(1-q)p = -q$)

$$\|f\, g\|_1^p = \Big| \int_A |f|\, |g|^{1-q} |g|^q d\mu \Big|^p \le \int_A (|f|\, |g|^{1-q})^p |g|^q d\mu = \int_A |f|^p d\mu \le \|f\|_p^p,$$

where we have used Jensen's inequality with $\varphi(x) = |x|^p$ applied to the function $h = |f|\, |g|^{1-q}$ and measure $d\nu = |g|^q d\mu$ (note $\nu(X) = \int |g|^q d\mu = \|g\|_q^q = 1$). □

Note that in the special case $p = 2$ we have $q = 2$ and Hölder's inequality reduces to the Cauchy–Schwarz inequality. Moreover, in the case $1 < p < \infty$ the function x^p is strictly convex and equality will only occur if $|f|$ is a multiple of $|g|^{q-1}$ or g is trivial.

As a consequence we also get

Theorem 8.5 (Minkowski's inequality)**.** *Let $f, g \in L^p(X, d\mu)$. Then*

$$\|f + g\|_p \le \|f\|_p + \|g\|_p. \tag{8.14}$$

Proof. Since the cases $p = 1, \infty$ are straightforward, we only consider $1 < p < \infty$. Using $|f + g|^p \le |f|\, |f + g|^{p-1} + |g|\, |f + g|^{p-1}$, we obtain from Hölder's inequality (note $(p-1)q = p$)

$$\begin{aligned} \|f + g\|_p^p &\le \|f\|_p \||f + g|^{p-1}\|_q + \|g\|_p \||f + g|^{p-1}\|_q \\ &= (\|f\|_p + \|g\|_p) \|(f + g)\|_p^{p-1}. \end{aligned}$$

□

This shows that $L^p(X, d\mu)$ is a normed vector space.

Problem 8.4. *Prove*

$$\prod_{k=1}^n x_k^{\alpha_k} \le \sum_{k=1}^n \alpha_k x_k, \qquad \textit{if} \quad \sum_{k=1}^n \alpha_k = 1, \tag{8.15}$$

for $\alpha_k > 0$, $x_k > 0$. (Hint: Take a sum of Dirac-measures and use that the exponential function is convex.)

Problem 8.5. *Show the following generalization of Hölder's inequality:*

$$\|f\, g\|_r \le \|f\|_p \|g\|_q, \qquad \frac{1}{p} + \frac{1}{q} = \frac{1}{r}. \tag{8.16}$$

Problem 8.6. *Show the iterated Hölder's inequality:*

$$\|f_1 \cdots f_m\|_r \le \prod_{j=1}^{m} \|f_j\|_{p_j}, \qquad \frac{1}{p_1} + \cdots + \frac{1}{p_m} = \frac{1}{r}. \tag{8.17}$$

Problem 8.7. *Show that*

$$\|u\|_{p_0} \le \mu(X)^{\frac{1}{p_0} - \frac{1}{p}} \|u\|_p, \qquad 1 \le p_0 \le p.$$

(Hint: Hölder's inequality.)

8.3. Nothing missing in L^p

Finally it remains to show that $L^p(X, d\mu)$ is complete.

Theorem 8.6 (Riesz–Fischer)**.** *The space $L^p(X, d\mu)$, $1 \le p \le \infty$, is a Banach space.*

Proof. We begin with the case $1 \le p < \infty$. Suppose f_n is a Cauchy sequence. It suffices to show that some subsequence converges (show this). Hence we can drop some terms such that

$$\|f_{n+1} - f_n\|_p \le \frac{1}{2^n}.$$

Now consider $g_n = f_n - f_{n-1}$ (set $f_0 = 0$). Then

$$G(x) = \sum_{k=1}^{\infty} |g_k(x)|$$

is in L^p. This follows from

$$\Big\| \sum_{k=1}^{n} |g_k| \Big\|_p \le \sum_{k=1}^{n} \|g_k\|_p \le \|f_1\|_p + 1$$

using the monotone convergence theorem. In particular, $G(x) < \infty$ almost everywhere and the sum

$$\sum_{n=1}^{\infty} g_n(x) = \lim_{n \to \infty} f_n(x)$$

is absolutely convergent for those x. Now let $f(x)$ be this limit. Since $|f(x) - f_n(x)|^p$ converges to zero almost everywhere and $|f(x) - f_n(x)|^p \le (2G(x))^p \in L^1$, dominated convergence shows $\|f - f_n\|_p \to 0$.

In the case $p = \infty$ note that the Cauchy sequence property $|f_n(x) - f_m(x)| < \varepsilon$ for $n, m > N$ holds except for sets $A_{m,n}$ of measure zero. Since

$A = \bigcup_{n,m} A_{n,m}$ is again of measure zero, we see that $f_n(x)$ is a Cauchy sequence for $x \in X \backslash A$. The pointwise limit $f(x) = \lim_{n\to\infty} f_n(x)$, $x \in X\backslash A$, is the required limit in $L^\infty(X, d\mu)$ (show this). □

In particular, in the proof of the last theorem we have seen:

Corollary 8.7. *If $\|f_n - f\|_p \to 0$, then there is a subsequence (of representatives) which converges pointwise almost everywhere.*

Note that the statement is not true in general without passing to a subsequence (Problem 8.8).

It even turns out that L^p is separable.

Lemma 8.8. *Suppose X is a second countable topological space (i.e., it has a countable basis) and μ is an outer regular Borel measure. Then $L^p(X, d\mu)$, $1 \le p < \infty$, is separable. In particular, for every countable basis which is closed under finite unions the set of characteristic functions $\chi_O(x)$ with O in this basis is total.*

Proof. The set of all characteristic functions $\chi_A(x)$ with $A \in \Sigma$ and $\mu(A) < \infty$ is total by construction of the integral (cf. (7.26)). Now our strategy is as follows: Using outer regularity, we can restrict A to open sets and using the existence of a countable base, we can restrict A to open sets from this base.

Fix A. By outer regularity, there is a decreasing sequence of open sets O_n such that $\mu(O_n) \to \mu(A)$. Since $\mu(A) < \infty$, it is no restriction to assume $\mu(O_n) < \infty$, and thus $\mu(O_n\backslash A) = \mu(O_n) - \mu(A) \to 0$. Now dominated convergence implies $\|\chi_A - \chi_{O_n}\|_p \to 0$. Thus the set of all characteristic functions $\chi_O(x)$ with O open and $\mu(O) < \infty$ is total. Finally let $\mathcal{B}$ be a countable basis for the topology. Then, every open set O can be written as $O = \bigcup_{j=1}^\infty \tilde{O}_j$ with $\tilde{O}_j \in \mathcal{B}$. Moreover, by considering the set of all finite unions of elements from $\mathcal{B}$, it is no restriction to assume $\bigcup_{j=1}^n \tilde{O}_j \in \mathcal{B}$. Hence there is an increasing sequence $\tilde{O}_n \nearrow O$ with $\tilde{O}_n \in \mathcal{B}$. By monotone convergence, $\|\chi_O - \chi_{\tilde{O}_n}\|_p \to 0$ and hence the set of all characteristic functions $\chi_{\tilde{O}}$ with $\tilde{O} \in \mathcal{B}$ is total. □

To finish this chapter, let us show that continuous functions are dense in L^p.

Theorem 8.9. *Let X be a locally compact metric space and let μ be a regular Borel measure. Then the set $C_c(X)$ of continuous functions with compact support is dense in $L^p(X, d\mu)$, $1 \le p < \infty$.*

Proof. As in the previous proof the set of all characteristic functions $\chi_K(x)$ with K compact is total (using inner regularity). Hence it suffices to show that $\chi_K(x)$ can be approximated by continuous functions. By outer regularity there is an open set $O \supset K$ such that $\mu(O\backslash K) \le \varepsilon$. By Urysohn's lemma (Lemma 1.16) there is a continuous function $f_\varepsilon : X \to [0,1]$ with compact support which is 1 on K and 0 outside O. Since

$$\int_X |\chi_K - f_\varepsilon|^p d\mu = \int_{O\backslash K} |f_\varepsilon|^p d\mu \le \mu(O\backslash K) \le \varepsilon,$$

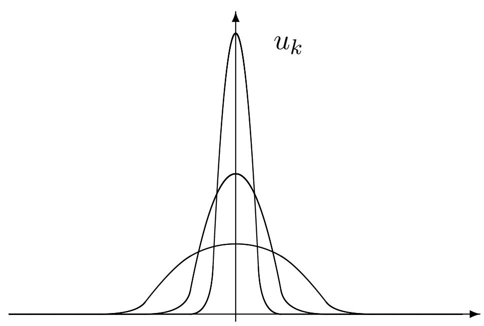
we have $\|f_\varepsilon - \chi_K\| \to 0$ and we are done. □

If X is some subset of $\mathbb{R}^n$, we can do even better. A nonnegative function $u \in C_c^\infty(\mathbb{R}^n)$ is called a **mollifier** if

$$\int_{\mathbb{R}^n} u(x)dx = 1. \tag{8.18}$$

The standard mollifier is $u(x) = \exp(\frac{1}{|x|^2-1})$ for $|x| < 1$ and $u(x) = 0$ otherwise.

If we scale a mollifier according to $u_k(x) = k^n u(k\,x)$ such that its mass is preserved ($\|u_k\|_1 = 1$) and it concentrates more and more around the origin,



we have the following result (Problem 8.9):

Lemma 8.10. *Let u be a mollifier in $\mathbb{R}^n$ and set $u_k(x) = k^n u(k\,x)$. Then for every (uniformly) continuous function $f : \mathbb{R}^n \to \mathbb{C}$ we have that*

$$f_k(x) = \int_{\mathbb{R}^n} u_k(x-y) f(y) dy \tag{8.19}$$

is in $C^\infty(\mathbb{R}^n)$ and converges to f (uniformly).

Now we are ready to prove

Theorem 8.11. *If $X \subseteq \mathbb{R}^n$ is open and μ is a regular Borel measure, then the set $C_c^\infty(X)$ of all smooth functions with compact support is dense in $L^p(X, d\mu)$, $1 \le p < \infty$.*

Proof. By our previous result it suffices to show that every continuous function $f(x)$ with compact support can be approximated by smooth ones. By setting $f(x) = 0$ for $x \notin X$, it is no restriction to assume $X = \mathbb{R}^n$.

Now choose a mollifier u and observe that f_k has compact support (since f has). Moreover, since f has compact support, it is uniformly continuous and $f_k \to f$ uniformly. But this implies $f_k \to f$ in L^p. □

We say that $f \in L^p_{loc}(X)$ if $f \in L^p(K)$ for every compact subset $K \subset X$.

Lemma 8.12. *Suppose $f \in L^1_{loc}(\mathbb{R}^n)$. Then*

$$\int_{\mathbb{R}^n} \varphi(x) f(x) dx = 0, \qquad \forall \varphi \in C_c^\infty(\mathbb{R}^n), \tag{8.20}$$

if and only if $f(x) = 0$ (a.e.).

Proof. Consider the Borel measure $\nu(A) = \int_A f(x)dx$. Let K be some product of compact intervals and consider some sequence $\varphi_m \in C_c^\infty(\mathbb{R}^n)$ which has support in K with $0 \le \varphi_m \le 1$ and $\varphi_m(x) \to \chi_K(x)$ pointwise for a.e. x. Then by dominated convergence $\nu(K) = \lim_{m\to\infty} \int \varphi_m f\, dx = 0$ and thus $\nu = 0$ by Theorem 7.5. □

Problem 8.8. *Find a sequence f_n which converges to 0 in $L^p([0,1], dx)$, $1 \le p < \infty$, but for which $f_n(x) \to 0$ for a.e. $x \in [0,1]$ does not hold. (Hint: Every $n \in \mathbb{N}$ can be uniquely written as $n = 2^m + k$ with $0 \le m$ and $0 \le k < 2^m$. Now consider the characteristic functions of the intervals $I_{m,k} = [k2^{-m}, (k+1)2^{-m}]$.)*

Problem 8.9. *Prove Lemma 8.10. (Hint: To show that f_k is smooth, use Problems 7.15 and 7.16.)*

Problem 8.10 (Convolution). *Show that for $f \in L^1(\mathbb{R}^n)$ and $g \in L^p(\mathbb{R}^n)$, the* **convolution**

$$(g * f)(x) = \int_{\mathbb{R}^n} g(x-y) f(y) dy = \int_{\mathbb{R}^n} g(y) f(x-y) dy \tag{8.21}$$

is in $L^p(\mathbb{R}^n)$ and satisfies **Young's inequality**

$$\|f * g\|_p \le \|f\|_1 \|g\|_p. \tag{8.22}$$

(Hint: Without restriction $\|f\|_1 = 1$. Now use Jensen with $\varphi(x) = |x|^p$, $d\mu = |f|dx$ and Fubini.)

Problem 8.11 (Smoothing). *Suppose $f \in L^p(\mathbb{R}^n)$. Show that f_k defined as in (8.19) converges to f in L^p. (Hint: Use Lemma 4.21 and Young's inequality.)*

Problem 8.12. *Let μ_j be σ-finite regular Borel measures on some second countable topological spaces X_j, $j = 1, 2$. Show that the set of characteristic functions $\chi_{A_1 \times A_2}$ with A_j Borel sets is total in $L^p(X_1 \times X_2, d(\mu_1 \otimes \mu_2))$ for $1 \le p < \infty$. (Hint: Problem 7.17 and Lemma 8.8.)*

8.4. Integral operators

Using Hölder's inequality, we can also identify a class of bounded operators in $L^p(X, d\mu)$.

Lemma 8.13 (Schur criterion). *Let μ, ν be σ-finite measures. Consider $L^p(X, d\mu)$ and $L^p(Y, d\nu)$ and let $\frac{1}{p} + \frac{1}{q} = 1$. Suppose that $K(x, y)$ is measurable and there are nonnegative measurable functions $K_1(x, y)$, $K_2(x, y)$ such that $|K(x, y)| \le K_1(x, y)K_2(x, y)$ and*

$$\|K_1(x, .)\|_{L^q(Y,d\nu)} \le C_1, \qquad \|K_2(., y)\|_{L^p(X,d\mu)} \le C_2 \tag{8.23}$$

for μ-almost every x, respectively, for ν-almost every y. Then the operator $K : L^p(Y, d\nu) \to L^p(X, d\mu)$, defined by

$$(Kf)(x) = \int_Y K(x, y)f(y)d\nu(y), \tag{8.24}$$

for μ-almost every x is bounded with $\|K\| \le C_1C_2$.

Proof. We assume $1 < p < \infty$ for simplicity and leave the cases $p = 1, \infty$ to the reader. Choose $f \in L^p(Y, d\nu)$. By Fubini's theorem $\int_Y |K(x, y)f(y)|d\nu(y)$ is measurable and by Hölder's inequality we have

$$\begin{aligned}\int_Y |K(x, y)f(y)|d\nu(y) &\le \int_Y K_1(x, y)K_2(x, y)|f(y)|d\nu(y)\\ &\le \left(\int_Y K_1(x, y)^q d\nu(y)\right)^{1/q} \left(\int_Y |K_2(x, y)f(y)|^p d\nu(y)\right)^{1/p}\\ &\le C_1 \left(\int_Y |K_2(x, y)f(y)|^p d\nu(y)\right)^{1/p}\end{aligned}$$

for μ a.e. x (if $K_2(x, .)f(.) \notin L^p(X, d\nu)$, the inequality is trivially true). Now take this inequality to the p'th power and integrate with respect to x using Fubini

$$\begin{aligned}\int_X \left(\int_Y |K(x, y)f(y)|d\nu(y)\right)^p d\mu(x) &\le C_1^p \int_X \int_Y |K_2(x, y)f(y)|^p d\nu(y)d\mu(x)\\ &= C_1^p \int_Y \int_X |K_2(x, y)f(y)|^p d\mu(x)d\nu(y) \le C_1^pC_2^p\|f\|_p^p.\end{aligned}$$

Hence $\int_Y |K(x, y)f(y)|d\nu(y) \in L^p(X, d\mu)$ and in particular it is finite for μ-almost every x. Thus $K(x, .)f(.)$ is ν integrable for μ-almost every x and $\int_Y K(x, y)f(y)d\nu(y)$ is measurable. □

Note that the assumptions are for example satisfied if $\|K(x, .)\|_{L^1(Y,d\nu)} \le C$ and $\|K(., y)\|_{L^1(X,d\mu)} \le C$ which follows by choosing $K_1(x, y) = |K(x, y)|^{1/q}$ and $K_2(x, y) = |K(x, y)|^{1/p}$.

Another case of special importance is the case of integral operators

$$(Kf)(x) = \int_X K(x,y) f(y) d\mu(y), \qquad f \in L^2(X, d\mu), \tag{8.25}$$

where $K(x,y) \in L^2(X \times X, d\mu \otimes d\mu)$. Such an operator is called a **Hilbert–Schmidt operator**.

Lemma 8.14. *Let K be a Hilbert–Schmidt operator in $L^2(X, d\mu)$. Then*

$$\int_X \int_X |K(x,y)|^2 d\mu(x) d\mu(y) = \sum_{j \in J} \|K u_j\|^2 \tag{8.26}$$

for every orthonormal basis $\{u_j\}_{j \in J}$ in $L^2(X, d\mu)$.

Proof. Since $K(x,.) \in L^2(X, d\mu)$ for μ-almost every x we infer from Parseval's relation

$$\sum_j \left| \int_X K(x,y) u_j(y) d\mu(y) \right|^2 = \int_X |K(x,y)|^2 d\mu(y)$$

for μ-almost every x and thus

$$\begin{aligned} \sum_j \|K u_j\|^2 &= \sum_j \int_X \left| \int_X K(x,y) u_j(y) d\mu(y) \right|^2 d\mu(x) \\ &= \int_X \sum_j \left| \int_X K(x,y) u_j(y) d\mu(y) \right|^2 d\mu(x) \\ &= \int_X \int_X |K(x,y)|^2 d\mu(x) d\mu(y) \end{aligned}$$

as claimed. □

Hence in combination with Lemma 5.5 this shows that our definition for integral operators agrees with our previous definition from Section 5.2. In particular, this gives us an easy to check test for compactness of an integral operator.

Example. Let $[a,b]$ be some compact interval and suppose $K(x,y)$ is bounded. Then the corresponding integral operator in $L^2(a,b)$ is Hilbert–Schmidt and thus compact. This generalizes Lemma 3.4. ⋄

Problem 8.13. *Show that the integral operator with kernel $K(x,y) = \mathrm{e}^{|x-y|}$ in $L^2(\mathbb{R})$ is bounded.*

Chapter 9

More measure theory

9.1. Decomposition of measures

Let μ, ν be two measures on a measurable space (X, Σ). They are called **mutually singular** (in symbols $\mu \perp \nu$) if they are supported on disjoint sets. That is, there is a measurable set N such that $\mu(N) = 0$ and $\nu(X \backslash N) = 0$.

Example. Let λ be the Lebesgue measure and Θ the Dirac measure (centered at 0). Then $\lambda \perp \Theta$: Just take $N = \{0\}$; then $\lambda(\{0\}) = 0$ and $\Theta(\mathbb{R} \backslash \{0\}) = 0$. ⋄

On the other hand, ν is called **absolutely continuous** with respect to μ (in symbols $\nu \ll \mu$) if $\mu(A) = 0$ implies $\nu(A) = 0$.

Example. The prototypical example is the measure $d\nu = f\, d\mu$ (compare Lemma 7.19). Indeed by Lemma 7.23 $\mu(A) = 0$ implies

$$\nu(A) = \int_A f\, d\mu = 0 \tag{9.1}$$

and shows that ν is absolutely continuous with respect to μ. In fact, we will show below that every absolutely continuous measure is of this form. ⋄

The two main results will follow as simple consequence of the following result:

Theorem 9.1. *Let μ, ν be σ-finite measures. Then there exists a nonnegative function f and a set N of μ measure zero, such that*

$$\nu(A) = \nu(A \cap N) + \int_A f\, d\mu. \tag{9.2}$$

Proof. We first assume μ, ν to be finite measures. Let $\alpha = \mu + \nu$ and consider the Hilbert space $L^2(X, d\alpha)$. Then

$$\ell(h) = \int_X h\, d\nu$$

is a bounded linear functional on $L^2(X, d\alpha)$ by Cauchy–Schwarz:

$$|\ell(h)|^2 = \left| \int_X 1 \cdot h\, d\nu \right|^2 \le \left(\int |1|^2\, d\nu \right) \left(\int |h|^2 d\nu \right)$$
$$\le \nu(X) \left(\int |h|^2 d\alpha \right) = \nu(X) \|h\|^2.$$

Hence by the Riesz lemma (Theorem 2.10) there exists a $g \in L^2(X, d\alpha)$ such that

$$\ell(h) = \int_X hg\, d\alpha.$$

By construction

$$\nu(A) = \int \chi_A\, d\nu = \int \chi_A g\, d\alpha = \int_A g\, d\alpha. \tag{9.3}$$

In particular, g must be positive a.e. (take A the set where g is negative). Furthermore, let $N = \{x | g(x) \ge 1\}$. Then

$$\nu(N) = \int_N g\, d\alpha \ge \alpha(N) = \mu(N) + \nu(N),$$

which shows $\mu(N) = 0$. Now set

$$f = \frac{g}{1-g} \chi_{N'}, \qquad N' = X \backslash N.$$

Then, since (9.3) implies $d\nu = g\, d\alpha$, respectively, $d\mu = (1-g) d\alpha$, we have

$$\int_A f d\mu = \int \chi_A \frac{g}{1-g} \chi_{N'}\, d\mu = \int \chi_{A\cap N'} g\, d\alpha = \nu(A \cap N')$$

as desired.

To see the σ-finite case, observe that $Y_n \nearrow X$, $\mu(Y_n) < \infty$ and $Z_n \nearrow X$, $\nu(Z_n) < \infty$ implies $X_n = Y_n \cap Z_n \nearrow X$ and $\alpha(X_n) < \infty$. Hence when restricted to X_n, we have sets N_n and functions f_n. Now take $N = \bigcup N_n$ and choose f such that $f|_{X_n} = f_n$ (this is possible since $f_{n+1}|_{X_n} = f_n$ a.e.). Then $\mu(N) = 0$ and

$$\nu(A \cap N') = \lim_{n\to\infty} \nu(A \cap (X_n \backslash N)) = \lim_{n\to\infty} \int_{A\cap X_n} f\, d\mu = \int_A f\, d\mu,$$

which finishes the proof. □

Now the anticipated results follow with no effort:

Theorem 9.2 (Radon–Nikodym)**.** *Let μ, ν be two σ-finite measures on a measurable space (X, Σ). Then ν is absolutely continuous with respect to μ if and only if there is a positive measurable function f such that*

$$\nu(A) = \int_A f\, d\mu \tag{9.4}$$

for every $A \in \Sigma$. The function f is determined uniquely a.e. with respect to μ and is called the **Radon–Nikodym derivative** $\frac{d\nu}{d\mu}$ *of ν with respect to μ.*

Proof. Just observe that in this case $\nu(A \cap N) = 0$ for every A. Uniqueness will be shown in the next theorem. □

Theorem 9.3 (Lebesgue decomposition)**.** *Let μ, ν be two σ-finite measures on a measurable space (X, Σ). Then ν can be uniquely decomposed as $\nu = \nu_{ac} + \nu_{sing}$, where ν_{ac} and ν_{sing} are mutually singular and ν_{ac} is absolutely continuous with respect to μ.*

Proof. Taking $\nu_{sing}(A) = \nu(A \cap N)$ and $d\nu_{ac} = f\, d\mu$ from the previous theorem, there is at least one such decomposition. To show uniqueness assume there is another one, $\nu = \tilde{\nu}_{ac} + \tilde{\nu}_{sing}$, and let $\tilde{N}$ be such that $\mu(\tilde{N}) = 0$ and $\tilde{\nu}_{sing}(\tilde{N}') = 0$. Then $\nu_{sing}(A) - \tilde{\nu}_{sing}(A) = \int_A (\tilde{f} - f) d\mu$. In particular, $\int_{A \cap N' \cap \tilde{N}'} (\tilde{f} - f) d\mu = 0$ and hence $\tilde{f} = f$ a.e. away from $N \cup \tilde{N}$. Since $\mu(N \cup \tilde{N}) = 0$, we have $\tilde{f} = f$ a.e. and hence $\tilde{\nu}_{ac} = \nu_{ac}$ as well as $\tilde{\nu}_{sing} = \nu - \tilde{\nu}_{ac} = \nu - \nu_{ac} = \nu_{sing}$. □

Problem 9.1. *Let μ be a Borel measure on $\mathfrak{B}$ and suppose its distribution function $\mu(x)$ is continuously differentiable. Show that the Radon–Nikodym derivative equals the ordinary derivative $\mu'(x)$.*

Problem 9.2. *Suppose μ is a Borel measure on $\mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ is continuous. Show that $f_\star \mu$ is absolutely continuous if μ is. (Hint: Problem 7.11)*

Problem 9.3. *Suppose μ and ν are inner regular measures. Show that $\nu \ll \mu$ if and only if $\mu(C) = 0$ implies $\nu(C) = 0$ for every compact set.*

Problem 9.4. *Suppose $\nu(A) \le C\mu(A)$ for all $A \in \Sigma$. Then $d\nu = f\, d\mu$ with $0 \le f \le C$ a.e.*

Problem 9.5. *Let $d\nu = f\, d\mu$. Suppose $f > 0$ a.e. with respect to μ. Then $\mu \ll \nu$ and $d\mu = f^{-1} d\nu$.*

Problem 9.6 (Chain rule)**.** *Show that $\nu \ll \mu$ is a transitive relation. In particular, if $\omega \ll \nu \ll \mu$, show that*

$$\frac{d\omega}{d\mu} = \frac{d\omega}{d\nu} \frac{d\nu}{d\mu}.$$

Problem 9.7. *Suppose $\nu \ll \mu$. Show that for every measure ω we have*

$$\frac{d\omega}{d\mu} d\mu = \frac{d\omega}{d\nu} d\nu + d\zeta,$$

where ζ is a positive measure (depending on ω) which is singular with respect to ν. Show that $\zeta = 0$ if and only if $\mu \ll \nu$.

9.2. Derivatives of measures

If μ is a Borel measure on $\mathfrak{B}$ and its distribution function $\mu(x)$ is continuously differentiable, then the Radon–Nikodym derivative is just the ordinary derivative $\mu'(x)$ (Problem 9.1). Our aim in this section is to generalize this result to arbitrary Borel measures on $\mathfrak{B}^n$.

We call

$$(D\mu)(x) = \lim_{\varepsilon \downarrow 0} \frac{\mu(B_\varepsilon(x))}{|B_\varepsilon(x)|} \tag{9.5}$$

the derivative of μ at $x \in \mathbb{R}^n$ provided the above limit exists. (Here $B_r(x) \subset \mathbb{R}^n$ is a ball of radius r centered at $x \in \mathbb{R}^n$ and $|A|$ denotes the Lebesgue measure of $A \in \mathfrak{B}^n$.)

Note that for a Borel measure on $\mathfrak{B}$, $(D\mu)(x)$ exists if $\mu(x)$ (as defined in (7.3)) is differentiable at x and $(D\mu)(x) = \mu'(x)$ in this case.

To compute the derivative of μ, we introduce the **upper** and **lower derivative**,

$$(\overline{D}\mu)(x) = \limsup_{\varepsilon \downarrow 0} \frac{\mu(B_\varepsilon(x))}{|B_\varepsilon(x)|} \quad \text{and} \quad (\underline{D}\mu)(x) = \liminf_{\varepsilon \downarrow 0} \frac{\mu(B_\varepsilon(x))}{|B_\varepsilon(x)|}. \tag{9.6}$$

Clearly μ is differentiable at x if $(\overline{D}\mu)(x) = (\underline{D}\mu)(x) < \infty$. First of all note that they are measurable: In fact, this follows from

$$(\overline{D}\mu)(x) = \lim_{n \to \infty} \sup_{0 < \varepsilon < 1/n} \frac{\mu(B_\varepsilon(x))}{|B_\varepsilon(x)|} \tag{9.7}$$

since the supremum on the right-hand side is lower semicontinuous with respect to x (cf. Problem 7.9) as $x \mapsto \mu(B_\varepsilon(x))$ is continuous (show this!). Similarly for $(\underline{D}\mu)(x)$.

Next, the following geometric fact of $\mathbb{R}^n$ will be needed.

Lemma 9.4 (Wiener covering lemma)**.** *Given open balls $B_1, \dots, B_m$ in $\mathbb{R}^n$, there is a subset of disjoint balls $B_{j_1}, \dots, B_{j_k}$ such that*

$$\left| \bigcup_{i=1}^m B_i \right| \le 3^n \sum_{i=1}^k |B_{j_i}|. \tag{9.8}$$

Proof. Assume that the balls B_j are ordered by decreasing radius. Start with $B_{j_1} = B_1 = B_{r_1}(x_1)$ and remove all balls from our list which intersect B_{j_1}. Observe that the removed balls are all contained in $3B_1 = B_{3r_1}(x_1)$. Proceeding like this, we obtain B_{j_1}, B_{j_2}, ..., such that

$$\bigcup_j B_j \subseteq \bigcup_k 3B_{j_k}$$

and the claim follows since $|3B| = 3^n |B|$. □

Now we can show

Lemma 9.5. *Let $\alpha > 0$. For every Borel set A we have*

$$|\{x \in A \,|\, (\overline{D}\mu)(x) > \alpha\}| \le 3^n \frac{\mu(A)}{\alpha} \tag{9.9}$$

and

$$|\{x \in A \,|\, (\overline{D}\mu)(x) > 0\}| = 0, \ \text{whenever } \mu(A) = 0. \tag{9.10}$$

Proof. Let $A_\alpha = \{x \in A | (\overline{D}\mu)(x) > \alpha\}$. We will show

$$|A_\alpha| \le 3^n \frac{\mu(O)}{\alpha}$$

for every open set O with $A \subseteq O$. The first claim then follows from outer regularity of μ and the Lebesgue measure.

Given fixed K, O, for every $x \in K$ there is some r_x such that $B_{r_x}(x) \subseteq O$ and $|B_{r_x}(x)| < \alpha^{-1}\mu(B_{r_x}(x))$. Since K is compact, we can choose a finite subcover of K. Moreover, by Lemma 9.4 we can refine our set of balls such that

$$|K| \le 3^n \sum_{i=1}^k |B_{r_i}(x_i)| < \frac{3^n}{\alpha} \sum_{i=1}^k \mu(B_{r_i}(x_i)) \le 3^n \frac{\mu(O)}{\alpha}.$$

To see the second claim, observe that

$$\{x \in A \,|\, (\overline{D}\mu)(x) > 0\} = \bigcup_{j=1}^\infty \{x \in A \,|\, (\overline{D}\mu)(x) > \frac{1}{j}\}$$

and by the first part $|\{x \in A \,|\, (\overline{D}\mu)(x) > \frac{1}{j}\}| = 0$ for every j if $\mu(A) = 0$. □

Theorem 9.6 (Lebesgue)**.** *Let f be (locally) integrable, then for a.e. $x \in \mathbb{R}^n$ we have*

$$\lim_{r \downarrow 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)| dy = 0. \tag{9.11}$$

Proof. Decompose f as $f = g + h$, where g is continuous and $\|h\|_1 < \varepsilon$ (Theorem 8.9) and abbreviate

$$D_r(f)(x) = \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)| dy.$$

Then, since $\lim D_r(g)(x) = 0$ (for every x) and $D_r(f) \le D_r(g) + D_r(h)$, we have

$$\limsup_{r\downarrow 0} D_r(f)(x) \le \limsup_{r\downarrow 0} D_r(h)(x) \le (\overline{D}\mu)(x) + |h(x)|,$$

where $d\mu = |h|dx$. This implies

$$\{x \,|\, \limsup_{r\downarrow 0} D_r(f)(x) \ge 2\alpha\} \subseteq \{x|(\overline{D}\mu)(x) \ge \alpha\} \cup \{x \,|\, |h(x)| \ge \alpha\}$$

and using the first part of Lemma 9.5 plus $|\{x \,|\, |h(x)| \ge \alpha\}| \le \alpha^{-1}\|h\|_1$ (Problem 9.11), we see

$$|\{x \,|\, \limsup_{r\downarrow 0} D_r(f)(x) \ge 2\alpha\}| \le (3^n + 1)\frac{\varepsilon}{\alpha}.$$

Since ε is arbitrary, the Lebesgue measure of this set must be zero for every α. That is, the set where the lim sup is positive has Lebesgue measure zero. □

The points where (9.11) holds are called **Lebesgue points** of f.

Note that the balls can be replaced by more general sets: A sequence of sets $A_j(x)$ is said to shrink to x nicely if there are balls $B_{r_j}(x)$ with $r_j \to 0$ and a constant $\varepsilon > 0$ such that $A_j(x) \subseteq B_{r_j}(x)$ and $|A_j| \ge \varepsilon |B_{r_j}(x)|$. For example $A_j(x)$ could be some balls or cubes (not necessarily containing x). However, the portion of $B_{r_j}(x)$ which they occupy must not go to zero! For example the rectangles $(0, \frac{1}{j}) \times (0, \frac{2}{j}) \subset \mathbb{R}^2$ do shrink nicely to 0, but the rectangles $(0, \frac{1}{j}) \times (0, \frac{2}{j^2})$ do not.

Lemma 9.7. *Let f be (locally) integrable. Then at every Lebesgue point we have*

$$f(x) = \lim_{j\to\infty} \frac{1}{|A_j(x)|} \int_{A_j(x)} f(y)dy \tag{9.12}$$

whenever $A_j(x)$ shrinks to x nicely.

Proof. Let x be a Lebesgue point and choose some nicely shrinking sets $A_j(x)$ with corresponding $B_{r_j}(x)$ and ε. Then

$$\frac{1}{|A_j(x)|} \int_{A_j(x)} |f(y) - f(x)|dy \le \frac{1}{\varepsilon |B_{r_j}(x)|} \int_{B_{r_j}(x)} |f(y) - f(x)|dy$$

and the claim follows. □

Corollary 9.8. *Let μ be a Borel measure on $\mathbb{R}$ which is absolutely continuous with respect to Lebesgue measure. Then its distribution function is differentiable a.e. and $d\mu(x) = \mu'(x)dx$.*

Proof. By assumption $d\mu(x) = f(x)dx$ for some locally integrable function f. In particular, the distribution function $\mu(x) = \int_0^x f(y)dy$ is continuous. Moreover, since the sets $(x, x+r)$ shrink nicely to x as $r \to 0$, Lemma 9.7 implies

$$\lim_{r\to 0} \frac{\mu((x,x+r))}{r} = \lim_{r\to 0} \frac{\mu(x+r)-\mu(x)}{r} = f(x)$$

at every Lebesgue point of f. Since the same is true for the sets $(x-r,x)$, $\mu(x)$ is differentiable at every Lebesgue point and $\mu'(x) = f(x)$. □

As another consequence we obtain

Theorem 9.9. *Let μ be a Borel measure on $\mathbb{R}^n$. The derivative $D\mu$ exists a.e. with respect to Lebesgue measure and equals the Radon–Nikodym derivative of the absolutely continuous part of μ with respect to Lebesgue measure; that is,*

$$\mu_{ac}(A) = \int_A (D\mu)(x)dx. \tag{9.13}$$

Proof. If $d\mu = f\,dx$ is absolutely continuous with respect to Lebesgue measure, then $(D\mu)(x) = f(x)$ at every Lebesgue point of f by Lemma 9.7 and the claim follows from Theorem 9.6. To see the general case, use the Lebesgue decomposition of μ and let N be a support for the singular part with $|N| = 0$. Then $(\overline{D}\mu_{sing})(x) = 0$ for a.e. $x \in \mathbb{R}^n \backslash N$ by the second part of Lemma 9.5. □

In particular, μ is singular with respect to Lebesgue measure if and only if $D\mu = 0$ a.e. with respect to Lebesgue measure.

Using the upper and lower derivatives, we can also give supports for the absolutely and singularly continuous parts.

Theorem 9.10. *The set $\{x | (\underline{D}\mu)(x) = \infty\}$ is a support for the singular and $\{x | 0 < (D\mu)(x) < \infty\}$ is a support for the absolutely continuous part.*

Proof. First suppose μ is purely singular. Let us show that the set $A_k = \{x \,|\, (\underline{D}\mu)(x) < k\}$ satisfies $\mu(A_k) = 0$ for every $k \in \mathbb{N}$.

Let $K \subset A_k$ be compact, and let $V_j \supset K$ be some open set such that $|V_j \backslash K| \le \frac{1}{j}$. For every $x \in K$ there is some $\varepsilon = \varepsilon(x)$ such that $B_\varepsilon(x) \subseteq V_j$ and $\mu(B_\varepsilon(x)) \le k|B_\varepsilon(x)|$. By compactness, finitely many of these balls cover K and hence

$$\mu(K) \le \sum_i \mu(B_{\varepsilon_i}(x_i)) \le k \sum_i |B_{\varepsilon_i}(x_i)|.$$

Selecting disjoint balls as in Lemma 9.4 further shows

$$\mu(K) \le k3^n \sum_\ell |B_{\varepsilon_{i_\ell}}(x_{i_\ell})| \le k3^n |V_j|.$$

Letting $j \to \infty$, we see $\mu(K) \leq k3^n|K|$ and by regularity we even have $\mu(A) \leq k3^n|A|$ for every $A \subseteq A_k$. Hence μ is absolutely continuous on A_k and since we assumed μ to be singular, we must have $\mu(A_k) = 0$.

Thus $(\underline{D}\mu_{sing})(x) = \infty$ for a.e. x with respect to μ_{sing} and we are done. □

Finally, we note that these supports are minimal. Here a support M of some measure μ is called a **minimal support** (it is sometimes also called an **essential support**) if every subset $M_0 \subseteq M$ which does not support μ (i.e., $\mu(M_0) = 0$) has Lebesgue measure zero.

Lemma 9.11. *The set $M_{ac} = \{x | 0 < (D\mu)(x) < \infty\}$ is a minimal support for μ_{ac}.*

Proof. Suppose $M_0 \subseteq M_{ac}$ and $\mu_{ac}(M_0) = 0$. Set $M_\varepsilon = \{x \in M_0 | \varepsilon < (D\mu)(x)\}$ for $\varepsilon > 0$. Then $M_\varepsilon \nearrow M_0$ and

$$|M_\varepsilon| = \int_{M_\varepsilon} dx \leq \frac{1}{\varepsilon} \int_{M_\varepsilon} (D\mu)(x) dx = \frac{1}{\varepsilon} \mu_{ac}(M_\varepsilon) \leq \frac{1}{\varepsilon} \mu_{ac}(M_0) = 0$$

shows $|M_0| = \lim_{\varepsilon \downarrow 0} |M_\varepsilon| = 0$. □

Note that the set $M = \{x | 0 < (D\mu)(x)\}$ is a minimal support of μ.

Example. The **Cantor function** is constructed as follows: Take the sets C_n used in the construction of the Cantor set C: C_n is the union of 2^n closed intervals with $2^n - 1$ open gaps in between. Set f_n equal to $j/2^n$ on the j'th gap of C_n and extend it to $[0,1]$ by linear interpolation. Note that, since we are creating precisely one new gap between every old gap when going from C_n to C_{n+1}, the value of f_{n+1} is the same as the value of f_n on the gaps of C_n. In particular, $\|f_n - f_m\|_\infty \leq 2^{-\min(n,m)}$ and hence we can define the Cantor function as $f = \lim_{n\to\infty} f_n$. By construction f is a continuous function which is constant on every subinterval of $[0,1]\backslash C$. Since C is of Lebesgue measure zero, this set is of full Lebesgue measure and hence $f' = 0$ a.e. in $[0,1]$. In particular, the corresponding measure, the **Cantor measure**, is supported on C and is purely singular with respect to Lebesgue measure. ⋄

Problem 9.8. *Show that $M = \{x | 0 < (D\mu)(x)\}$ is a minimal support of μ.*

Problem 9.9. *Suppose $\overline{D}\mu \leq \alpha$. Show that $d\mu = f\, dx$ with $\|f\|_\infty \leq \alpha$.*

Problem 9.10 (Hardy–Littlewood maximal function)**.** *For $f \in L^1(\mathbb{R}^n)$ consider*

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

Show that $M(f)(.)$ is lower semicontinuous. Furthermore, show

$$|\{x \in A | M(f)(x) > \alpha\}| \le \frac{3^n}{\alpha} \int_A |f(y)| dy.$$

Problem 9.11 (Chebyshev's inequality). *For $f \in L^1(\mathbb{R}^n)$ show*

$$|\{x \in A| \, |f(x)| \ge \alpha\}| \le \frac{1}{\alpha} \int_A |f(x)| dx.$$

9.3. Complex measures

Let (X, Σ) be some measurable space. A map $\nu : \Sigma \to \mathbb{C}$ is called a **complex measure** if

$$\nu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \nu(A_n), \qquad A_n \cap A_m = \emptyset, \quad n \ne m. \tag{9.14}$$

Choosing $A_n = \emptyset$ for all n in (9.14) shows $\nu(\emptyset) = 0$.

Note that a positive measure is a complex measure only if it is finite (the value ∞ is not allowed for complex measures). Moreover, the definition implies that the sum is independent of the order of the sets A_j, that is, it converges unconditionally and thus absolutely by the Riemann series theorem.

Example. Let μ be a positive measure. For every $f \in L^1(X, d\mu)$ we have that $f\,d\mu$ is a complex measure (compare the proof of Lemma 7.19 and use dominated in place of monotone convergence). In fact, we will show that every complex measure is of this form. ⋄

Example. Let ν_1, ν_2 be two complex measures and α_1, α_2 two complex numbers. Then $\alpha_1\nu_1 + \alpha_2\nu_2$ is again a complex measure. Clearly we can extend this to any finite linear combination of complex measures. ⋄

Given a complex measure ν it seems natural to consider the set function $A \mapsto |\nu(A)|$. However, considering the simple example $d\nu(x) = \operatorname{sign}(x)dx$ on $X = [-1, 1]$ one sees that this set function is not additive and this simple approach does not provide a positive measure associated with ν. To this end we introduce the **total variation** of a measure defined as

$$|\nu|(A) = \sup \Big\{ \sum_{k=1}^{n} |\nu(A_k)| \Big| A_k \in \Sigma \text{ disjoint, } A = \bigcup_{k=1}^{n} A_k \Big\}. \tag{9.15}$$

Note that by construction we have

$$|\nu(A)| \le |\nu|(A). \tag{9.16}$$

Theorem 9.12. *The total variation $|\nu|$ of a complex measure ν is a finite positive measure.*

Proof. We begin by showing that $|\nu|$ is a positive measure. Suppose $A = \bigcup_{n=1}^{\infty} A_n$. We need to show $|\nu|(A) = \sum_{n=1}^{\infty} |\nu|(A_n)$ for any partition of A into disjoint sets A_n. If $|\nu|(A_n) = \infty$ for some n it is not hard to see that $|\nu|(A) = \infty$ and hence we can assume $|\nu|(A_n) < \infty$ for all n.

Let $\varepsilon > 0$ be fixed and for each A_n choose a disjoint partition $B_{n,k}$ such that

$$|\nu|(A_n) \le \sum_{k=1}^{m} |\nu(B_{n,k})| + \frac{\varepsilon}{2^n}.$$

Then

$$\sum_{n=1}^{\infty} |\nu|(A_n) \le \sum_{n,k} |\nu(B_{n,k})| + \varepsilon \le |\nu|(A) + \varepsilon$$

since $\bigcup_{n,k} B_{n,k} = A$. Since ε was arbitrary this shows $|\nu|(A) \ge \sum_{n=1}^{\infty} |\nu|(A_n)$.

Conversely, given a finite partition B_k of A, then

$$\begin{aligned}\sum_{k=1}^{m} |\nu(B_k)| &= \sum_{k=1}^{m} \Big| \sum_{n=1}^{\infty} \nu(B_k \cap A_n) \Big| \le \sum_{k=1}^{m} \sum_{n=1}^{\infty} |\nu(B_k \cap A_n)| \\ &= \sum_{n=1}^{\infty} \sum_{k=1}^{m} |\nu(B_k \cap A_n)| \le \sum_{n=1}^{\infty} |\nu|(A_n).\end{aligned}$$

Taking the supremum over all partitions B_k shows $|\nu|(A) \le \sum_{n=1}^{\infty} |\nu|(A_n)$.

Hence $|\nu|$ is a positive measure and it remains to show that it is finite. Splitting ν into its real and imaginary part, it is no restriction to assume that ν is real-valued since $|\nu|(A) \le |\mathrm{Re}(\nu)|(A) + |\mathrm{Im}(\nu)|(A)$.

The idea is as follows: Suppose we can split any given set A with $|\nu|(A) = \infty$ into two subsets B and $A\backslash B$ such that $|\nu(B)| \ge 1$ and $|\nu|(A\backslash B) = \infty$. Then we can construct a sequence B_n of disjoint sets with $|\nu(B_n)| \ge 1$ for which

$$\sum_{n=1}^{\infty} \nu(B_n)$$

diverges (the terms of a convergent series must converge to zero). But σ-additivity requires that the sum converges to $\nu(\bigcup_n B_n)$, a contradiction.

It remains to show existence of this splitting. Let A with $|\nu|(A) = \infty$ be given. Then there are disjoint sets A_j such that

$$\sum_{j=1}^{n} |\nu(A_j)| \ge 2 + |\nu(A)|.$$

Now let $A_+ = \bigcup\{A_j | \nu(A_j) \ge 0\}$ and $A_- = A\backslash A_+ = \bigcup\{A_j | \nu(A_j) < 0\}$. Then for both of them we have $|\nu(A_\pm)| \ge 1$ and by $|\nu|(A) = |\nu|(A_+) + |\nu|(A_-)$ either A_+ or A_- must have infinite $|\nu|$ measure. □

Note that this implies that every complex measure ν can be written as a linear combination of four positive measures. In fact, first we can split ν into its real and imaginary part

$$\nu = \nu_r + \mathrm{i}\nu_i, \qquad \nu_r(A) = \mathrm{Re}(\nu(A)),\ \nu_i(A) = \mathrm{Im}(\nu(A)). \tag{9.17}$$

Second we can split every real (also called **signed**) measure according to

$$\nu = \nu_+ - \nu_-, \qquad \nu_\pm(A) = \frac{|\nu|(A) \pm \nu(A)}{2}. \tag{9.18}$$

By (9.16) both ν_- and ν_+ are positive measures. This splitting is also known as **Hahn decomposition** of a signed measure.

Of course such a decomposition of a signed measure is not unique (we can always add a positive measure to both parts), however, the Hahn decomposition is unique in the sense that it is the smallest possible decomposition.

Lemma 9.13. *Let ν be a complex measure and μ a positive measure satisfying $|\nu(A)| \le \mu(A)$ for all measurable sets A. Then $\mu \ge |\nu|$. (Here $\nu \le \mu$ has to be understood as $\nu(A) \le \mu(A)$ for every measurable set A.)*

Furthermore, let ν be a signed measure and $\nu = \tilde{\nu}_+ - \tilde{\nu}_-$ a decomposition into positive measures. Then $\tilde{\nu}_\pm \ge \nu_\pm$, where $\nu_\pm$ is the Hahn decomposition.

Proof. It suffices to prove the first part since the second is a special case. But for every measurable set A and a corresponding finite partition A_k we have $\sum_k |\nu(A_k)| \le \sum \mu(A_k) = \mu(A)$ implying $|\nu|(A) \le \mu(A)$. □

Moreover, we also have:

Theorem 9.14. *The set of all complex measures $\mathcal{M}(X)$ together with the norm $\|\nu\| = |\nu|(X)$ is a Banach space.*

Proof. Clearly $\mathcal{M}(X)$ is a vector space and it is straightforward to check that $|\nu|(X)$ is a norm. Hence it remains to show that every Cauchy sequence ν_k has a limit.

First of all, by $|\nu_k(A) - \nu_j(A)| = |(\nu_k - \nu_j)(A)| \le |\nu_k - \nu_j|(A) \le \|\nu_k - \nu_j\|$, we see that $\nu_k(A)$ is a Cauchy sequence in $\mathbb{C}$ for every $A \in \Sigma$ and we can define

$$\nu(A) = \lim_{k\to\infty} \nu_k(A).$$

Moreover, $C_j = \sup_{k \ge j} \|\nu_k - \nu_j\| \to 0$ as $j \to \infty$ and we have

$$|\nu_j(A) - \nu(A)| \le C_j.$$

Next we show that ν satisfies (9.14). Let A_m be given disjoint sets and set $\tilde{A}_n = \bigcup_{m=1}^n A_m$, $A = \bigcup_{m=1}^\infty A_m$. Since we can interchange limits with finite sums, (9.14) holds for finitely many sets. Hence it remains to show $\nu(\tilde{A}_n) \to \nu(A)$. This follows from

$$\begin{aligned}|\nu(\tilde{A}_n) - \nu(A)| &\le |\nu(\tilde{A}_n) - \nu_k(\tilde{A}_n)| + |\nu_k(\tilde{A}_n) - \nu_k(A)| + |\nu_k(A) - \nu(A)| \\ &\le 2C_k + |\nu_k(\tilde{A}_n) - \nu_k(A)|.\end{aligned}$$

Finally, $\nu_k \to \nu$ since $|\nu_k(A) - \nu(A)| \le C_k$ implies $\|\nu_k - \nu\| \le 4C_k$ (Problem 9.16). □

If μ is a positive and ν a complex measure we say that ν is absolutely continuous with respect to μ if $\mu(A) = 0$ implies $\nu(A) = 0$.

Lemma 9.15. *If μ is a positive and ν a complex measure then $\nu \ll \mu$ if and only if $|\nu| \ll \mu$.*

Proof. If $\nu \ll \mu$, then $\mu(A) = 0$ implies $\mu(B) = 0$ for every $B \subseteq A$ and hence $|\nu|(A) = 0$. Conversely, if $|\nu| \ll \mu$, then $\mu(A) = 0$ implies $|\nu(A)| \le |\nu|(A) = 0$. □

Now we can prove the complex version of the Radon–Nikodym theorem:

Theorem 9.16 (Complex Radon–Nikodym)**.** *Let (X, Σ) be a measurable space, μ a positive σ-finite measure and ν a complex measure which is absolutely continuous with respect to μ. Then there is a unique $f \in L^1(X, d\mu)$ such that*

$$\nu(A) = \int_A f\, d\mu. \tag{9.19}$$

Proof. By treating the real and imaginary part separately it is no restriction to assume that ν is real-valued. Let $\nu = \nu_+ - \nu_-$ be its Hahn decomposition. Then both ν_+ and ν_- are absolutely continuous with respect to μ and by the Radon–Nikodym theorem there are functions $f_\pm$ such that $d\nu_\pm = f_\pm d\mu$. By construction

$$\int_X f_\pm d\mu = \nu_\pm(X) \le |\nu|(X) < \infty,$$

which shows $f = f_+ - f_- \in L^1(X, d\mu)$. Moreover, $d\nu = d\nu_+ - d\nu_- = f\, d\mu$ as required. □

In this case the total variation of $d\nu = f\, d\mu$ is just $d|\nu| = |f| d\mu$:

Lemma 9.17. *Suppose $d\nu = f\, d\mu$, where μ is a positive measure and $f \in L^1(X, d\mu)$. Then*

$$|\nu|(A) = \int_A |f| d\mu. \tag{9.20}$$

Proof. If A_n are disjoint sets and $A = \bigcup_n A_n$ we have

$$\sum_n |\nu(A_n)| = \sum_n \Big| \int_{A_n} f\, d\mu \Big| \le \sum_n \int_{A_n} |f| d\mu = \int_A |f| d\mu.$$

Hence $|\nu|(A) \le \int_A |f| d\mu$. To show the converse define

$$A_k^n = \{x | \frac{k-1}{n} < \frac{\arg(f(x)) + \pi}{2\pi} \le \frac{k}{n}\}, \quad 1 \le k \le n.$$

Then the simple functions

$$s_n(x) = \sum_{k=1}^n \mathrm{e}^{-2\pi \mathrm{i} \frac{k-1}{n}} \chi_{A_k^n}(x)$$

converge to $\mathrm{sign}(f(x)^*)$ pointwise and hence

$$\lim_{n\to\infty} \int_A s_n f\, d\mu = \int_A |f| d\mu$$

by dominated convergence. Moreover,

$$\Big| \int_A s_n f\, d\mu \Big| \le \sum_{k=1}^n \Big| \int_{A_k^n} s_n f\, d\mu \Big| = \sum_{k=1}^n |\nu(A_k^n)| \le |\nu|(A)$$

shows $\int_A |f| d\mu \le |\nu|(A)$. □

As a consequence we obtain (Problem 9.12):

Corollary 9.18. *If ν is a complex measure, then $d\nu = h\, d|\nu|$, where $|h| = 1$. Moreover, if ν is real-valued, then $d\nu_\pm = \chi_{A_\pm} d|\nu|$, where $A_\pm = h^{-1}(\{\pm 1\})$.*

In particular, note that

$$\Big| \int_A f\, d\nu \Big| \le \|f\|_\infty |\nu|(A). \tag{9.21}$$

Finally, there is an interesting equivalent definition of absolute continuity:

Lemma 9.19. *If μ is a positive and ν a complex measure then $\nu \ll \mu$ if and only if for every $\varepsilon > 0$ there is a corresponding $\delta > 0$ such that*

$$\mu(A) < \delta \quad \Rightarrow \quad |\nu(A)| < \varepsilon, \qquad \forall A \in \Sigma. \tag{9.22}$$

Proof. Suppose $\nu \ll \mu$ implying that it is of the from (9.19). Let $X_n = \{x \in X | \, |f(x)| \le n\}$ and note that $|\nu|(X \backslash X_n) \to 0$ since $X_n \nearrow X$ and $|\nu|(X) < \infty$. Given $\varepsilon > 0$ we can choose n such that $|\nu|(X \backslash X_n) \le \frac{\varepsilon}{2}$ and $\delta = \frac{\varepsilon}{2n}$. Then, if $\mu(A) < \delta$ we have

$$|\nu(A)| \le |\nu|(A \cap X_n) + |\nu|(X \backslash X_n) \le n\, \mu(A) + \frac{\varepsilon}{2} < \varepsilon.$$

The converse direction is obvious. □

It is important to emphasize that the fact that $|\nu|(X) < \infty$ is crucial for the above lemma to hold. In act, it can fail for positive measures as the simple counter example $d\nu(\lambda) = \lambda^2 d\lambda$ on $\mathbb{R}$ shows.

Problem 9.12. *Prove Corollary 9.18. (Hint: Use the complex Radon–Nikodym theorem to get existence of h. Then show that $1 - |h|$ vanishes a.e.)*

Problem 9.13 (Chebyshev's inequality)**.** *Let ν be a complex and μ a positive measure. If f denotes the Radon–Nikodym derivative of ν with respect to μ, then show that*

$$\mu(\{x \in A \,|\, |f(x)| \geq \alpha\}) \leq \frac{|\nu|(A)}{\alpha}.$$

Problem 9.14. *Let ν be a complex and μ a positive measure and suppose $|\nu(A)| \leq C\mu(A)$ for all $A \in \Sigma$. Then $d\nu = f\, d\mu$ with $\|f\|_\infty \leq C$. (Hint: First show $|\nu|(A) \leq C\mu(A)$ and then use Problem 9.4.)*

Problem 9.15. *Let ν be a signed measure and $\nu_\pm$ its Hahn decomposition. Show*

$$\nu_+(A) = \max_{B\in\Sigma, B\subseteq A} \nu(B), \qquad \nu_-(A) = -\min_{B\in\Sigma, B\subseteq A} \nu(B)$$

Problem 9.16. *Let ν be a complex measure and let*

$$\nu = \nu_{r,+} - \nu_{r,-} + \mathrm{i}(\nu_{i,+} - \nu_{i,-})$$

be its decomposition into positive measures. Show the estimate

$$\frac{1}{\sqrt{2}}\nu_s(A) \leq |\nu|(A) \leq \nu_s(A), \quad \nu_s = \nu_{r,+} + \nu_{r,-} + \nu_{i,+} + \nu_{i,-}.$$

Conclude that $|\nu(A)| \leq C$ for all measurable sets A implies $\|\nu\| \leq 4C$.

9.4. Appendix: Functions of bounded variation and absolutely continuous functions

Let $[a,b] \subseteq \mathbb{R}$ be some compact interval and $f : [a,b] \to \mathbb{C}$. Given a partition $P = \{a = x_0, \dots, x_n = b\}$ of $[a,b]$ we define the **variation** of f with respect to the partition P by

$$V(P,f) = \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|. \tag{9.23}$$

The supremum over all partitions

$$V_a^b(f) = \sup_{\text{partitions } P \text{ of } [a,b]} V(P,f) \tag{9.24}$$

is called the **total variation** of f over $[a,b]$. If the total variation is finite, f is called of **bounded variation**. Since we clearly have

$$V_a^b(\alpha f) = |\alpha| V_a^b(f), \qquad V_a^b(f+g) \le V_a^b(f) + V_a^b(g) \tag{9.25}$$

the space $BV[a,b]$ of all functions of finite total variation is a vector space. However, the total variation is not a norm since (consider the partition $P = \{a, x, b\}$)

$$V_a^b(f) = 0 \quad \Leftrightarrow \quad f(x) \equiv c. \tag{9.26}$$

Moreover, any function of bounded variation is in particular bounded (consider again the partition $P = \{a, x, b\}$)

$$\sup_{x\in[a,b]} |f(x)| \le |f(a)| + V_a^b(f). \tag{9.27}$$

Example. Any real-valued nondecreasing function f is of bounded variation with variation given by $V_a^b(f) = f(b) - f(a)$. Similarly, every real-valued nonincreasing function f is of bounded variation with variation given by $V_a^b(f) = f(a) - f(b)$. ⋄

Furthermore, observe $V_a^a(f) = 0$ as well as (Problem 9.18)

$$V_a^b(f) = V_a^c(f) + V_c^b(f), \qquad c \in [a,b], \tag{9.28}$$

and it will be convenient to set

$$V_b^a(f) = -V_a^b(f). \tag{9.29}$$

Theorem 9.20 (Jordan)**.** *Let $f : [a,b] \to \mathbb{R}$ be of bounded variation, then f can be decomposed as*

$$f(x) = f_+(x) - f_-(x), \qquad f_\pm(x) = \frac{1}{2}\left(V_a^x(f) \pm f(x)\right), \tag{9.30}$$

where $f_\pm$ are nondecreasing functions. Moreover, $V_a^b(f_\pm) \le V_a^b(f)$.

Proof. From

$$f(y) - f(x) \le |f(y) - f(x)| \le V_x^y(f) = V_a^y(f) - V_a^x(f)$$

for $x \le y$ we infer $V_a^x(f) - f(x) \le V_a^y(f) - f(y)$, that is, f_+ is nondecreasing. Moreover, replacing f by $-f$ shows that f_- is nondecreasing and the claim follows. □

In particular, we see that functions of bounded variation have at most countably many discontinuities and at every discontinuity the limits from the left and right exist.

For functions $f : (a,b) \to \mathbb{C}$ (including the case where (a,b) is unbounded) we will set

$$V_a^b(f) = \lim_{c\downarrow a, d\uparrow b} V_c^d(f). \tag{9.31}$$

In this respect the following lemma is of interest (whose proof is left as an exercise):

Lemma 9.21. *Suppose $f \in BV[a,b]$. We have $\lim_{c\uparrow b} V_a^c(f) = V_a^b(f)$ if and only if $f(b) = f(b-)$ and $\lim_{c\downarrow a} V_c^b(f) = V_a^b(f)$ if and only if $f(a) = f(a+)$. In particular, $V_a^x(f)$ is left, right continuous if and only f is.*

If $f : \mathbb{R} \to \mathbb{C}$ is of bounded variation, then we can write it as a linear combination of four nondecreasing functions and hence associate a complex measure df with f via Theorem 7.3 (since all four functions are bounded, so are the associated measures).

Theorem 9.22. *There is a one-to-one correspondence between functions in $f \in BV(\mathbb{R})$ which are right continuous and normalized by $f(0) = 0$ and complex Borel measures ν on $\mathbb{R}$ such that f is the distribution function of ν as defined in (7.3). Moreover, in this case the distribution function of the total variation of ν is $|\nu|(x) = V_0^x(f)$.*

Proof. We have already seen how to associate a complex measure df with a function of bounded variation. If f is right continuous and normalized, it will be equal to the distribution function of df by construction. Conversely, let $d\nu$ be a complex measure with distribution function ν. Then for every $a < b$ we have

$$\begin{aligned} V_a^b(\nu) &= \sup_{P=\{a=x_0,\dots,x_n=b\}} V(P,\nu) \\ &= \sup_{P=\{a=x_0,\dots,x_n=b\}} \sum_{k=1}^n |\nu((x_{k-1},x_k])| \le |\nu|((a,b]) \end{aligned}$$

and thus the distribution function is of bounded variation. Furthermore, consider the measure μ whose distribution function is $\mu(x) = V_0^x(\nu)$. Then we see $|\nu((a,b])| = |\nu(b) - \nu(a)| \le V_a^b(\nu) = \mu((a,b]) \le |\nu|((a,b])$. Hence we obtain $|\nu(A)| \le \mu(A) \le |\nu|(A)$ for all intervals A, thus for all open sets (by Problem 1.6), and thus for all Borel sets by outer regularity. Hence Lemma 9.13 implies $\mu = |\nu|$ and hence $|\nu|(x) = V_0^x(f)$. □

We will call a function $f : [a,b] \to \mathbb{C}$ **absolutely continuous** if for every $\varepsilon > 0$ there is a corresponding $\delta > 0$ such that

$$\sum_k |y_k - x_k| < \delta \quad \Rightarrow \quad \sum_k |f(y_k) - f(x_k)| < \varepsilon \tag{9.32}$$

for every countable collection of pairwise disjoint intervals $(x_k, y_k) \subset [a,b]$. The set of all absolutely continuous functions on $[a,b]$ will be denoted by $AC[a,b]$. The special choice of just one interval shows that every absolutely continuous function is (uniformly) continuous, $AC[a,b] \subseteq C[a,b]$.

Theorem 9.23. *A complex Borel measure ν on $\mathbb{R}$ is absolutely continuous with respect to Lebesgue measure if and only if its distribution function is locally absolutely continuous (i.e., absolutely continuous on every compact sub-interval). Moreover, in this case the distribution function $\nu(x)$ is differentiable almost everywhere and*

$$\nu(x) = \nu(0) + \int_0^x \nu'(y)dy \tag{9.33}$$

with ν' integrable, $\int_{\mathbb{R}} |\nu'(y)|dy = |\nu|(\mathbb{R})$.

Proof. Suppose the measure ν is absolutely continuous. Since we can write ν as a sum of four positive measures, we can suppose ν is positive. Now (9.32) follows from (9.22) in the special case where A is a union of pairwise disjoint intervals.

Conversely, suppose $\nu(x)$ is absolutely continuous on $[a, b]$. We will verify (9.22). To this end fix ε and choose δ such that $\nu(x)$ satisfies (9.32). By outer regularity it suffices to consider the case where A is open. Moreover, by Problem 1.6, every open set $O \subset (a, b)$ can be written as a countable union of disjoint intervals $I_k = (x_k, y_k)$ and thus $|O| = \sum_k |y_k - x_k| \le \delta$ implies

$$\nu(O) = \sum_k \big(\nu(y_k) - \nu(x_k)\big) \le \sum_k |\nu(y_k) - \nu(x_k)| \le \varepsilon$$

as required.

The rest follows from Corollary 9.8. □

As a simple consequence of this result we can give an equivalent definition of absolutely continuous functions as precisely the functions for which the **fundamental theorem of calculus** holds.

Theorem 9.24. *A function $f : [a, b] \to \mathbb{C}$ is absolutely continuous if and only if it is of the form*

$$f(x) = f(a) + \int_a^x g(y)dy \tag{9.34}$$

for some integrable function g. Moreover, in this case f is differentiable a.e with respect to Lebesgue measure and $f'(x) = g(x)$. In addition, f is of bounded variation and

$$V_a^x(f) = \int_a^x |g(y)|dy. \tag{9.35}$$

Proof. This is just a reformulation of the previous result. To see the last claim combine the last part of Theorem 9.22 with Lemma 9.17. □

Finally, we note that in this case the **integration by parts formula** continues to hold.

Lemma 9.25. *Let $f, g \in BV[a,b]$, then*

$$\int_{[a,b)} f(x-)dg(x) = f(b-)g(b-) - f(a-)g(a-) - \int_{[a,b)} g(x+)df(x) \quad (9.36)$$

as well as

$$\int_{(a,b]} f(x+)dg(x) = f(b+)g(b+) - f(a+)g(a+) - \int_{[a,b)} g(x-)df(x). \quad (9.37)$$

Proof. Since the formula is linear in f and holds if f is constant, we can assume $f(a-) = 0$ without loss of generality. Similarly we can assume $g(b-) = 0$. Plugging $f(x-) = \int_{[a,x)} df(y)$ into the left-hand side of the first formula we obtain from Fubini

$$\begin{aligned}\int_{[a,b)} f(x-)dg(x) &= \int_{[a,b)} \int_{[a,x)} df(y)dg(x)\\ &= \int_{[a,b)} \int_{[a,b)} \chi_{\{(x,y)|y<x\}}(x,y)df(y)dg(x)\\ &= \int_{[a,b)} \int_{[a,b)} \chi_{\{(x,y)|y<x\}}(x,y)dg(x)df(y)\\ &= \int_{[a,b)} \int_{(y,b)} dg(x)df(y) = -\int_{[a,b)} g(y+)df(y).\end{aligned}$$

The second formula is shown analogously. □

Problem 9.17. *Compute $V_a^b(f)$ for $f(x) = \operatorname{sign}(x)$ on $[a,b] = [-1,1]$.*

Problem 9.18. *Show (9.28).*

Problem 9.19. *Consider $f_j(x) = x^j \cos(\pi/x)$ for $j \in \mathbb{N}$. Show that $f_j \in C[0,1]$ if we set $f_j(0) = 0$. Show that f_j is of bounded variation for $j \geq 2$ but not for $j = 1$.*

Problem 9.20. *Show that if $f \in BV[a,b]$ then so is f^*, $|f|$ and*

$$V_a^b(f^*) = V_a^b(f), \qquad V_a^b(|f|) \leq V_a^b(f).$$

Moreover, show

$$V_a^b(\mathrm{Re}(f)) \leq V_a^b(f), \qquad V_a^b(\mathrm{Im}(f)) \leq V_a^b(f).$$

Problem 9.21. *Show that if $f, g \in BV[a,b]$ then so is $f\,g$ and*

$$V_a^b(f\,g) \leq V_a^b(f) \sup|g| + V_a^b(g) \sup|f|.$$

Problem 9.22. *Show that $BV[a,b]$ together with the norm*

$$|f(a)| + V_a^b(f)$$

is a Banach space.

Problem 9.23 (Product rule for absolutely continuous functions). *Show that if $f, g \in AC[a,b]$ then so is $f\,g$ and $(f\,g)' = f'g + f\,g'$. (Hint: Integration by parts.)*

Problem 9.24. *Let $X \subseteq \mathbb{R}$, Y be some measure space, and $f : X \times Y \to \mathbb{R}$ some measurable function. Suppose $x \mapsto f(x,y)$ is absolutely continuous for a.e. y such that*

$$\int_a^b \int_A \left| \frac{\partial}{\partial x} f(x,y) \right| d\mu(y) dx < \infty \tag{9.38}$$

for every compact interval $[a,b] \subseteq X$ and $\int_A |f(c,y)| d\mu(y) < \infty$ for one $c \in X$.

Show that

$$F(x) = \int_A f(x,y)\, d\mu(y) \tag{9.39}$$

is absolutely continuous and

$$F'(x) = \int_A \frac{\partial}{\partial x} f(x,y)\, d\mu(y) \tag{9.40}$$

in this case. (Hint: Fubini.)

Problem 9.25. *Show that if $f \in AC(a,b)$ and $f' \in L^p(a,b)$, then f is Hölder continuous:*

$$|f(x) - f(y)| \le \|f'\|_p |x-y|^{1-\frac{1}{p}}.$$

Chapter 10

The dual of L^p

10.1. The dual of L^p, $p < \infty$

By the Hölder inequality every $g \in L^q(X, d\mu)$ gives rise to a linear functional on $L^p(X, d\mu)$ and this clearly raises the question if every linear functional is of this form. In order to answer this question we begin with a lemma.

Lemma 10.1. *Consider $L^p(X, d\mu)$ and let q be the corresponding dual index, $\frac{1}{p} + \frac{1}{q} = 1$. Then for every measurable function f*

$$\|f\|_p = \sup_{g \text{ simple}, \ \|g\|_q = 1} \left| \int_X f g \, d\mu \right|$$

if $1 \le p < \infty$. If μ is σ-finite the claim also holds for $p = \infty$.

Proof. We begin with the case $1 \le p < \infty$. Choosing $h = |f|^{p-1} \operatorname{sign}(f^*)$ (where $|f|^0 = 1$ in the case $p = 1$) it follows that $h \in L^q$ and $g = h/\|h\|_q$ satisfies $\int_X f g \, d\mu = \|f\|_p$. Now since the simple functions are dense in L^q there is a sequence of simple functions $s_n \to g$ in L^q and Hölder's inequality implies $\int_X f s_n \, d\mu \to \int_X f g \, d\mu$.

Now let us turn to the case $p = \infty$. For every $\varepsilon > 0$ the set $A_\varepsilon = \{x | \, |f(x)| \ge \|f\|_\infty - \varepsilon\}$ has positive measure. Moreover, considering $X_n \nearrow X$ with $\mu(X_n) < \infty$ there must be some n such that $B_\varepsilon = A_\varepsilon \cap X_n$ satisfies $0 < \mu(B_\varepsilon) < \infty$. Then $g_\varepsilon = \operatorname{sign}(f^*) \chi_{B_\varepsilon} / \mu(B_\varepsilon)$ satisfies $\int_X f g_\varepsilon \, d\mu \ge \|f\|_\infty - \varepsilon$. Finally, choosing a sequence of simple functions $s_n \to g_\varepsilon$ in L^1 finishes the proof. □

After these preparations we are able to identify the dual of L^p for $p < \infty$.

Theorem 10.2. *Consider $L^p(X, d\mu)$ with some σ-finite measure μ and let q be the corresponding dual index, $\frac{1}{p} + \frac{1}{q} = 1$. Then the map $g \in L^q \mapsto \ell_g \in (L^p)^*$ given by*

$$\ell_g(f) = \int_X g f \, d\mu \tag{10.1}$$

is an isometric isomorphism for $1 \le p < \infty$. If $p = \infty$ it is at least isometric.

Proof. Given $g \in L^q$ it follows from Hölder's inequality that ℓ_g is a bounded linear functional with $\|\ell_g\| \le \|g\|_q$. Moreover, $\|\ell_g\| = \|g\|_q$ follows from Lemma 10.1.

To show that this map is surjective if $1 \le p < \infty$, first suppose $\mu(X) < \infty$ and choose some $\ell \in (L^p)^*$. Since $\|\chi_A\|_p = \mu(A)^{1/p}$, we have $\chi_A \in L^p$ for every $A \in \Sigma$ and we can define

$$\nu(A) = \ell(\chi_A).$$

Suppose $A = \bigcup_{j=1}^\infty A_j$, where the A_j's are disjoint. Then, by dominated convergence, $\| \sum_{j=1}^n \chi_{A_j} - \chi_A\|_p \to 0$ (this is false for $p = \infty$!) and hence

$$\nu(A) = \ell(\sum_{j=1}^\infty \chi_{A_j}) = \sum_{j=1}^\infty \ell(\chi_{A_j}) = \sum_{j=1}^\infty \nu(A_j).$$

Thus ν is a complex measure. Moreover, $\mu(A) = 0$ implies $\chi_A = 0$ in L^p and hence $\nu(A) = \ell(\chi_A) = 0$. Thus ν is absolutely continuous with respect to μ and by the complex Radon–Nikodym theorem $d\nu = g\, d\mu$ for some $g \in L^1(X, d\mu)$. In particular, we have

$$\ell(f) = \int_X f g \, d\mu$$

for every simple function f. Next let $A_n = \{x | |g| < n\}$, then $g_n = g\chi_{A_n} \in L^q$ and by Lemma 10.1 we conclude $\|g_n\|_q \le \|\ell\|$. Letting $n \to \infty$ shows $g \in L^q$ and finishes the proof for finite μ.

If μ is σ-finite, let $X_n \nearrow X$ with $\mu(X_n) < \infty$. Then for every n there is some g_n on X_n and by uniqueness of g_n we must have $g_n = g_m$ on $X_n \cap X_m$. Hence there is some g and by $\|g_n\| \le \|\ell\|$ independent of n, we have $g \in L^q$. □

Corollary 10.3. *Let μ be some σ-finite measure. Then $L^p(X, d\mu)$ is reflexive for $1 < p < \infty$.*

Proof. Identify $L^p(X, d\mu)^*$ with $L^q(X, d\mu)$ and choose $h \in L^p(X, d\mu)^{**}$. Then there is some $f \in L^p(X, d\mu)$ such that

$$h(g) = \int g(x) f(x) d\mu(x), \qquad g \in L^q(X, d\mu) \cong L^p(X, d\mu)^*.$$

But this implies $h(g) = g(f)$, that is, $h = J(f)$, and thus J is surjective. □

10.2. The dual of L^∞ and the Riesz representation theorem

In the last section we have computed the dual space of L^p for $p < \infty$. Now we want to investigate the case $p = \infty$. Recall that we already know that the dual of L^∞ is much larger than L^1 since it cannot be separable in general.

Example. Let ν be a complex measure. Then

$$\ell_\nu(f) = \int_X f d\nu \tag{10.2}$$

is a bounded linear functional on $B(X)$ (the Banach space of bounded measurable functions) with norm

$$\|\ell_\nu\| = |\nu|(X) \tag{10.3}$$

by (9.21) and Corollary 9.18. If ν is absolutely continuous with respect to μ, then it will even be a bounded linear functional on $L^\infty(X, d\mu)$ since the integral will be independent of the representative in this case. ⋄

So the dual of $B(X)$ contains all complex measures. However, this is still not all of $B(X)^*$. In fact, it turns out that it suffices to require only finite additivity for ν.

Let (X, Σ) be a measurable space. A **complex content** ν is a map $\nu : \Sigma \to \mathbb{C}$ such that (finite additivity)

$$\nu(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n \nu(A_k), \qquad A_j \cap A_k = \emptyset,\ j \neq k. \tag{10.4}$$

Given a content ν we can define the corresponding integral for simple functions $s(x) = \sum_{k=1}^n \alpha_k \chi_{A_k}$ as usual

$$\int_A s\, d\nu = \sum_{k=1}^n \alpha_k \nu(A_k \cap A). \tag{10.5}$$

As in the proof of Lemma 7.17 one shows that the integral is linear. Moreover,

$$|\int_A s\, d\nu| \le |\nu|(A)\, \|s\|_\infty, \tag{10.6}$$

where $|\nu|(A)$ is defined as in (9.15) and the same proof as in Theorem 9.12 shows that $|\nu|$ is a content. However, since we do not require σ-additivity, it is not clear that $|\nu|(X)$ is finite. Hence we will call ν finite if $|\nu|(X) < \infty$. Moreover, for a finite content this integral can be extended to all of $B(X)$ such that

$$|\int_X f\, d\nu| \le |\nu|(X)\, \|f\|_\infty \tag{10.7}$$

by Theorem 1.29 (compare Problem 7.13). However, note that our convergence theorems (monotone convergence, dominated convergence) will no longer hold in this case (unless ν happens to be a measure).

In particular, every complex content gives rise to a bounded linear functional on $B(X)$ and the converse also holds:

Theorem 10.4. *Every bounded linear functional $\ell \in B(X)^*$ is of the form*

$$\ell(f) = \int_X f\, d\nu \tag{10.8}$$

for some unique complex content ν and $\|\ell\| = |\nu|(X)$.

Proof. Let $\ell \in B(X)^*$ be given. If there is a content ν at all it is uniquely determined by $\nu(A) = \ell(\chi_A)$. Using this as definition for ν, we see that finite additivity follows from linearity of ℓ. Moreover, (10.8) holds for characteristic functions and by

$$\ell(\sum_{k=1}^n \alpha_k \chi_{A_k}) = \sum_{k=1}^n \alpha_k \nu(A_k) = \sum_{k=1}^n |\nu(A_k)|, \qquad \alpha_k = \mathrm{sign}(\nu(A_k)),$$

we see $|\nu|(X) \le \|\ell\|$.

Since the characteristic functions are total, (10.8) holds everywhere by continuity and (10.7) shows $\|\ell\| \le |\nu|(X)$. □

Remark: To obtain the dual of $L^\infty(X, d\mu)$ from this you just need to restrict to those linear functionals which vanish on $\mathcal{N}(X, d\mu)$ (cf. Problem 10.1), that is, those whose content is *absolutely continuous* with respect to μ (note that the Radon–Nikodym theorem does not hold unless the content is a measure).

Example. Consider $B(\mathbb{R})$ and define

$$\ell(f) = \lim_{\varepsilon \downarrow 0} (\lambda f(-\varepsilon) + (1-\lambda) f(\varepsilon)), \qquad \lambda \in [0,1], \tag{10.9}$$

for f in the subspace of bounded measurable functions which have left and right limits at 0. Since $\|\ell\| = 1$ we can extend it to all of $B(\mathbb{R})$ using the Hahn–Banach theorem. Then the corresponding content ν is no measure:

$$\lambda = \nu([-1,0)) = \nu(\bigcup_{n=1}^\infty [-\frac{1}{n}, -\frac{1}{n+1})) \ne \sum_{n=1}^\infty \nu([-\frac{1}{n}, -\frac{1}{n+1})) = 0. \tag{10.10}$$

Observe that the corresponding distribution function (defined as in (7.3)) is nondecreasing but not right continuous! If we render ν right continuous, we get the the distribution function of the Dirac measure (centered at 0). In addition, the Dirac measure has the same integral at least for continuous functions! ◇

Theorem 10.5 (Riesz representation)**.** *Let $I = [a, b] \subseteq \mathbb{R}$ be a compact interval. Every bounded linear functional $\ell \in C(I)^*$ is of the form*

$$\ell(f) = \int_I f\, d\nu \tag{10.11}$$

for some unique complex Borel measure ν and $\|\ell\| = |\nu|(I)$.

Moreover, in the case $I = \mathbb{R}$ every bounded linear functional $\ell \in C_\infty(\mathbb{R})^*$ is of the above form.

Proof. Extending ℓ to a bounded linear functional $\overline{\ell} \in B(I)^*$ we have a corresponding content $\tilde{\nu}$. Splitting this content into real and imaginary part we see that it is no restriction to assume that $\tilde{\nu}$ is real. Moreover, splitting $\tilde{\nu}$ into $\tilde{\nu}_\pm = (|\tilde{\nu}| \pm \tilde{\nu})/2$ it is no restriction to assume $\tilde{\nu}$ is positive.

Now the idea is as follows: Define a distribution function for $\tilde{\nu}$ as in (7.3). By finite additivity of $\tilde{\nu}$ it will be nondecreasing and we can use Theorem 7.3 to obtain an associated measure ν whose distribution function coincides with $\tilde{\nu}$ except possibly at points where ν is discontinuous. It remains to show that the corresponding integral coincides with ℓ for continuous functions.

Let $f \in C(I)$ be given. Fix points $a < x_0^n < x_1^n < \dots x_n^n < b$ such that $x_0^n \to a$, $x_n^n \to b$, and $\sup_k |x_{k-1}^n - x_k^n| \to 0$ as $n \to \infty$. Then the sequence of simple functions

$$f_n(x) = f(x_0^n)\chi_{[x_0^n, x_1^n)} + f(x_1^n)\chi_{[x_1^n, x_2^n)} + \cdots + f(x_{n-1}^n)\chi_{[x_{n-1}^n, x_n^n]}.$$

converges uniformly to f by continuity of f (and the fact that f vanishes as $x \to \pm\infty$ in the case $I = \mathbb{R}$). Moreover,

$$\begin{aligned}
\int_I f\, d\nu &= \lim_{n\to\infty} \int_I f_n\, d\nu = \lim_{n\to\infty} \sum_{k=1}^n f(x_{k-1}^n)(\nu(x_k^n) - \nu(x_{k-1}^n)) \\
&= \lim_{n\to\infty} \sum_{k=1}^n f(x_{k-1}^n)(\tilde{\nu}(x_k^n) - \tilde{\nu}(x_{k-1}^n)) = \lim_{n\to\infty} \int_I f_n\, d\tilde{\nu} \\
&= \int_I f\, d\tilde{\nu} = \ell(f)
\end{aligned}$$

provided the points x_k^n are chosen to stay away from all discontinuities of $\nu(x)$ (recall that there are at most countably many).

To see $\|\ell\| = |\nu|(I)$ recall $d\nu = h d|\nu|$ where $|h| = 1$ (Corollary 9.18). Now choose continuous functions $h_n(x) \to h(x)$ pointwise a.e. Using $\tilde{h}_n = \max(1, |h_n|)\operatorname{sign}(h_n)$ we see even get such a sequence with $|\tilde{h}_n| \le 1$. Hence $\ell(\tilde{h}_n) = \int \tilde{h}_n^* h d\|\nu| \to \int |h|^2 d|\nu| = |\nu|(I)$ implying $\|\ell\| \ge |\nu|(I)$. The converse follows from (10.7). $\square$

Note that ν will be a positive measure if ℓ is a **positive functional**, that is, $\ell(f) \geq 0$ whenever $f \geq 0$.

Problem 10.1. *Let M be a closed subspace of a Banach space X. Show that $(X/M)^* \cong \{\ell \in X^* | M \subseteq \mathrm{Ker}(\ell)\}$.*

Problem 10.2 (Vague convergence of measures)**.** *Let I be a compact interval. A sequence of measures ν_n is said to converge vaguely to a measure ν if*

$$\int_I f d\nu_n \to \int_I f d\nu, \qquad f \in C(I). \tag{10.12}$$

Show that every bounded sequence of measures has a vaguely convergent subsequence. Show that the limit ν is a positive measure if all ν_n are. (Hint: Compare this definition to the definition of weak-$*$ convergence in Section 4.3.)

Chapter 11

The Fourier transform

11.1. The Fourier transform on L^2

For $f \in L^1(\mathbb{R}^n)$ we define its **Fourier transform** via

$$\mathcal{F}(f)(p) \equiv \hat{f}(p) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}px} f(x) d^n x. \tag{11.1}$$

Lemma 11.1. *The Fourier transform is a bounded map from $L^1(\mathbb{R}^n)$ into $C_b(\mathbb{R}^n)$ satisfying*

$$\|\hat{f}\|_\infty \le (2\pi)^{-n/2} \|f\|_1. \tag{11.2}$$

Proof. Since $|\mathrm{e}^{-\mathrm{i}px}| = 1$ the estimate (11.2) is immediate from

$$|\hat{f}(p)| \le \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} |\mathrm{e}^{-\mathrm{i}px} f(x)| d^n x = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} |f(x)| d^n x.$$

Moreover, a straightforward application of the dominated convergence theorem shows that $\hat{f}$ is continuous. □

Three more simple properties are left as an exercise.

Lemma 11.2. *Let $f \in L^1(\mathbb{R}^n)$. Then*

$$(f(x+a))^\wedge(p) = \mathrm{e}^{\mathrm{i}ap} \hat{f}(p), \qquad a \in \mathbb{R}^n, \tag{11.3}$$

$$(\mathrm{e}^{\mathrm{i}xa} f(x))^\wedge(p) = \hat{f}(p-a), \qquad a \in \mathbb{R}^n, \tag{11.4}$$

$$(f(\lambda x))^\wedge(p) = \frac{1}{\lambda^n} \hat{f}(\frac{p}{\lambda}), \qquad \lambda > 0. \tag{11.5}$$

Next we look at the connection with differentiation.

Lemma 11.3. *Suppose $f \in C^1(\mathbb{R}^n)$ such that $\lim_{|x|\to\infty} f(x) = 0$ and $f, \partial_j f \in L^1(\mathbb{R}^n)$ for some $1 \le j \le n$. Then*

$$(\partial_j f)^\wedge(p) = \mathrm{i} p_j \hat{f}(p). \tag{11.6}$$

Similarly, if $f(x), x_j f(x) \in L^1(\mathbb{R}^n)$ for some $1 \le j \le n$, then $\hat{f}(p)$ is differentiable with respect to p_j and

$$(x_j f(x))^\wedge(p) = \mathrm{i}\partial_j \hat{f}(p). \tag{11.7}$$

Proof. First of all, by integration by parts, we see

$$\begin{aligned}(\partial_j f)^\wedge(p) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}px} \frac{\partial}{\partial x_j} f(x) d^n x \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(-\frac{\partial}{\partial x_j} \mathrm{e}^{-\mathrm{i}px} \right) f(x) d^n x \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{i} p_j \mathrm{e}^{-\mathrm{i}px} f(x) d^n x = \mathrm{i} p_j \hat{f}(p).\end{aligned}$$

Similarly, the second formula follows from

$$\begin{aligned}(x_j f(x))^\wedge(p) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} x_j \mathrm{e}^{-\mathrm{i}px} f(x) d^n x \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\mathrm{i} \frac{\partial}{\partial p_j} \mathrm{e}^{-\mathrm{i}px} \right) f(x) d^n x = \mathrm{i} \frac{\partial}{\partial p_j} \hat{f}(p),\end{aligned}$$

where interchanging the derivative and integral is permissible by Problem 7.16. In particular, $\hat{f}(p)$ is differentiable. □

This result immediately extends to higher derivatives. To this end let $C^\infty(\mathbb{R}^n)$ be the set of all complex-valued functions which have partial derivatives of arbitrary order. For $f \in C^\infty(\mathbb{R}^n)$ and $\alpha \in \mathbb{N}_0^n$ we set

$$\partial_\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}, \quad x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \quad |\alpha| = \alpha_1 + \cdots + \alpha_n. \tag{11.8}$$

An element $\alpha \in \mathbb{N}_0^n$ is called a **multi-index** and $|\alpha|$ is called its **order**. We will also set $(\lambda x)^\alpha = \lambda^{|\alpha|} x^\alpha$ for $\lambda \in \mathbb{R}$. Recall the Schwartz space

$$\mathcal{S}(\mathbb{R}^n) = \{ f \in C^\infty(\mathbb{R}^n) | \sup_x |x^\alpha (\partial_\beta f)(x)| < \infty, \, \alpha, \beta \in \mathbb{N}_0^n \} \tag{11.9}$$

which is a subspace of $L^p(\mathbb{R}^n)$ and which is dense for $1 \le p < \infty$ (since $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$). Together with the seminorms $\|x^\alpha (\partial_\beta f)(x)\|_\infty$ it is a Fréchet space. Note that if $f \in \mathcal{S}(\mathbb{R}^n)$, then the same is true for $x^\alpha f(x)$ and $(\partial_\alpha f)(x)$ for every multi-index α.

Lemma 11.4. *The Fourier transform satisfies $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$. Furthermore, for every multi-index $\alpha \in \mathbb{N}_0^n$ and every $f \in \mathcal{S}(\mathbb{R}^n)$ we have*

$$(\partial_\alpha f)^\wedge(p) = (\mathrm{i}p)^\alpha \hat{f}(p), \qquad (x^\alpha f(x))^\wedge(p) = \mathrm{i}^{|\alpha|} \partial_\alpha \hat{f}(p). \tag{11.10}$$

Proof. The formulas are immediate from the previous lemma. To see that $\hat{f} \in \mathcal{S}(\mathbb{R}^n)$ if $f \in \mathcal{S}(\mathbb{R}^n)$, we begin with the observation that $\hat{f}$ is bounded by (11.2). But then $p^\alpha(\partial_\beta \hat{f})(p) = \mathrm{i}^{-|\alpha|-|\beta|}(\partial_\alpha x^\beta f(x))^\wedge(p)$ is bounded since $\partial_\alpha x^\beta f(x) \in \mathcal{S}(\mathbb{R}^n)$ if $f \in \mathcal{S}(\mathbb{R}^n)$. □

Hence we will sometimes write $pf(x)$ for $-\mathrm{i}\partial f(x)$, where $\partial = (\partial_1, \dots, \partial_n)$ is the **gradient**. As an application of the previous lemma we can compute the Fourier transform of a Gaussian function.

Lemma 11.5. *We have* $\mathrm{e}^{-zx^2/2} \in \mathcal{S}(\mathbb{R}^n)$ *for* $\mathrm{Re}(z) > 0$ *and*

$$\mathcal{F}(\mathrm{e}^{-zx^2/2})(p) = \frac{1}{z^{n/2}} \mathrm{e}^{-p^2/(2z)}. \tag{11.11}$$

Here $z^{n/2}$ *is the standard branch with branch cut along the negative real axis.*

Proof. Due to the product structure of the exponential, one can treat each coordinate separately, reducing the problem to the case $n = 1$.

Let $\phi_z(x) = \exp(-zx^2/2)$. Then $\phi_z'(x) + zx\phi_z(x) = 0$ and hence $\mathrm{i}(p\hat{\phi}_z(p) + z\hat{\phi}_z'(p)) = 0$. Thus $\hat{\phi}_z(p) = c\phi_{1/z}(p)$ and (Problem 11.2)

$$c = \hat{\phi}_z(0) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-zx^2/2) dx = \frac{1}{\sqrt{z}}$$

at least for $z > 0$. However, since the integral is holomorphic for $\mathrm{Re}(z) > 0$ by Problem 7.18, this holds for all z with $\mathrm{Re}(z) > 0$ if we choose the branch cut of the root along the negative real axis. □

Now we can also show

Theorem 11.6. *Suppose* $f, \hat{f} \in L^1(\mathbb{R}^n)$. *Then*

$$(\hat{f})^\vee = f, \tag{11.12}$$

where

$$\check{f}(p) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}px} f(x) d^n x = \hat{f}(-p). \tag{11.13}$$

In particular, $\mathcal{F} : F^1(\mathbb{R}^n) \to F^1(\mathbb{R}^n)$ *is a bijection, where* $F^1(\mathbb{R}^n) = \{f \in L^1(\mathbb{R}^n) | \hat{f} \in L^1(\mathbb{R}^n)\}$. *Moreover,* $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ *is a bijection.*

Proof. Abbreviate $\phi_\varepsilon(x) = \exp(-\varepsilon x^2/2)$. By dominated convergence we have

$$\begin{aligned}
(\hat{f}(p))^\vee(x) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}px} \hat{f}(p) d^n p \\
&= \lim_{\varepsilon \to 0} \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \phi_\varepsilon(p) \mathrm{e}^{\mathrm{i}px} \hat{f}(p) d^n p \\
&= \lim_{\varepsilon \to 0} \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \phi_\varepsilon(p) \mathrm{e}^{\mathrm{i}px} f(y) \mathrm{e}^{-\mathrm{i}py} d^n y d^n p,
\end{aligned}$$

and, invoking Fubini and Lemma 11.2, we further see

$$= \lim_{\varepsilon\to 0} \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} (\phi_\varepsilon(p)\mathrm{e}^{\mathrm{i}px})^\wedge(y) f(y) d^n y$$
$$= \lim_{\varepsilon\to 0} \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \frac{1}{\varepsilon^{n/2}} \phi_{1/\varepsilon}(y-x) f(y) d^n y.$$

But the last integral converges to f in $L^1(\mathbb{R}^n)$ by Problem 8.11 which finishes the proof. □

For the next lemma let $C_\infty(\mathbb{R}^n)$ denote the Banach space of all continuous functions $f: \mathbb{R}^n \to \mathbb{C}$ which vanish at ∞ equipped with the sup norm (Problem 11.6).

Lemma 11.7 (Riemann-Lebesgue)**.** *The Fourier transform is a bounded injective map from $L^1(\mathbb{R}^n)$ into $C_\infty(\mathbb{R}^n)$.*

Proof. Clearly we have $\hat{f} \in C_\infty(\mathbb{R}^n)$ if $f \in \mathcal{S}(\mathbb{R}^n)$. Moreover, since $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$, the estimate (11.2) shows that the Fourier transform extends to a continuous map from $L^1(\mathbb{R}^n)$ into $C_\infty(\mathbb{R}^n)$.

To see that the Fourier transform is injective, suppose $\hat{f} = 0$. Then Fubini implies

$$0 = \int_{\mathbb{R}^n} \varphi(x) \hat{f}(x) d^n x = \int_{\mathbb{R}^n} \hat{\varphi}(x) f(x) d^n x$$

for every $\varphi \in \mathcal{S}(\mathbb{R}^n)$. Hence Lemma 8.12 implies $f = 0$. □

Note that $\mathcal{F}: L^1(\mathbb{R}^n) \to C_\infty(\mathbb{R}^n)$ is not onto (cf. Problem 11.7).

Lemma 11.8. *Suppose $f, \hat{f} \in L^1(\mathbb{R}^n)$. Then $f, \hat{f} \in L^2(\mathbb{R}^n)$ and*

$$\|f\|_2^2 = \|\hat{f}\|_2^2 \le (2\pi)^{-n/2} \|f\|_1 \|\hat{f}\|_1 \tag{11.14}$$

holds.

Proof. This follows from Fubini's theorem since

$$\int_{\mathbb{R}^n} |\hat{f}(p)|^2 d^n p = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x)^* \hat{f}(p) \mathrm{e}^{\mathrm{i}px} d^n p \, d^n x$$
$$= \int_{\mathbb{R}^n} |f(x)|^2 d^n x \tag{11.15}$$

for $f, \hat{f} \in L^1(\mathbb{R}^n)$. □

The identity $\|f\|_2 = \|\hat{f}\|_2$ is known as the **Plancherel identity**. Thus, by Theorem 1.29, we can extend $\mathcal{F}$ to all of $L^2(\mathbb{R}^n)$ by setting $\mathcal{F}(f) = \lim_{m\to\infty} \mathcal{F}(f_m)$, where f_m is an arbitrary sequence from, say, $\mathcal{S}(\mathbb{R}^n)$ converging to f in the L^2 norm.

Theorem 11.9 (Plancherel)**.** *The Fourier transform $\mathcal{F}$ extends to a unitary operator $\mathcal{F}: L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$.*

Proof. As already noted, $\mathcal{F}$ extends uniquely to a bounded operator on $L^2(\mathbb{R}^n)$. Since Plancherel's identity remains valid by continuity of the norm and since its range is dense, this extension is a unitary operator. □

We also note that this extension is still given by (11.1) whenever the right-hand side is integrable.

Lemma 11.10. *Let $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, then (11.1) and (11.13) continue to hold, where $\mathcal{F}$ now denotes the extension of the Fourier transform from $\mathcal{S}(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$.*

Proof. Fix a bounded set $X \subset \mathbb{R}^n$ and let $f \in L^2(X)$. Then we can approximate f by functions $f_n \in C_c^\infty(X)$ in the L^2 norm. Since $L^2(X)$ is continuously embedded into $L^1(X)$ (Problem 8.7), this sequence will also converge in $L^1(X)$. Extending all functions to $\mathbb{R}^n$ by setting them zero outside X we see that the claim holds for $f \in L^2(\mathbb{R}^n)$ with compact support. Finally, for general $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ consider $f_m = f\chi_{B_m(0)}$. Then $f_m \to f$ in both $L^1(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$ and the claim follows. □

In particular,

$$\hat{f}(p) = \lim_{m\to\infty} \frac{1}{(2\pi)^{n/2}} \int_{|x|\le m} \mathrm{e}^{-\mathrm{i}px} f(x) d^n x, \tag{11.16}$$

where the limit has to be understood in $L^2(\mathbb{R}^n)$ and can be omitted if $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$.

Another useful property is the convolution formula.

Lemma 11.11. *The* **convolution**

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y) g(x-y) d^n y = \int_{\mathbb{R}^n} f(x-y) g(y) d^n y \tag{11.17}$$

of two functions $f, g \in L^1(\mathbb{R}^n)$ is again in $L^1(\mathbb{R}^n)$ and we have **Young's inequality**

$$\|f * g\|_1 \le \|f\|_1 \|g\|_1. \tag{11.18}$$

Moreover, its Fourier transform is given by

$$(f * g)^\wedge(p) = (2\pi)^{n/2} \hat{f}(p) \hat{g}(p). \tag{11.19}$$

Proof. The fact that $f * g$ is in L^1 together with Young's inequality follows by applying Fubini's theorem to $h(x,y) = f(x-y)g(y)$. For the last claim we compute

$$(f*g)^\wedge(p) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}px} \int_{\mathbb{R}^n} f(y)g(x-y) d^n y\, d^n x$$
$$= \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}py} f(y) \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}p(x-y)} g(x-y) d^n x\, d^n y$$
$$= \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}py} f(y)\hat{g}(p) d^n y = (2\pi)^{n/2} \hat{f}(p)\hat{g}(p),$$

where we have again used Fubini's theorem. □

In other words, $L^1(\mathbb{R}^n)$ together with convolution as a product is a Banach algebra (without identity). As a consequence we can also deal with the case of convolution on $\mathcal{S}(\mathbb{R}^n)$ as well as on $L^2(\mathbb{R}^n)$.

Corollary 11.12. *The convolution of two $\mathcal{S}(\mathbb{R}^n)$ functions as well as their product is in $\mathcal{S}(\mathbb{R}^n)$ and*

$$(f*g)^\wedge = (2\pi)^{n/2}\hat{f}\hat{g}, \qquad (fg)^\wedge = (2\pi)^{-n/2}\hat{f}*\hat{g}$$

in this case.

Proof. Clearly the product of two functions in $\mathcal{S}(\mathbb{R}^n)$ is again in $\mathcal{S}(\mathbb{R}^n)$ (show this!). Since $\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$ the previous lemma implies $(f*g)^\wedge = (2\pi)^{n/2}\hat{f}\hat{g} \in \mathcal{S}(\mathbb{R}^n)$. Moreover, since the Fourier transform is injective on $L^1(\mathbb{R}^n)$ we conclude $f*g = (2\pi)^{n/2}(\hat{f}\hat{g})^\vee \in \mathcal{S}(\mathbb{R}^n)$. Replacing f, g by $\check{f}, \check{g}$ in the last fomula finally shows $\check{f}*\check{g} = (2\pi)^{n/2}(fg)^\vee$ and the claim follows by a simple change of variables using $\check{f}(p) = \hat{f}(-p)$. □

Corollary 11.13. *The convolution of two $L^2(\mathbb{R}^n)$ functions is in $C_\infty(\mathbb{R}^n)$ and we have $\|f*g\|_\infty \le \|f\|_2\|g\|_2$ as well as*

$$(fg)^\wedge = (2\pi)^{-n/2}\hat{f}*\hat{g}$$

in this case.

Proof. The inequality $\|f*g\|_\infty \le \|f\|_2\|g\|_2$ is immediate from Cauchy–Schwarz and shows that the convolution is a continuous bilinear form from $L^2(\mathbb{R}^n)$ to $L^\infty(\mathbb{R}^n)$. Now take sequences $f_n, g_n \in \mathcal{S}(\mathbb{R}^n)$ converging to $f, g \in L^2(\mathbb{R}^n)$. Then using the previous corollary together with continuity of the Fourier transform from $L^1(\mathbb{R}^n)$ to $C_\infty(\mathbb{R}^n)$ and on $L^2(\mathbb{R}^n)$ we obtain

$$(fg)^\wedge = \lim_{n\to\infty}(f_n g_n)^\wedge = (2\pi)^{-n/2}\lim_{n\to\infty}\hat{f}_n*\hat{g}_n = (2\pi)^{-n/2}\hat{f}*\hat{g}.$$

This also shows $\hat{f}*\hat{g} \in C_\infty(\mathbb{R}^n)$ by the Riemann–Lebesgue lemma. □

Finally, note that by looking at the Gaussian's $\phi_\lambda(x) = \exp(-\lambda x^2/2)$ one observes that a well centered peak transforms into a broadly spread peak and vice versa. This turns out to be a general property of the Fourier transform known as **uncertainty principle**. One quantitative way of measuring this fact is to look at

$$\|(x_j - x^0) f(x)\|_2^2 = \int_{\mathbb{R}^n} (x_j - x^0)^2 |f(x)|^2 d^n x \tag{11.20}$$

which will be small if f is well concentrated around ξ_j in the j'th coordinate direction.

Theorem 11.14 (Heisenberg uncertainty principle)**.** *Suppose $f \in \mathcal{S}(\mathbb{R}^n)$. Then for any $x^0, p^0 \in \mathbb{R}$ we have*

$$\|(x_j - x^0) f(x)\|_2 \|(p_j - p^0) \hat{f}(p)\|_2 \geq \frac{\|f\|_2^2}{2}. \tag{11.21}$$

Proof. Replacing $f(x)$ by $\mathrm{e}^{\mathrm{i} x_j p^0} f(x + x^0 e_j)$ (where e_j is the unit vector into the j'th coordinate direction) we can assume $x^0 = p^0 = 0$ by Lemma 11.2. Using integration by parts we have

$$\|f\|_2^2 = \int_{\mathbb{R}^n} |f(x)|^2 d^n x = - \int_{\mathbb{R}^n} x_j \partial_j |f(x)|^2 d^n x = -2\mathrm{Re} \int_{\mathbb{R}^n} x_j f(x)^* \partial_j f(x) d^n x.$$

Hence, by Cauchy–Schwarz,

$$\|f\|_2^2 \leq 2 \|x_j f(x)\|_2 \|\partial_j f(x)\|_2 = 2 \|x_j f(x)\|_2 \|p_j \hat{f}(p)\|_2$$

the claim follows. □

The name steams from quantum mechanics, where $|f(x)|^2$ is interpreted as the probability distribution for the position of a particle and $|\hat{f}(x)|^2$ is interpreted as the probability distribution for its momentum. Equation (11.21) says that the variance of both distributions cannot both be small and thus one cannot simultaneously measure position and momentum of a particle with arbitrary precision.

Another version states that f and $\hat{f}$ cannot both have compact support.

Theorem 11.15. *Suppose $f \in L^2(\mathbb{R}^n)$. If both f and $\hat{f}$ have compact support, then $f = 0$.*

Proof. Let $A, B \subset \mathbb{R}^n$ be two compact sets and consider the subspace of all functions with $\mathrm{supp}(f) \subseteq A$ and $\mathrm{supp}(\hat{f}) \subseteq B$. Then

$$f(x) = \int_{\mathbb{R}^n} K(x, y) f(y) d^n y,$$

where

$$K(x, y) = \frac{1}{(2\pi)^n} \int_B \mathrm{e}^{\mathrm{i}(x-p)} \chi_A(y) d^n p = (2\pi)^{-n/2} \hat{\chi}_B(y - x) \chi_A(y).$$

Since $K \in L^2(\mathbb{R}^n \times \mathbb{R}^n)$ the corresponding integral operator is Hilbert–Schmidt, and thus its eigenspace corresponding to the eigenvalue 1 can be at most finite dimensional.

Now if there is a nonzero f, we can find a sequence of vectors $x^n \to 0$ such the functions $f_n(x) = f(x - x^n)$ are linearly independent (look at their supports) and satisfy $\mathrm{supp}(f_n) \subseteq 2A$, $\mathrm{supp}(\hat{f}_n) \subseteq B$. But this a contradiction by the first part applied to the sets $2A$ and B. □

Problem 11.1. *Show that $\mathcal{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$. (Hint: If $f \in \mathcal{S}(\mathbb{R}^n)$, then $|f(x)| \le C_m \prod_{j=1}^n (1+x_j^2)^{-m}$ for every m.)*

Problem 11.2. *Show that $\int_{\mathbb{R}} \exp(-x^2/2)dx = \sqrt{2\pi}$. (Hint: Square the integral and evaluate it using polar coordinates.)*

Problem 11.3. *Compute the Fourier transform of the following functions $f : \mathbb{R} \to \mathbb{C}$:*

(i) $f(x) = \chi_{(-1,1)}(x)$. (ii) $f(p) = \frac{1}{p^2+k^2}$, $\mathrm{Re}(k) > 0$.

Problem 11.4. *Suppose $f(x) \in L^1(\mathbb{R})$ and $g(x) = -\mathrm{i}xf(x) \in L^1(\mathbb{R})$. Then $\hat{f}$ is differentiable and $\hat{f}' = \hat{g}$.*

Problem 11.5. *A function $f : \mathbb{R}^n \to \mathbb{C}$ is called* **spherically symmetric** *if it is invariant under rotations; that is, $f(Ox) = f(x)$ for all $O \in SO(\mathbb{R}^n)$ (equivalently, f depends only on the distance to the origin $|x|$). Show that the Fourier transform of a spherically symmetric function is again spherically symmetric.*

Problem 11.6. *Show that $C_\infty(\mathbb{R}^n)$ is indeed a Banach space. Show that $C_c^\infty(\mathbb{R}^n)$ is dense. (Hint: Lemma 8.10.)*

Problem 11.7. *Show that $\mathcal{F} : L^1(\mathbb{R}^n) \to C_\infty(\mathbb{R}^n)$ is not onto as follows:*

(i) *The range of $\mathcal{F}$ is dense.*

(ii) *$\mathcal{F}$ is onto if and only if it has a bounded inverse.*

(iii) *$\mathcal{F}$ has no bounded inverse.*

(Hint for (iii) *in the case $n = 1$: Suppose $\varphi \in C_c^\infty(0,1)$ and set $f_m(x) = \sum_{k=1}^m \mathrm{e}^{\mathrm{i}kx}\varphi(x-k)$. Then $\|f_m\|_1 = m\|\varphi\|_1$ and $\|\hat{f}_m\|_\infty \le const$ since $\varphi \in \mathcal{S}(\mathbb{R})$ and hence $|\varphi(p)| \le const(1+|p|)^{-2}$).*

Problem 11.8 (Wiener)**.** *Suppose $f \in L^2(\mathbb{R}^n)$. Then the set $\{f(x+a) | a \in \mathbb{R}^n\}$ is total in $L^2(\mathbb{R}^n)$ if and only if $\hat{f}(p) \neq 0$ a.e. (Hint: Use Lemma 11.2 and the fact that a subspace is total if and only if its orthogonal complement is zero.)*

Problem 11.9. *Suppose $f(x)\mathrm{e}^{k|x|} \in L^1(\mathbb{R})$ for some $k > 0$. Then $\hat{f}(p)$ has an analytic extension to the strip $|\mathrm{Im}(p)| < k$.*

11.2. Interpolation and the Fourier transform on L^p

If $f \in L^{p_0} \cap L^{p_1}$ for some $p_0 < p_1$ then it is not hard to see that $f \in L^p$ for every $p \in [p_0, p_1]$ and we have (Problem 11.10) the **Lyapunov inequality**

$$\|f\|_p \le \|f\|_{p_0}^{\theta} \|f\|_{p_1}^{1-\theta}, \tag{11.22}$$

where $\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$, $\theta \in (0,1)$. Note that $L^{p_0} \cap L^{p_1}$ contains all integrable simple functions which are dense in L^p for $1 \le p < \infty$ (for $p = \infty$ this is only true if the measure is finite).

This is a first occurrence of an interpolation technique. Moreover, we have defined the Fourier transform as an operator from $L^1 \to L^\infty$ as well as from $L^2 \to L^2$ and it is hence not surprising that this can be used to extend the Fourier transform to the spaces in between.

To this end we begin with a result from complex analysis.

Theorem 11.16 (Hadamard three-lines theorem)**.** *Let S be the open strip $\{z \in \mathbb{C} | 0 < \mathrm{Re}(z) < 1\}$ and let $F : \overline{S} \to \mathbb{C}$ be continuous and bounded on $\overline{S}$ and holomorphic in S. If*

$$|F(z)| \le \begin{cases} M_0, & \mathrm{Re}(z) = 0, \\ M_1, & \mathrm{Re}(z) = 1, \end{cases} \tag{11.23}$$

then

$$|F(z)| \le M_0^{1-\mathrm{Re}(z)} M_1^{\mathrm{Re}(z)} \tag{11.24}$$

for every $z \in \overline{S}$.

Proof. Without loss of generality we can assume $M_0, M_1 > 0$ and after the transformation $F(z) \to M_0^{z-1} M_1^{-z} F(z)$ even $M_0 = M_1 = 1$. Now we consider the auxiliary function

$$F_n(z) = \mathrm{e}^{(z^2-1)/n} F(z)$$

which still satisfies $|F_n(z)| \le 1$ for $\mathrm{Re}(z) = 0$ and $\mathrm{Re}(z) = 1$ since $\mathrm{Re}(z^2 - 1) \le -\mathrm{Im}(z)^2 \le 0$ for $z \in \overline{S}$. Moreover, by assumption $|F(z)| \le M$ implying $|F_n(z)| \le 1$ for $|\mathrm{Im}(z)| \ge \sqrt{\log(M) n}$. Since we also have $|F_n(z)| \le 1$ for $|\mathrm{Im}(z)| \le \sqrt{\log(M) n}$ by the maximum modulus principle we see $|F_n(z)| \le 1$ for all $z \in \overline{S}$. Finally, letting $n \to \infty$ the claim follows. □

Now we are abel to show the **Riesz–Thorin interpolation theorem**

Theorem 11.17 (Riesz–Thorin)**.** *Let $(X, d\mu)$ and $(Y, d\nu)$ measure spaces and $1 \le p_0, p_1, q_0, q_1 \le \infty$. If $q_0 = q_1 = \infty$ assume additionally that ν is σ-finite. If A is a linear operator on*

$$A : L^{p_0}(X, d\mu) \cap L^{p_1}(X, d\mu) \to L^{q_0}(Y, d\nu) \cap L^{q_1}(Y, d\nu) \tag{11.25}$$

satisfying

$$\begin{aligned} \|Af\|_{q_0} &\le M_0\|f\|_{p_0}, \\ \|Af\|_{q_1} &\le M_1\|f\|_{p_1}, \end{aligned} \qquad f \in L^{p_0}(X,d\mu)\cap L^{p_1}(X,d\mu), \tag{11.26}$$

then A *has a continuous extension*

$$A_\theta : L^{p_\theta}(X,d\mu) \to L^{q_\theta}(Y,d\nu), \qquad \frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \ \frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1} \tag{11.27}$$

satisfying $\|A_\theta\| \le M_0^{1-\theta} M_1^{\theta}$ *for every* $\theta \in (0,1)$.

Proof. In the case $p_0 = p_1 = \infty$ the claim is immediate from (11.22) and hence we can assume $p_\theta < \infty$ in which case the space of integrable simple functions is dense in L^{p_θ}. We will also temporarily assume $q_\theta < \infty$. Then, by Lemma 10.1 it suffices to show

$$\left| \int (Af)(y) g(y) d\nu(y) \right| \le M_0^{1-\theta} M_1^{\theta},$$

where f, g are simple functions with $\|f\|_{p_\theta} = \|g\|_{q'_\theta} = 1$ and $\frac{1}{q_\theta} + \frac{1}{q'_\theta} = 1$.

Now choose simple functions $f(x) = \sum_j \alpha_j \chi_{A_j}(x)$, $g(x) = \sum_k \beta_k \chi_{B_k}(x)$ with $\|f\|_1 = \|g\|_1 = 1$ and set $f_z(x) = \sum_j |\alpha_j|^{1/p_z} \operatorname{sign}(\alpha_j)\chi_{A_j}(x)$, $g_z(y) = \sum_k |\beta_k|^{1-1/q_z} \operatorname{sign}(\beta_k)\chi_{B_k}(y)$ such that $\|f_z\|_{p_\theta} = \|g_z\|_{q'_\theta} = 1$ for $\theta = \operatorname{Re}(z) \in [0,1]$. Moreover, note that both functions are entire and thus the function

$$F(z) = \int (Af_z)(y) g_z d\nu(y)$$

satisfies the assumptions of the three-lines theorem.

It remains to consider the case $p_0 < p_1$ and $q_0 = q_1 = \infty$. In this case we can proceed as before using again Lemma 10.1 and a simple function for $g = g_z$. □

As a consequence we get two important inequalities:

Corollary 11.18 (Hausdorff–Young inequality). *The Fourier transform extends to a continuous map* $\mathcal{F} : L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$, *for* $1 \le p \le 2$, $\frac{1}{p} + \frac{1}{q} = 1$, *satisfying*

$$(2\pi)^{-n/(2q)} \|\hat{f}\|_q \le (2\pi)^{-n/(2p)} \|f\|_p. \tag{11.28}$$

We remark that the Fourier transform does not extend to a continuous map $\mathcal{F} : L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$, for $p > 2$ (Problem 11.11). Moreover, its range is dense for $1 < p \le 2$ but not all of $L^q(\mathbb{R}^n)$ unless $p = q = 2$.

Corollary 11.19 (Young inequality). *Let* $f \in L^p(\mathbb{R}^n)$ *and* $g \in L^q(\mathbb{R}^n)$ *with* $\frac{1}{p} + \frac{1}{q} \ge 1$. *Then* $f(y)g(x-y)$ *is integrable with respect to* y *for a.e.* x *and the convolution satisfies* $f * g \in L^r(\mathbb{R}^n)$ *with*

$$\|f * g\|_r \le \|f\|_p \|g\|_q, \tag{11.29}$$

where $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$.

Proof. We consider the operator $A_g f = f * g$ which satisfies $\|A_g f\|_q \le \|g\|_q \|f\|_1$ for every $f \in L^1$ by Problem 8.10. Similarly, Hölder's inequality implies $\|A_g f\|_\infty \le \|g\|_q \|f\|_{q'}$ for every $f \in L^{q'}$, where $\frac{1}{q} + \frac{1}{q'} = 1$. Hence the Riesz–Thorin theorem implies that A_g extends to an operator $A_q : L^p \to L^r$, where $\frac{1}{p} = \frac{1-\theta}{1} + \frac{\theta}{q'} = 1 - \frac{\theta}{q}$ and $\frac{1}{r} = \frac{1-\theta}{q} + \frac{\theta}{\infty} = \frac{1}{p} + \frac{1}{q} - 1$. To see that $f(y)g(x-y)$ is integrable a.e. consider $f_n(x) = \chi_{|x|\le n}(x) \max(n, |f(x)|)$. Then the convolution $(f_n * |g|)(x)$ is finite and converges for every x by monotone convergence. Moreover, since $f_n \to |f|$ in L^p we have $f_n * |g| \to A_g f$ in L^r, which finishes the proof. □

Problem 11.10. *Show* (11.22)*. (Hint: Generalized Hölder inequality, Problem* 8.5*.)*

Problem 11.11. *Show that the Fourier transform does not extend to a continuous map* $\mathcal{F} : L^p(\mathbb{R}^n) \to L^q(\mathbb{R}^n)$*, for* $p > 2$*. Use the closed graph theorem to conclude that* $\mathcal{F}$ *is not onto for* $1 \le p \le 2$*. (Hint for the case* $n = 1$*: Consider* $\phi_z(x) = \exp(-zx^2/2)$ *for* $z = \lambda + \mathrm{i}\omega$ *with* $\lambda > 0$*.)*

11.3. Sobolev spaces

We begin by introducing the **Sobolev space**

$$H^r(\mathbb{R}^n) = \{f \in L^2(\mathbb{R}^n) | |p|^r \hat{f}(p) \in L^2(\mathbb{R}^n)\}. \tag{11.30}$$

The most important case is when r is an integer, however our definition makes sense for any $r \ge 0$. Moreover, note that $H^r(\mathbb{R})$ becomes a Hilbert space if we introduce the scalar product

$$\langle f, g\rangle = \int_{\mathbb{R}^n} \hat{f}(p)^* \hat{g}(p)(1 + |p|^2)^r d^n p. \tag{11.31}$$

In particular, note that by construction $\mathcal{F}$ maps $H^r(\mathbb{R}^n)$ unitarily onto $L^2(\mathbb{R}^n, (1 + |p|^2)^r d^n p)$. Clearly $H^r(\mathbb{R}^n) \subset H^{r+1}(\mathbb{R}^n)$ with the embedding being continuous. Moreover, $\mathcal{S}(\mathbb{R}^n) \subset H^r(\mathbb{R}^n)$ and this subset is dense (since $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n, (1 + |p|^2)^r d^n p)$).

The motivation for the definition (11.30) stems from Lemma 11.4 which allows us to extend differentiation to a larger class. In fact, every function in $H^r(\mathbb{R}^n)$ has partial derivatives up to order $\lfloor r \rfloor$, which are defined via

$$\partial_\alpha f = ((\mathrm{i}p)^\alpha \hat{f}(p))^\vee, \qquad f \in H^r(\mathbb{R}^n), \ |\alpha| \le r. \tag{11.32}$$

Example. Consider $f(x) = (1 - |x|)\chi_{[-1,1]}(x)$. Then $\hat{f}(p) = \sqrt{\frac{2}{\pi}} \frac{\cos(p) - 1}{p^2}$ and $f \in H^1(\mathbb{R})$. The weak derivative is $f'(x) = -\operatorname{sign}(x)\chi_{[-1,1]}(x)$. ⋄

By Lemma 11.4 this definition coincides with the usual one for every $f \in \mathcal{S}(\mathbb{R}^n)$ and we have

$$\begin{aligned}\int_{\mathbb{R}^n} g(x)(\partial_\alpha f)(x) d^n x &= \langle g^*, (\partial_\alpha f)\rangle = \langle \hat{g}(p)^*, (\mathrm{i}p)^\alpha \hat{f}(p)\rangle \\ &= (-1)^{|\alpha|}\langle (\mathrm{i}p)^\alpha \hat{g}(p)^*, \hat{f}(p)\rangle = (-1)^{|\alpha|}\langle \partial_\alpha g^*, f\rangle \\ &= (-1)^{|\alpha|}\int_{\mathbb{R}^n} (\partial_\alpha g)(x) f(x) d^n x, \qquad (11.33)\end{aligned}$$

for $f, g \in H^r(\mathbb{R}^n)$. Furthermore, recall that a function $h \in L^1_{loc}(\mathbb{R}^n)$ satisfying

$$\int_{\mathbb{R}^n} \varphi(x) h(x) d^n x = (-1)^{|\alpha|}\int_{\mathbb{R}^n} (\partial_\alpha \varphi)(x) f(x) d^n x, \quad \forall \varphi \in C_c^\infty(\mathbb{R}^n), \quad (11.34)$$

is also called the **weak derivative** or the derivative in the sense of distributions of f (by Lemma 8.12 such a function is unique if it exists). Hence, choosing $g = \varphi$ in (11.33), we see that $H^r(\mathbb{R}^n)$ is the set of all functions having partial derivatives (in the sense of distributions) up to order r, which are in $L^2(\mathbb{R}^n)$.

In this connection the following norm for $H^m(\mathbb{R}^n)$ with $m \in \mathbb{N}_0$ is more common:

$$\|f\|_{2,m}^2 = \sum_{|\alpha| \le m} \|\partial_\alpha f\|_2^2. \qquad (11.35)$$

By $|p^\alpha| \le |p|^{|\alpha|} \le (1 + |p|^2)^{m/2}$ it follows that this norm is equivalent to (11.31).

Of course a natural question to ask is when the weak derivatives are in fact classical derivatives. To this end observe that the Riemann–Lebesgue lemma implies that $\partial_\alpha f(x) \in C_\infty(\mathbb{R}^n)$ provided $p^\alpha f(p) \in L^1(\mathbb{R}^n)$. Moreover, in this situation the derivatives will exist as classical derivatives:

Lemma 11.20. *Suppose $f \in L^1(\mathbb{R}^n)$ or $f \in L^2(\mathbb{R}^n)$ with $(1 + |p|^k)\hat{f}(p) \in L^1(\mathbb{R}^n)$ for some $k \in \mathbb{N}_0$. Then $f \in C_\infty^k(\mathbb{R}^n)$, the set of functions with continuous partial derivatives of order k all of which vanish at ∞. Moreover,*

$$(\partial_\alpha f)^\wedge(p) = (\mathrm{i}p)^\alpha \hat{f}(p), \qquad |\alpha| \le k, \qquad (11.36)$$

in this case.

Proof. We begin by observing that by Theorem 11.6

$$f(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}px} \hat{f}(p) d^n p.$$

Now the claim follows as in the proof of Lemma 11.4 by differentiating the integral using Problem 7.16. □

Now we are able to prove the following embedding theorem.

Theorem 11.21 (Sobolev embedding)**.** *Suppose $r > k + \frac{n}{2}$ for some $k \in \mathbb{N}_0$. Then $H^r(\mathbb{R}^n)$ is continuously embedded into $C^k_\infty(\mathbb{R}^n)$ with*

$$\|\partial_\alpha f\|_\infty \le C_{n,r}\|f\|_{2,r}, \qquad |\alpha| \le k. \tag{11.37}$$

Proof. Abbreviate $\langle p\rangle = (1+|p|^2)^{1/2}$. Now use $|(\mathrm{i}p)^\alpha \hat{f}(p)| \le \langle p\rangle^{|\alpha|}|\hat{f}(p)| = \langle p\rangle^{-s}\cdot\langle p\rangle^{|\alpha|+s}|\hat{f}(p)|$. Now $\langle p\rangle^{-s} \in L^2(\mathbb{R}^n)$ if $s > \frac{n}{2}$ (use polar coordinates to compute the norm) and $\langle p\rangle^{|\alpha|+s}|\hat{f}(p)| \in L^2(\mathbb{R}^n)$ if $s + |\alpha| \le r$. Hence the claim follows from the previous lemma. □

In fact, we can even do a bit better.

Lemma 11.22 (Morrey inequality)**.** *Suppose $f \in H^{n/2+\gamma}(\mathbb{R}^n)$ for some $\gamma \in (0,1)$. Then $f \in C^{0,\gamma}_\infty(\mathbb{R}^n)$, the set of functions which are Hölder continuous with exponent γ and vanish at ∞. Moreover,*

$$|f(x) - f(y)| \le C_{n,\gamma}\|\hat{f}(p)\|_{2,n/2+\gamma}|x-y|^\gamma \tag{11.38}$$

in this case.

Proof. We begin with

$$f(x+y) - f(x) = \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}px}(\mathrm{e}^{\mathrm{i}py}-1)\hat{f}(p)d^np$$

implying

$$|f(x+y) - f(x)| \le \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \frac{|\mathrm{e}^{\mathrm{i}py}-1|}{\langle p\rangle^{n/2+\gamma}}\langle p\rangle^{n/2+\gamma}|\hat{f}(p)|d^np.$$

where again $\langle p\rangle = (1+|p|^2)^{1/2}$. Hence after applying Cauchy–Schwarz it remains to estimate (recall (7.56))

$$\begin{aligned}\int_{\mathbb{R}^n} \frac{|\mathrm{e}^{\mathrm{i}py}-1|^2}{\langle p\rangle^{n+2\gamma}}d^np \le{}& S_n\int_0^{1/|y|}\frac{(|y|r)^2}{\langle r\rangle^{n+2\gamma}}r^{n-1}dr\\ &+ S_n\int_{1/|y|}^\infty \frac{4}{\langle r\rangle^{n+2\gamma}}r^{n-1}dr\\ \le{}& \frac{S_n}{2(1-\gamma)}|y|^{2\gamma} + \frac{S_n}{2\gamma}|y|^{2\gamma} = \frac{S_n}{2\gamma(1-\gamma)}|y|^{2\gamma},\end{aligned}$$

where $S_n = nV_n$ is the volume of the unit sphere in $\mathbb{R}^n$. □

Using this lemma we immediately obtain:

Corollary 11.23. *Suppose $r \ge k + \gamma + \frac{n}{2}$ for some $k \in \mathbb{N}_0$ and $\gamma \in (0,1)$. Then $H^r(\mathbb{R}^n)$ is continuously embedded into $C^{k,\gamma}_\infty(\mathbb{R}^n)$ the set of functions in $C^k_\infty(\mathbb{R}^n)$ whose highest derivatives are Hölder continuous of exponent γ.*

Problem 11.12. *Suppose* $f \in L^2(\mathbb{R}^n)$ *show that* $\varepsilon^{-1}(f(x+e_j\varepsilon)-f(x)) \to g_j(x)$ *in* L^2 *if and only if* $p_j\hat{f}(p) \in L^2$, *where* e_j *is the unit vector into the* j*'th coordinate direction. Moreover, show* $g_j = \partial_j f$ *if* $f \in H^1(\mathbb{R}^n)$.

Problem 11.13. *Show that* u *is weakly differentiable in the interval* $(0,1)$ *if and only if* u *is absolutely continuous and* $u' = v$ *in this case. (Hint: You will need that* $\int_0^1 u(t)\varphi'(t)dt = 0$ *for all* $\varphi \in C_c^\infty(0,1)$ *if and only if* u *is constant. To see this choose some* $\varphi_0 \in C_c^\infty(0,1)$ *with* $I(\varphi_0) = \int_0^1 \varphi_0(t)dt = 1$. *Then invoke Lemma 8.12 and use that every* $\varphi \in C_c^\infty(0,1)$ *can be written as* $\varphi(t) = \Phi'(t) + I(\varphi)\varphi_0(t)$ *with* $\Phi(t) = \int_0^t \varphi(s)ds - I(\varphi)\int_0^t \varphi_0(s)ds$.*)*

Zorn's lemma

A **partial order** is a binary relation "$\preceq$" over a set $\mathcal{P}$ such that for all $A, B, C \in \mathcal{P}$:

- $A \preceq A$ (reflexivity),
- if $A \preceq B$ and $B \preceq A$ then $A = B$ (antisymmetry),
- if $A \preceq B$ and $B \preceq C$ then $A \preceq C$ (transitivity).

Example. Let $\mathcal{P}(X)$ be the collections of all subsets of a set X. Then $\mathcal{P}$ is partially ordered by inclusion $\subseteq$. $\diamond$

It is important to emphasize that two elements of $\mathcal{P}$ need not be comparable, that is, in general neither $A \preceq B$ nor $B \preceq A$ might hold. However, if any two elements are comparable, $\mathcal{P}$ will be called **totally ordered**.

Example. $\mathbb{R}$ with $\leq$ is totally ordered. $\diamond$

If $\mathcal{P}$ is partially ordered, then every totally ordered subset is also called a **chain**. If $\mathcal{Q} \subseteq \mathcal{P}$, then an element $M \in \mathcal{P}$ satisfying $A \preceq M$ for all $A \in \mathcal{Q}$ is called an **upper bound**.

Example. Let $\mathcal{P}(X)$ as before. Then a collection of subsets $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{P}(X)$ satisfying $A_n \subseteq A_{n+1}$ is a chain. The set $\bigcup_n A_n$ is an upper bound. $\diamond$

An element $M \in \mathcal{P}$ for which $M \preceq A$ for some $A \in \mathcal{P}$ is only possible if $M = A$ is called a **maximal element**.

Theorem A.1 (Zorn's lemma)**.** *Every partially ordered set in which every chain has an upper bound contains at least one maximal element.*

Zorn's lemma is one of the equivalent incarnations of the **Axiom of Choice** and we are going to take it, as well as the rest of set theory, for granted.

Bibliography

[1] M. Abramovitz, I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1972.

[2] H. W. Alt, Lineare Funktionalanalysis, 4th ed., Springer, Berlin, 2002.

[3] H. Bauer, Measure and Integration Theory, de Gruyter, Berlin, 2001.

[4] J.L. Kelly, *General Topology*, Springer, New York, 1955.

[5] E. Lieb and M. Loss, *Analysis*, Amer. Math. Soc., Providence, 1997.

[6] I. K. Rana, *An Introduction to Measure and Integration*, 2nd ed., Amer. Math. Soc., Providence, 2002.

[7] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I. Functional Analysis*, rev. and enl. edition, Academic Press, San Diego, 1980.

[8] J. R. Retherford, *Hilbert Space: Compact Operators and the Trace Theorem*, Cambridge University Press, Cambridge, 1993.

[9] H. Royden, *Real Analysis*, Prencite Hall, New Jersey, 1988.

[10] W. Rudin, *Real and Complex Analysis*, 3rd edition, McGraw-Hill, New York, 1987.

[11] J. Weidmann, *Lineare Operatoren in Hilberträumen I: Grundlagen*, B.G.Teubner, Stuttgart, 2000.

[12] D. Werner, *Funktionalanalysis*, 3rd edition, Springer, Berlin, 2000.

Glossary of notation

$AC[a,b]$	...absolutely continuous functions, 170
$\arg(z)$	...argument of $z \in \mathbb{C}$; $\arg(z) \in (-\pi, \pi]$, $\arg(0) = 0$
$B_r(x)$	...open ball of radius r around x, 7
$B(X)$	...Banach space of bounded measurable functions
$BV[a,b]$	...functions of bounded variation, 168
$\mathfrak{B}$	$= \mathfrak{B}^1$
$\mathfrak{B}^n$	...Borel σ-field of $\mathbb{R}^n$, 106
$\mathbb{C}$	...the set of complex numbers
$\mathfrak{C}(\mathfrak{H})$	...set of compact operators, 51
$C(U)$	...set of continuous functions from U to $\mathbb{C}$
$C(U,V)$	...set of continuous functions from U to V
$C_\infty(\mathbb{R}^n)$	...set of continuous functions vanishing at ∞, 184
$C_c^\infty(U,V)$	...set of compactly supported smooth functions
$\chi_\Omega(.)$	...characteristic function of the set Ω
dim	...dimension of a vector space
$\mathrm{dist}(x,Y)$	$= \inf_{y\in Y} \|x-y\|$, distance between x and Y
$\mathfrak{D}(.)$	...domain of an operator
e	...exponential function, $\mathrm{e}^z = \exp(z)$
$\mathfrak{H}$	...a Hilbert space
i	...complex unity, $\mathrm{i}^2 = -1$
Im(.)	...imaginary part of a complex number
inf	...infimum
$\mathrm{Ker}(A)$	...kernel of an operator A, 30
$\mathfrak{L}(X,Y)$	...set of all bounded linear operators from X to Y, 32
$\mathfrak{L}(X)$	$= \mathfrak{L}(X,X)$
$L^p(X,d\mu)$	...Lebesgue space of p integrable functions, 144

$L^\infty(X, d\mu)$	... Lebesgue space of bounded functions, 144
$L^p_{loc}(X, d\mu)$	... locally p integrable functions, 151
$\mathcal{L}^1(X)$	... space of integrable functions, 124
$\mathcal{L}^2_{cont}(I)$	... space of continuous square integrable functions, 27
$\ell^p(\mathbb{N})$	... Banach space of p summable sequences, 19
$\ell^2(\mathbb{N})$	... Hilbert space of square summable sequences, 24
$\ell^\infty(\mathbb{N})$	... Banach space of bounded summable sequences, 19
max	... maximum
$\mathbb{N}$	... the set of positive integers
$\mathbb{N}_0$	$= \mathbb{N} \cup \{0\}$
$\mathbb{Q}$	... the set of rational numbers
$\mathbb{R}$	... the set of real numbers
$\mathrm{Ran}(A)$	... range of an operator A, 30
$\mathrm{Re}(.)$	... real part of a complex number
$\mathrm{sign}(z)$	$= z/\lvert z\rvert$ for $z \neq 0$ and 1 for $z = 0$; complex sign function
sup	... supremum
supp	... support of a function, 12
$\mathrm{span}(M)$	... set of finite linear combinations from M, 20
$\mathbb{Z}$	... the set of integers
$\mathbb{I}$	... identity operator
$\sqrt{z}$	... square root of z with branch cut along $(-\infty, 0)$
z^*	... complex conjugation
$\Vert.\Vert$	... norm, 24
$\Vert.\Vert_p$	... norm in the Banach space ℓ^p and L^p, 19, 143
$\langle ., ..\rangle$	... scalar product in $\mathfrak{H}$, 24
$\oplus$	... orthogonal sum of vector spaces or operators, 48
$\otimes$	... tensor product, 49
$\lfloor x \rfloor$	$= \max\{n \in \mathbb{Z} \mid n \leq x\}$ floor function
∂	... gradient
∂_α	... partial derivative
$M^\perp$	... orthogonal complement, 44
(λ_1, λ_2)	$= \{\lambda \in \mathbb{R} \,\vert\, \lambda_1 < \lambda < \lambda_2\}$, open interval
$[\lambda_1, \lambda_2]$	$= \{\lambda \in \mathbb{R} \,\vert\, \lambda_1 \leq \lambda \leq \lambda_2\}$, closed interval

Index

a.e., *see* almost everywhere
absolute convergence, 23
absolutely continuous
  function, 170
  measure, 155
accumulation point, 8
algebra, 105
almost everywhere, 111

B.L.T. theorem, 31
Baire category theorem, 61
Banach algebra, 32, 89
Banach space, 18
Banach–Steinhaus theorem, 62
base, 9
basis, 20
  orthonormal, 41
Bessel inequality, 40
bidual space, 70
bijective, 11
Bolzano–Weierstraß theorem, 15
Borel
  function, 117
  measure, 108
    regular, 108
  set, 106
  σ-algebra, 106
boundary condition, 6
boundary value problem, 6
bounded
  operator, 30
  sesquilinear form, 29
  set, 15
bounded variation, 169

Cantor
  function, 162
  measure, 162
  set, 111
Cauchy sequence, 10
Cauchy–Schwarz–Bunjakowski inequality, 25
characteristic function, 121
Chebyshev's inequality, 163, 168
closed set, 9
closure, 9
cluster point, 8
compact, 13
  locally, 15
  sequentially, 14
complete, 10, 18
completion, 29
content, 177
continuous, 11
convergence, 10
convex, 145
convolution, 151, 185
cover, 12
C^* algebra, 94

dense, 10
diffusion equation, 3
dimension, 43
Dirac measure, 110, 125
direct sum, 33
discrete set, 8
discrete topology, 8
distance, 7, 15
distribution function, 108
domain, 30
dominated convergence theorem, 125
double dual, 70

Dynkin system, 112
Dynkin's π-λ theorem, 113

eigenspace, 53
eigenvalue, 53
  simple, 54
eigenvector, 53
equivalent norms, 27
essential supremum, 144
Extreme value theorem, 15

finite intersection property, 13
form
  bounded, 29
Fourier
  transform, 181
Fourier series, 42
Fréchet space, 36
Fredholm alternative, 86
Fredholm operator, 87
Fubini theorem, 128
function
  open, 12
fundamental theorem of calculus, 171

generalized inverse, 135
gradient, 183
Gram–Schmidt orthogonalization, 42
graph, 64
graph norm, 66
Green function, 58

Hahn decomposition, 165
Hardy–Littlewood maximal function, 162
Hausdorff space, 9
Hausdorff–Young inequality, 190
heat equation, 3
Heine–Borel theorem, 15
Heisenberg uncertainty principle, 187
Hilbert space, 24
  dimension, 43
Hilbert–Schmidt
  operator, 153
Hilbert–Schmidt operator, 82
Hölder continuous, 36
Hölder's inequality, 19, 147
homeomorphism, 12

identity, 32, 89
index, 87
induced topology, 9
injective, 11
inner product, 24
inner product space, 24
integrable, 124
  Riemann, 139
integral, 121
integration by parts, 172
integration by substitution, 137
interior, 10
interior point, 7
involution, 94
isolated point, 8

Jensen's inequality, 146

kernel, 30
Kuratowski closure axioms, 10

λ-system, 112
Lebesgue
  decomposition, 157
  measure, 110
  point, 160
Lebesgue–Stieltjes measure, 108
lemma
  Riemann-Lebesgue, 184
Lidskij trace theorem, 84
limit point, 8
Lindelöf theorem, 13
linear
  functional, 32, 45
  operator, 30
linearly independent, 20
Lipschitz continuous, 36
lower semicontinuous, 118
Lyapunov inequality, 189

maximum norm, 17
measurable
  function, 117
  set, 106
  space, 106
measure, 106
  absolutely continuous, 155
  complete, 115
  complex, 163
  finite, 106
  Lebesgue, 110
  minimal support, 162
  mutually singular, 155
  product, 127
  space, 106
  support, 110
metric space, 7
Minkowski inequality, 147
mollifier, 150
monotone convergence theorem, 122
Morrey inequality, 193
multi-index, 35, 182
  order, 35, 182

neighborhood, 8
Neumann series, 92

Noether operator, 87
norm, 17
  operator, 30
normal, 16, 95
normalized, 25
normed space, 17
nowhere dense, 61
null space, 30

onto, 11
open
  ball, 7
  function, 12
  set, 8
operator
  adjoint, 46
  bounded, 30
  closeable, 65
  closed, 64
  closure, 65
  compact, 51
  domain, 30
  finite rank, 80
  Hilbert–Schmidt, 153
  linear, 30
  nonnegative, 47
  self-adjoint, 53
  strong convergence, 76
  symmetric, 53
  unitary, 43
  weak convergence, 76
order
  partial, 195
  total, 195
orthogonal, 25
orthogonal complement, 44
orthogonal projection, 45
orthogonal projections, 98
orthogonal sum, 48
outer measure, 114

parallel, 25
parallelogram law, 26
Parseval relation, 42
partial order, 195
partition, 138
partition of unity, 16
perpendicular, 25
π-system, 113
Plancherel identity, 184
polar coordinates, 131
polarization identity, 26
premeasure, 107
product measure, 127
product rule, 173
product topology, 12
projection valued measure, 98
pseudometric, 7
Pythagorean theorem, 25

quadrangle inequality, 17
quotient space, 33

Radon measure, 120
Radon–Nikodym
  derivative, 157
  theorem, 157
range, 30
rank, 80
reflexive, 70
relative σ-algebra, 106
relative topology, 9
resolution of the identity, 98
resolvent, 57, 91
resolvent set, 91
Riemann integrable, 139
Riemann integral, 139
  lower, 139
  upper, 139
Riesz lemma, 45

scalar product, 24
Schatten p-class, 82
Schauder basis, 20
Schur criterion, 152
second countable, 9
self-adjoint, 95
seminorm, 18
separable, 10, 21
separation of variables, 4
series
  absolutely convergent, 23
sesquilinear form, 24
  bounded, 29
  parallelogram law, 28
  polarization identity, 29
σ-algebra, 105
σ-finite, 106
signed measure, 165
simple function, 121
singular values, 79
Sobolev space, 191
span, 20
spectral measure, 96
spectral projections, 98
spectral radius, 92
spectrum, 91
spherical coordinates, 132
spherically symmetric, 188
$*$-subalgebra, 94
Stone–Weierstraß theorem, 100
strong convergence, 76
Sturm–Liouville problem, 6
subcover, 12

subspace topology, 9
support, 12
surjective, 11

tensor product, 49
theorem
  Arzelà–Ascoli, 52
  B.L.T., 31
  Bair, 61
  Banach–Steinhaus, 62
  Bolzano–Weierstraß, 15
  closed graph, 64
  dominated convergence, 125
  Dynkin's π-λ, 113
  Fatou, 123
  Fubini, 128
  fundamental thm. of calculus, 171
  Hadamard three-lines, 189
  Hahn–Banach, 68
  Heine–Borel, 15
  Hellinger–Toeplitz, 66
  Jordan, 169
  Jordan–von Neumann, 26
  Lax–Milgram, 48
  Lebesgue, 125, 140
  Lebesgue decomposition, 157
  Levi, 122
  Lindelöf, 13
  monotone convergence, 122
  open mapping, 63
  Plancherel, 185
  Pythagorean, 25
  Radon–Nikodym, 157
  Riesz representation, 179
  Riesz–Fischer, 148
  Riesz–Thorin, 189
  Schur, 152
  Sobolev embedding, 193
  Stone–Weierstraß, 100
  Urysohn, 16
  Weierstraß, 15, 22
  Wiener, 188
  Zorn, 195
topological space, 8
topology
  base, 9
  product, 12
total, 21
total order, 195
total variation, 163, 169
trace, 84
  class, 82
trace topology, 9
triangel inequality, 7, 17
  inverse, 7, 18
trivial topology, 8

uncertainty principle, 187
uniform boundedness principle, 62
uniformly convex space, 28
unit vector, 25
unitary, 95
upper semicontinuous, 118
Urysohn lemma, 16

vague convergence
  measures, 180
Vandermonde determinant, 23
variation, 168
Vitali set, 112

wave equation, 5
weak
  derivative, 192
weak convergence, 73
weak topology, 73
weak-$*$ convergence, 78
weak-$*$ topology, 78
Weierstraß approximation, 22
Weierstraß theorem, 15
Wiener covering lemma, 158

Young inequality, 151, 185, 190

Zorn's lemma, 195