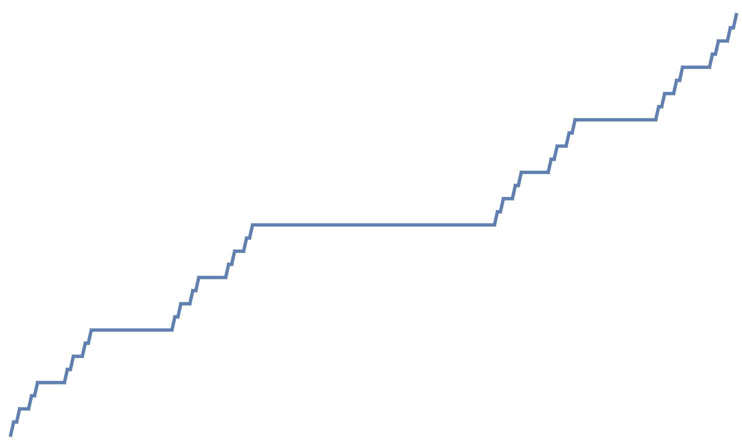
Topics in Real Analysis

Gerald Teschl



Graduate Studies
in Mathematics

Volume (to appear)

American Mathematical Society
Providence, Rhode Island

Gerald Teschl
Fakultät für Mathematik
Oskar-Mogenstern-Platz 1
Universität Wien
1090 Wien, Austria

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

2010 Mathematics subject classification. 28-01, 42-01

Abstract. This manuscript provides a brief introduction to Real Analysis. It covers basic measure theory including Lebesgue and Sobolev spaces and the Fourier transform.

Keywords and phrases. Measure theory, Lebesgue spaces, Sobolev spaces, Fourier transform

Typeset by $\mathcal{AMS}$-LaTeX and Makeindex.
Version: April 26, 2024

Contents

Preface

The present manuscript was written for my course *Real Analysis* given in summer 2011, 2013, and 2018. It assumes some basic familiarity with Functional Analysis, for which there is an accompanying part [**22**], where these topics are covered.

It is updated whenever I find some errors and extended from time to time. 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-ra/`

Goals

The main goal of the present book is to give students a concise introduction which gets to some interesting results without much ado while using a sufficiently general approach suitable for further studies. Still I have tried to always start with some interesting special cases and then work my way up to the general theory. While this unavoidably leads to some duplications, it usually provides much better motivation and implies that the core material always comes first (while the more general results are then optional). Moreover, this book is *not* written under the assumption that it will be read linearly starting with the first chapter and ending with the last. Consequently, I have tried to separate core and optional materials as much as possible while keeping the optional parts as independent as possible.

Furthermore, my aim is not to present an encyclopedic treatment but to provide the reader with a versatile toolbox for further study. Moreover, in contradistinction to many other books, I do not have a particular direction in mind and hence I am trying to give a broad introduction which should prepare you for diverse fields such as spectral theory, partial differential equations, or probability theory. This is related to the fact that I am working in mathematical physics, an area where you never know what mathematical theory you will need next.

I have tried to keep a balance between verbosity and clarity in the sense that I have tried to provide sufficient detail for being able to follow the arguments but without drowning the key ideas in boring details. In particular, you will find a *show this* from time to time encouraging the reader to check the claims made (these tasks typically involve only simple routine calculations). Moreover, to make the presentation student friendly, I have tried to include many worked out examples within the main text. Some of them are standard counterexamples pointing out the limitations of theorems (and explaining why the assumptions are important). Others show how to use the theory in the investigation of practical examples.

Preliminaries

The present manuscript is intended to be gentle when it comes to required background. Of course I assume basic familiarity with analysis (real and complex numbers, limits, differentiation, basic (Riemann) integration, open sets) and linear algebra (finite dimensional vector spaces, matrices).

Apart from this natural assumptions I also expect some familiarity with metric spaces and point set topology. However, only a few basic things are required to begin with. All (and much more) is collected in Appendix B from [**22**] Moreover, you should already know what a Banach/Hilbert space is, but Chapter 1 from [**22**] will be sufficient to get you started.

Content

Below follows a short description of each chapter together with some hints which parts can be skipped.

Chapter 1 introduces the concept of a measure and constructs Borel measures on $\mathbb{R}^n$ via distribution functions (the case of $n = 1$ is done first) which should meet the needs of partial differential equations, spectral theory, and probability theory. I have chosen the Carathéodory approach because I feel that it is the most versatile one.

Chapter 2 discusses the core results of integration theory including the change of variables formula, surface measure and the Gauss–Green theorem. The latter two are only done in a smooth (C^1) setting, but these topics are needed in the chapter on Sobolev spaces and I wanted to be self-contained here. Two optional appendices discuss transforming one-dimensional measures (which should be useful in both spectral theory and probability theory) and the connection with the Riemann integral.

Chapter 3 contains the core material on L^p spaces including basic inequalities and ends with an optional section on integral operators.

Chapter 4 collects some further results. Except for the first section, the results are mostly optional and independent of each other. Lebesgue points discussed in the second section are also used at some places later on. There is also a final section on functions of bounded variation and absolutely continuous functions.

Chapter 5 collects even further results. Again the results are mostly optional and independent of each other.

Chapter 6 discusses the dual space of L^p for $1 \le p < \infty$ and $p = \infty$ as well as some variants of the Riesz–Markov representation theorem.

Chapter 7 contains some core material on Sobolev spaces. I feel that it should be sufficient as a background for applications to partial differential equations (PDE).

Chapter 8 covers the Fourier transform on $\mathbb{R}^n$. It also gives an independent approach to L^2 based Sobolev spaces on $\mathbb{R}^n$. Of course the Fourier transform is vital in the treatment of constant coefficient PDE. However, in many introductory courses some of the technical details are swept under the carpet. I try to discuss these examples with full rigor.

Chapter 9 finally discusses two basic interpolation techniques including applications to Calderón–Zygmund operators.

Sometimes also the historic development of the subject is of interest. This is however not covered in the present book.

To the teacher

The book can either serve as a succinct introduction to the Lebesgue integral or as a full course focusing either on measure theory (covering Chapters 1 to 6 and possibly also Chapters 7 or 8) or on real analysis (skipping Chapter 4 except Section 4.1, 4.8 and including Chapters 7 and 8).

For a brief introduction to Lebesgue spaces, it suffices to cover Chapters 1,2, and 3. If one does not want to provide full proofs for everything, I recommend to cover Section 1.2 (Section 1.1 contains just motivation), give an outline of Section 1.3 (by covering Dynkin's π-λ theorem, the uniqueness theorem for measures, and then quoting the existence theorem for Lebesgue measure), cover Section 1.5. The core material from Chapter 2 are the first two sections and from Chapter 3 the first three sections.

Problems relevant for the main text are marked with a "*". A Solutions Manual will be available electronically for instructors only.

Acknowledgments

I wish to thank my readers, Kerstin Ammann, Phillip Bachler, Batuhan Bayır, Alexander Beigl, Mikhail Botchkarev, Ho Boon Suan, Peng Du, Christian Ekstrand, Damir Ferizović, Michael Fischer, Raffaello Giulietti, Melanie Graf, Josef Greilhuber, Julian Grüber, Matthias Hammerl, Jona Marie Hassenbach, Nobuya Kakehashi, Jerzy Knopik, Nikolas Knotz, Florian Kogelbauer, Helge Krüger, Reinhold Küstner, Oliver Leingang, Juho Leppäkangas, Joris Mestdagh, Alice Mikikits-Leitner, Claudiu Mîndrilă, Jakob Möller, Caroline Moosmüller, Frederick Moscatelli, Matthias Ostermann, Piotr Owczarek, Jakub Petr, Martina Pflegpeter, Mateusz Piorkowski, Tobias Preinerstorfer, Maximilian H. Ruep, Tidhar Sariel, Christian Schmid, Laura Shou, Bertram Tschiderer, Vincent Valmorin, David Wallauch, Richard Welke, David Wimmesberger, Song Xiaojun, Markus Youssef, Rudolf Zeidler, and colleagues Pierre-Antoine Absil, Roland Donninger, Nils C. Framstad, Fritz Gesztesy, Heinz Hanßmann, Günther Hörmann, Aleksey Kostenko, Wallace Lam, Daniel Lenz, Johanna Michor, Viktor Qvarfordt, Alex Strohmaier, David C. Ullrich, Hendrik Vogt, Marko Stautz, Maxim Zinchenko who have pointed out several typos and made useful suggestions for improvements. I am also grateful to Volker Enß for making his lecture notes on nonlinear Functional Analysis available to me.

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
March, 2020

Chapter 1

Measures

1.1. Prelude: The problem of measuring sets

The Riemann[1] integral starts straight with the definition of the integral by considering functions which can be sandwiched between step functions. This is based on the idea that for a function defined on an interval (or a rectangle in higher dimensions) the domain can be easily subdivided into smaller intervals (or rectangles). Moreover, for nice functions the variation of the values (difference between maximum and minimum) should decrease with the length of the intervals. Of course, this fails for rough functions whose variations cannot be decreased by subdividing the domain into sets of decreasing size. The Lebesgue[2] integral remedies this by subdividing the range of the function. This shifts the problem from controlling the variations of the function to defining the content of the preimage of the subdivisions for the range. Note that this problem does not occur in the Riemann approach since only the length of an interval (or the area of an rectangle) is needed. Consequently, the outset of Lebesgue theory is the problem of defining the content for a sufficiently large class of sets.

The Riemann-style approach to this problem in $\mathbb{R}^n$ is to start with a big rectangle containing the set under consideration and then take subdivisions thereby approximating the measure of the set from the inside and outside by the measure of the rectangles which lie inside and those which cover the set, respectively. If the difference tends to zero, the set is called *measurable* and the common limit is its measure.

[1]Bernhard Riemann (1826–1866), German mathematician
[2]Henri Lebesgue (1875-1941), French mathematician

To this end let $\mathcal{S}^n$ be the set of all **half-closed rectangles** of the form $(a,b] := (a_1,b_1] \times \cdots \times (a_n,b_n] \subseteq \mathbb{R}^n$ with $a < b$ augmented by the empty set. Here $a < b$ should be read as $a_j < b_j$ for all $1 \le j \le n$ (and similarly for $a \le b$). Moreover, we allow the intervals to be unbounded, that is $a, b \in \overline{\mathbb{R}}^n$ (with $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$).

Of course one could as well take open or closed rectangles but our half-closed rectangles have the advantage that they tile nicely. In particular, one can subdivide a half-closed rectangle into smaller ones without gaps or overlap.

The somewhat dissatisfactory situation with the Riemann integral alluded to before led Cantor[3], Peano[4], and particularly Jordan[5] to the following attempt of measuring arbitrary sets: Define the measure of a rectangle via

$$|(a,b]| := \prod_{j=1}^{n} (b_j - a_j). \tag{1.1}$$

Note that the measure will be infinite if the rectangle is unbounded. Furthermore, define the **inner, outer Jordan content** of a set $A \subseteq \mathbb{R}^n$ as

$$J_*(A) := \sup \Big\{ \sum_{j=1}^{m} |R_j| \,\Big|\, \dot{\bigcup_{j=1}^{m}} R_j \subseteq A,\ R_j \in \mathcal{S}^n \Big\}, \tag{1.2}$$

$$J^*(A) := \inf \Big\{ \sum_{j=1}^{m} |R_j| \,\Big|\, A \subseteq \dot{\bigcup_{j=1}^{m}} R_j,\ R_j \in \mathcal{S}^n \Big\}, \tag{1.3}$$

respectively. Here the dot inside the union indicates that we only allow unions of mutually disjoint sets (for the outer measure this is not relevant since overlap between rectangles will only lead to a larger sum). If $J_*(A) = J^*(A)$ the set A is called **Jordan measurable**.

Unfortunately this approach turned out to have several shortcomings (essentially identical to those of the Riemann integral). Its limitation stems from the fact that one only allows finite covers. Switching to countable covers will produce the much more flexible Lebesgue measure.

Example 1.1. To understand this limitation let us look at the classical example of a non Riemann integrable function, the characteristic function of the rational numbers inside $[0,1]$. If we want to cover $\mathbb{Q} \cap [0,1]$ by a finite number of intervals, we always end up covering all of $[0,1]$ since the rational numbers are dense. Conversely, if we want to find the inner content, no single (nontrivial) interval will fit into our set since it has empty interior. In summary, $J_*(\mathbb{Q} \cap [0,1]) = 0 \neq 1 = J^*(\mathbb{Q} \cap [0,1])$.

---

[3]Georg Cantor (1845–1918), German mathematician and founder of set theory

[4]Giuseppe Peano (1858–1932), Italian mathematician

[5]Camille Jordan (1838–1922), French mathematician

On the other hand, if we are allowed to take a countable number of intervals, we can enumerate the points in $\mathbb{Q} \cap [0,1]$ and cover the j'th point by an interval of length $\varepsilon 2^{-j}$ such that the total length of this cover is less than ε, which can be arbitrarily small. $\diamond$

The previous example also hints at what is going on in general. When computing the outer content you will always end up covering the closure of A and when computing the inner content you will never get more than the interior of A. Hence a set should be Jordan measurable if the difference between the closure and the interior, which is by definition the boundary $\partial A = \overline{A} \setminus A^\circ$, is small. However, we do not want to pursue this further at this point and hence we defer the interested reader to Appendix 1.7.

Rather we will make the anticipated change and define the **Lebesgue outer measure** via

$$\lambda^{n,*}(A) := \inf \Big\{ \sum_{j=1}^{\infty} |R_j| \,\Big|\, A \subseteq \bigcup_{j=1}^{\infty} R_j,\ R_j \in \mathcal{S}^n \Big\}. \tag{1.4}$$

In particular, we will call N a Lebesgue **null set** if $\lambda^{n,*}(N) = 0$.

Example 1.2. As shown in the previous example, the set of rational numbers inside $[0,1]$ is a null set. In fact, the same argument shows that every countable set is a null set.

Consequently we expect the irrational numbers inside $[0,1]$ to be a set of measure one. But if we try to approximate this set from the inside by half-closed intervals we are bound to fail as no single (nonempty) interval will fit into this set. This explains why we did not define a corresponding inner measure. $\diamond$

Next, observe that if $f(x) = Mx + a$ is an affine transformation, then

$$\lambda^{n,*}(MA + a) = \det(M)\lambda^{n,*}(A). \tag{1.5}$$

In fact, that translations do not change the outer measure is immediate since $\mathcal{S}^n$ is invariant under translations and the same is true for $|R|$. Moreover, every matrix can be written as $M = O_1 D O_2$, where O_j are orthogonal and D is diagonal (Problem 2.22). So it reduces the problem to showing this for diagonal matrices and for orthogonal matrices. The case of diagonal matrices follows as before but the case of orthogonal matrices is more involved (it can be shown by verifying that rectangles can be replaced by open balls in the definition of the outer measure). Again we postpone this to Appendix 1.8.

For now we will use this fact only to explain why our outer measure is still not good enough. The reason is that it lacks one key property, namely additivity! Of course this will be crucial for the corresponding integral to be linear and hence is indispensable. Now here comes the bad news: A classical paradox by Banach[6] and Tarski[7] 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, and reassemble them using only rotations and translations to get two copies of the unit ball. Hence our outer measure (as well as any other reasonable notion of size which is translation and rotation invariant) cannot be additive when defined for all sets! If you think that the situation in one dimension is better, I have to disappoint you as well: Problem 1.1.

So our only hope left is that additivity at least holds on a suitable class of sets. In fact, even finite additivity is not sufficient for us since limiting operations will require that we are able to handle countable operations. It was Lebesgue who eventually was successful with the following idea: As pointed out before, there is no corresponding inner Lebesgue measure since approximation by intervals from the inside does not work well. However, instead you can try to approximate the complement from the outside thereby setting

$$\lambda^1_*(A) := (b-a) - \lambda^{1,*}([a,b] \setminus A) \tag{1.6}$$

for every bounded set $A \subseteq [a,b]$. Now you can call a bounded set A measurable if $\lambda^1_*(A) = \lambda^{1,*}(A)$. We will however use a somewhat different approach due to Carathéodory. In this respect note that if we set $E = [a,b]$ then $\lambda^1_*(A) = \lambda^{1,*}(A)$ can be written as

$$\lambda^{1,*}(E) = \lambda^{1,*}(A \cap E) + \lambda^{1,*}(A' \cap E) \tag{1.7}$$

which should be compared with the Carathéodory[8] condition (1.12). Here $A' := [a,b] \setminus A$ denotes the complement of A.

Problem* 1.1 (Vitali[9] 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?)*

Problem 1.2. *Show that $J_*(A) \le \lambda^{n,*}(A) \le J^*(A)$ and hence $J(A) = \lambda^{n,*}(A)$ for every Jordan measurable set.*

Problem 1.3. *Let $X := \mathbb{N}$ and define the set function*

$$\mu(A) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \chi_A(n) \in [0,1]$$

[6] Stefan Banach (1892–1945), Polish mathematician

[7] Alfred Tarski (1901–1983), Polish-American mathematician

[8] Constantin Carathéodory (1802–1879), Greek mathematician

[9] Giuseppe Vitali (1875–1932)), Italian mathematician

on the collection $\mathcal{S}$ of all sets for which the above limit exists. Show that this collection is closed under disjoint unions and complements but not under intersections. Show that there is an extension to the σ-algebra generated by $\mathcal{S}$ (i.e. to $\mathfrak{P}(\mathbb{N})$) which is additive but no extension which is σ-additive. (Hints: To show that $\mathcal{S}$ is not closed under intersections take a set of even numbers $A_1 \notin \mathcal{S}$ and let A_2 be the missing even numbers. Then let $A = A_1 \cup A_2 = 2\mathbb{N}$ and $B = A_1 \cup \tilde{A}_2$, where $\tilde{A}_2 = A_2 + 1$. To obtain an extension to $\mathfrak{P}(\mathbb{N})$ consider χ_A, $A \in \mathcal{S}$, as vectors in $\ell^\infty(\mathbb{N})$ and μ as a linear functional such that you can apply the Hahn–Banach theorem — compare also the construction of the Banach limit, Problem 4.23 from [**22**].*)*

1.2. Sigma algebras and measures

In the previous section we have seen that even in the case of Lebesgue measure, one cannot define the measure for all sets. Hence we start by looking at collections of subsets of a given set X suitable for measure theory. Of course any reasonable definition of a measure μ will require additivity $\mu(A_1 \cup\!\!\!\cdot\, A_2) = \mu(A_1) + \mu(A_2)$ for disjoint subsets $A_1, A_2 \subseteq X$. Consequently, a suitable collection of subsets should be closed under the usual set operations like taking unions, intersections, or complements. Such a collection is known as an **algebra** $\mathcal{A}$ provided $\emptyset \in \mathcal{A}$.

Note that $X = \emptyset' \in \mathcal{A}$ and that $\mathcal{A}$ is also closed under relative complements since $A \setminus B = A \cap B'$, where $A' = X \setminus A$ denotes the complement. Moreover, by De Morgan's rule[10] $A \cup B = (A' \cap B')'$ it suffices to check that $\mathcal{A}$ is closed under finite intersections and complements.

Example 1.3. Let $X := \{1, 2, 3\}$, then $\mathcal{A} := \{\emptyset, \{1\}, \{2, 3\}, X\}$ is an algebra. ⋄

However, we have also seen in the previous section, that we need additivity for countably many sets. This leads us to the following definition:

If an algebra is closed under countable intersections, it is called a **σ-algebra**. Hence a σ-algebra is a family of subsets Σ of a given set X such that

- $\emptyset \in \Sigma$,
- Σ is closed under countable intersections, and
- Σ is closed under complements.

By De Morgan's rule Σ is also closed under countable unions.

Example 1.4. The power set $\mathfrak{P}(X)$ is clearly the largest σ-algebra and $\{\emptyset, X\}$ is the smallest. ⋄

[10] Augustus De Morgan (1806–1871), British mathematician and logician

Moreover, the intersection of any family of (σ-)algebras $\{\Sigma_\alpha\}$ is again a (σ-)algebra (check this) 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 1.5. For a given set X and a subset $A \subseteq X$ we have $\Sigma(\{A\}) = \{\emptyset, A, A', X\}$. Moreover, every finite algebra is also a σ-algebra and if S is finite, so will be $\Sigma(S)$ (Problem 1.5). $\diamond$

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 (equivalently, all closed) sets. In fact, if X is second countable, any countable base will suffice to generate the Borel σ-algebra (since every open set can we written as a union of elements from the base; Lemma B.2 from [**22**]). Sets in the Borel σ-algebra are called **Borel sets**.

Example 1.6. 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 $\mathfrak{B}^n$, open balls with rational center and rational radius suffice. In fact, any base for the topology will suffice. Moreover, since open balls can be written as a countable union of smaller closed balls with increasing radii, we could also use compact balls instead. $\diamond$

Example 1.7. 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). $\diamond$

In the sequel we will frequently meet unions of disjoint sets and hence we will introduce the following short hand notation for the union of mutually disjoint sets:

$$\bigcupdot_{j\in J} A_j := \bigcup_{j\in J} A_j \quad \text{with} \quad A_j \cap A_k = \emptyset \text{ for all } j \neq k. \tag{1.8}$$

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

- $\mu(\emptyset) = 0$,
- $\mu(\bigcupdot_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$, $A_n \in \Sigma$ (σ-additivity).

Here the sum is set equal to ∞ if one of the summands is ∞ or if it diverges.

The measure μ is called **finite** if $\mu(X) < \infty$ and a **probability measure** if $\mu(X) = 1$. It is called **σ-finite** if there is a countable cover $\{X_n\}_{n=1}^{\infty}$ of X

[10] Émil Borel (1871–1956), French mathematician and politician

such that $X_n \in \Sigma$ and $\mu(X_n) < \infty$ for all n. (Note that it is no restriction to assume $X_n \subseteq X_{n+1}$.) The sets in Σ are called **measurable sets** and the triple (X, Σ, μ) is referred to as a **measure space**.

Example 1.8. Take a set X with $\Sigma = \mathfrak{P}(X)$ and set $\mu(A) := \#A$ 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 1.9. 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**[11] centered at x. It is also frequently written as δ_x. $\diamond$

Example 1.10. 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 1.11. Let (X_1, Σ_1, μ_1) and (X_2, Σ_2, μ_2) be two measure spaces. Then the **direct sum** is defined by taking $X := X_1 \uplus X_2$ to be the disjoint union, Σ such that $A \in \Sigma$ if and only if $A \cap X_j \in \Sigma_j$ for $j = 1, 2$ and $\mu(A) := \mu_1(A \cap X_1) + \mu_2(A \cap X_2)$. We will write $(X_1 \oplus X_2, \Sigma_1 \oplus \Sigma_2, \mu_1 \oplus \mu_2)$. The construction extends to countably many measure spaces (show this). $\diamond$

Example 1.12. Let μ be a measure on (X, Σ) and $Y \in \Sigma$ a measurable subset. Then

$$\nu(A) := \mu(A \cap Y)$$

is again a measure on (X, Σ) (show this). $\diamond$

Example 1.13. 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** (also **trace 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 $\Sigma \cap Y \subseteq \Sigma(S \cap Y)$. $\diamond$

Example 1.14. 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. $\diamond$

[11] Paul Dirac (1902–1984), English theoretical physicist

Finally, we will show that σ-additivity implies some crucial continuity properties for measures which eventually will lead to powerful limiting theorems for the corresponding integrals. We will write $A_n \nearrow A$ if $A_n \subseteq A_{n+1}$ with $A = \bigcup_n A_n$ and $A_n \searrow A$ if $A_{n+1} \subseteq A_n$ with $A = \bigcap_n A_n$.

Theorem 1.1. *Any measure μ satisfies the following properties:*

(i) $A \subseteq B$ *implies* $\mu(A) \le \mu(B)$ *(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 $\mu(B) = \mu(A) + \mu(B \setminus A)$. To see the second define $\tilde{A}_1 = A_1$, $\tilde{A}_n = A_n \setminus A_{n-1}$ 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 $\tilde{A}_n = A_1 \setminus A_n \nearrow A_1 \setminus A$ implying $\mu(\tilde{A}_n) = \mu(A_1) - \mu(A_n) \to \mu(A_1 \setminus A) = \mu(A_1) - \mu(A)$. □

Example 1.15. Consider the counting measure on $\mathbb{N}$ and let $A_n = \{j \in \mathbb{N} | j \ge n\}$. Then $\mu(A_n) = \infty$, but $\mu(\bigcap_n A_n) = \mu(\emptyset) = 0$ which shows that the requirement $\mu(A_1) < \infty$ in item (iii) of Theorem 1.1 is not superfluous. ⋄

Problem 1.4. *Find all algebras over $X := \{1, 2, 3\}$.*

Problem 1.5. *Let $\{A_j\}_{j=1}^n$ be a finite family of subsets of a given set X. Show that $\Sigma(\{A_j\}_{j=1}^n)$ has at most 4^n elements. (Hint: Let X have 2^n elements and look at the case $n = 2$ to get an idea. Consider sets of the form $B_1 \cap \cdots \cap B_n$ with $B_j \in \{A_j, A_j'\}$.)*

Problem 1.6. *Show that $\mathcal{A} := \{A \subseteq X | A$ or $X \setminus A$ is finite$\}$ is an algebra (with X some fixed set). Show that $\Sigma := \{A \subseteq X | A$ or $X \setminus A$ is countable$\}$ is a σ-algebra. (Hint: To verify closedness under unions consider the cases where all sets are finite and where one set has finite complement.)*

Problem 1.7. *Show that the union of two σ-algebras is a σ-algebra if and only if one is contained in the other.*

Problem 1.8. *Take some set X and $\Sigma := \{A \subseteq X | A$ or $X \setminus A$ is countable$\}$. Show that*

$$\nu(A) := \begin{cases} 0, & \text{if } A \text{ is countable}, \\ 1, & \text{else}. \end{cases}$$

is a measure

Problem 1.9. *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.*

1.3. Extending a premeasure to a measure

Now we are ready to show how to construct a measure starting from a small collection of sets which generate the σ-algebra and for which it is clear what the measure should be. The crucial requirement for such a collection of sets is the requirement that it is closed under intersections as the following example shows.

Example 1.16. Let us consider $X := \{1,2,3\}$ with two measures $\mu(\{1\}) := \mu(\{2\}) := \mu(\{3\}) := \frac{1}{3}$ and $\nu(\{1\}) := \nu(\{3\}) := \frac{1}{6}$, $\nu(\{2\}) := \frac{1}{2}$. Then μ and ν agree on $S := \{\{1,2\},\{2,3\}\}$ but not on $\Sigma(S) = \mathfrak{P}(X)$. Note that S is not closed under intersections. If we take $S := \{\emptyset,\{1\},\{2,3\}\}$, which is closed under intersections, and $\nu(\{1\}) := \frac{1}{3}$, $\nu(\{2\}) := \frac{1}{2}$, $\nu(\{3\}) := \frac{1}{6}$, then μ and ν agree on $\Sigma(S) = \{\emptyset,\{1\},\{2,3\},X\}$ but not on $\mathfrak{P}(X)$. $\diamond$

Hence we begin with the question when a family of sets determines a measure uniquely. To this end 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**[12] (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 $B \setminus A = (B' \cup A)' \in \mathcal{D}$ 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 $A = \dot\bigcup_j B_j$ where the sets $B_j = A_j \setminus \bigcup_{k<j} A_k$ are disjoint.

Example 1.17. Let $X := \{1,2,3,4\}$. Then $\mathcal{D} := \{A \subseteq X | \#A \text{ is even}\}$ is a Dynkin system but no algebra. $\diamond$

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 1.2 (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.

[12] Eugene Dynkin (1924–2014), Soviet–American mathematician

First of all $X \in D(A)$ since $A \cap X = A \in \mathcal{D}$. Next, if $B \in D(A)$ then $A \cap B' = A \setminus (B \cap A) \in \mathcal{D}$ (since $\mathcal{D}$ is closed under proper relative complements) implying $B' \in D(A)$. Finally if $B = \dot\bigcup_j B_j$ with $B_j \in D(A)$ disjoint, then $A \cap B = \dot\bigcup_j (A \cap B_j) \in \mathcal{D}$ with $A \cap B_j \in \mathcal{D}$ 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. □

The typical use of this lemma is as follows: First verify some property for sets in a collection 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. This is illustrated in our first result.

Theorem 1.3 (Uniqueness of measures)**.** *Let $S \subseteq \Sigma$ be a collection of sets which generates 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. Furthermore, by σ-additivity it is also closed under countable disjoint unions. Since $\mathcal{D}$ contains S by assumption, we conclude $\mathcal{D} = \Sigma(S) = \Sigma$ from Lemma 1.2. 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. □

Example 1.18. The counting measure as well as the measure which assigns every nonempty set the value ∞ both agree on $\mathcal{S}^1$. Hence the finiteness assumption in the previous theorem is crucial. ◇

Inspired by the collection of rectangles $\mathcal{S}^n$ we call a collection of subsets $\mathcal{S}$ of a given set X a **semialgebra** if $\emptyset \in \mathcal{S}$, $\mathcal{S}$ is closed under finite intersections, and the complement of a set in $\mathcal{S}$ can be written as a finite union of disjoint sets from $\mathcal{S}$. Of course a semialgebra which is closed under complements is an algebra.

In fact, considering finite disjoint unions from a semialgebra we always have a corresponding algebra.

Lemma 1.4. *Let $\mathcal{S}$ be a semialgebra, then the set of all finite disjoint unions $\bar{\mathcal{S}} := \{\dot\bigcup_{j=1}^n A_j | A_j \in \mathcal{S}\}$ is an algebra.*

Proof. Suppose $A = \dot\bigcup_{j=1}^n A_j \in \bar{\mathcal{S}}$ and $B = \dot\bigcup_{k=1}^m B_k \in \bar{\mathcal{S}}$. Then $A \cap B = \dot\bigcup_{j,k}(A_j \cap B_k) \in \bar{\mathcal{S}}$. Concerning complements we have $A' = \bigcap_j A_j' \in \bar{\mathcal{S}}$ since $A_j' \in \bar{\mathcal{S}}$ by definition of a semialgebra and since $\bar{\mathcal{S}}$ is closed under finite intersections by the first part. $\square$

Example 1.19. The collection of all intervals (augmented by the empty set) clearly is a semialgebra. Moreover, the collection $\mathcal{S}^1$ of half-open intervals of the form $(a, b] \subseteq \mathbb{R}$, $-\infty \le a < b \le \infty$ augmented by the empty set is a semialgebra. Since the product of semialgebras is again a semialgebra (Problem 1.10), the same is true for the collection of rectangles $\mathcal{S}^n$. $\diamond$

A set function $\mu : \mathcal{A} \to [0, \infty]$ on an algebra is called a **premeasure** if it satisfies

- $\mu(\emptyset) = 0$,
- $\mu(\dot\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mu(A_j)$, if $A_j \in \mathcal{A}$ and $\dot\bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$ (σ-additivity).

Note that, in contradistinction to a σ-algebra, the fact $\dot\bigcup_{j=1}^\infty A_j \in \mathcal{A}$ does not come for free from $A_j \in \mathcal{A}$ but is part of the requirement.

In terms of a premeasure Theorem 1.3 reads as follows:

Corollary 1.5. *Let μ be a σ-finite premeasure on an algebra $\mathcal{A}$. Then there is at most one extension as a measure to $\Sigma(\mathcal{A})$.*

The following lemma gives conditions when the natural extension of a set function on a semialgebra $\mathcal{S}$ to its associated algebra $\bar{\mathcal{S}}$ will be a premeasure.

Lemma 1.6. *Let $\mathcal{S}$ be a semialgebra and let $\mu : \mathcal{S} \to [0, \infty]$ be additive, that is, $A = \dot\bigcup_{j=1}^n A_j$ with $A, A_j \in \mathcal{S}$ implies $\mu(A) = \sum_{j=1}^n \mu(A_j)$. Then the natural extension $\mu : \bar{\mathcal{S}} \to [0, \infty]$ given by*

$$\mu(A) := \sum_{j=1}^n \mu(A_j), \qquad A = \dot\bigcup_{j=1}^n A_j, \tag{1.9}$$

is (well defined and) additive on $\bar{\mathcal{S}}$. Moreover, it will be a premeasure if

$$\mu(\dot\bigcup_{j=1}^\infty A_j) \le \sum_{j=1}^\infty \mu(A_j) \tag{1.10}$$

whenever $\dot{\bigcup}_j A_j \in \mathcal{S}$ *and* $A_j \in \mathcal{S}$.

Proof. We begin by showing that μ is well defined. To this end let $A = \dot{\bigcup}_{j=1}^n A_j = \dot{\bigcup}_{k=1}^m B_k$ and set $C_{jk} := A_j \cap B_k$. Then

$$\sum_j \mu(A_j) = \sum_j \mu(\dot{\bigcup_k} C_{jk}) = \sum_{j,k} \mu(C_{jk}) = \sum_k \mu(\dot{\bigcup_j} C_{jk}) = \sum_k \mu(B_k)$$

by additivity on $\mathcal{S}$. Moreover, if $A = \dot{\bigcup}_{j=1}^n A_j$ and $B = \dot{\bigcup}_{k=1}^m B_k$ are two disjoint sets from $\bar{\mathcal{S}}$, then

$$\mu(A \dot{\cup} B) = \mu(\left(\dot{\bigcup_{j=1}^n} A_j\right) \dot{\cup} \left(\dot{\bigcup_{k=1}^m} B_k\right)) = \sum_j \mu(A_j) + \sum_k \mu(B_k) = \mu(A) + \mu(B)$$

which establishes additivity. Finally, let $A = \dot{\bigcup}_{j=1}^\infty A_j \in \mathcal{S}$ with $A_j \in \mathcal{S}$ and observe $B_n = \dot{\bigcup}_{j=1}^n A_j \in \bar{\mathcal{S}}$. Hence

$$\sum_{j=1}^n \mu(A_j) = \mu(B_n) \le \mu(B_n) + \mu(A \setminus B_n) = \mu(A)$$

and combining this with our assumption (1.10) shows σ-additivity when all sets are from $\mathcal{S}$. By finite additivity this extends to the case of sets from $\bar{\mathcal{S}}$. □

Using this lemma our set function $|R|$ for rectangles is easily seen to be a premeasure.

Lemma 1.7. *The set function* (1.1) *for rectangles extends to a premeasure on* $\bar{\mathcal{S}}^n$.

Proof. Finite additivity is left as an exercise (see the proof of Lemma 1.12 below) and it remains to verify (1.10). We can cover each $A_j := (a_j, b_j]$ by some slightly larger rectangle $B_j := (a^j, b^j + \delta^j]$ such that $|B_j| \le |A_j| + \frac{\varepsilon}{2^j}$. Then for any $r > 0$ we can find an m such that the open intervals $\{(a^j, b^j + \delta^j)\}_{j=1}^m$ cover the compact set $\overline{A \cap Q_r}$, where Q_r is a half-open cube of side length r. Hence

$$|A \cap Q_r| \le \Big|\bigcup_{j=1}^m B_j\Big| = \sum_{j=1}^m |B_j| \le \sum_{j=1}^\infty |A_j| + \varepsilon.$$

Letting $r \to \infty$ and since $\varepsilon > 0$ is arbitrary, we are done. □

The next step is to extend a premeasure to an outer measure: A function $\mu^* : \mathfrak{P}(X) \to [0, \infty]$ is an **outer measure** if it has the properties

- $\mu^*(\emptyset) = 0$,
- $A \subseteq B \;\Rightarrow\; \mu^*(A) \le \mu^*(B)$ (monotonicity), and

- $\mu^*(\bigcup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \mu^*(A_n)$ (subadditivity).

Here $\mathfrak{P}(X)$ is the power set (i.e., the collection of all subsets) of X. Note that null sets N (i.e., $\mu^*(N) = 0$) do not change the outer measure: $\mu^*(A) \le \mu^*(A \cup N) \le \mu^*(A) + \mu^*(N) = \mu^*(A)$.

It turns out that it is quite easy to get an outer measure from an arbitrary set function.

Lemma 1.8. *Let $\mathcal{E}$ be some family of subsets of X containing $\emptyset$. Suppose we have a set function $\rho : \mathcal{E} \to [0, \infty]$ such that $\rho(\emptyset) = 0$. Then*

$$\mu^*(A) := \inf \Big\{ \sum_{j=1}^\infty \rho(A_j) \Big| A \subseteq \bigcup_{j=1}^\infty A_j,\ A_j \in \mathcal{E} \Big\}$$

is an outer measure. Here the infimum extends over all countable covers from $\mathcal{E}$ with the convention that the infimum is infinite if no such cover exists.

Proof. $\mu^*(\emptyset) = 0$ is trivial since we can choose $A_j = \emptyset$ as a cover.

To see see monotonicity let $A \subseteq B$ and note that if $\{A_j\}$ is a cover for B then it is also a cover for A (if there is no cover for B, there is nothing to do). Hence

$$\mu^*(A) \le \sum_{j=1}^\infty \rho(A_j)$$

and taking the infimum over all covers for B shows $\mu^*(A) \le \mu^*(B)$.

To see subadditivity note that we can assume that all sets A_j have a cover (otherwise there is nothing to do) $\{B_{jk}\}_{k=1}^\infty$ for A_j such that $\sum_{k=1}^\infty \mu(B_{jk}) \le \mu^*(A_j) + \frac{\varepsilon}{2^j}$. Since $\{B_{jk}\}_{j,k=1}^\infty$ is a cover for $\bigcup_j A_j$ we obtain

$$\mu^*(A) \le \sum_{j,k=1}^\infty \rho(B_{jk}) \le \sum_{j=1}^\infty \mu^*(A_j) + \varepsilon$$

and since $\varepsilon > 0$ is arbitrary subadditivity follows. □

Note that while we clearly have $\mu^*(A) \le \rho(A)$ for $A \in \mathcal{E}$, equality will not hold in general.

Example 1.20. Consider the algebra $\mathcal{A} := \{A \subseteq \mathbb{N} | A \text{ or } \mathbb{N} \setminus A \text{ is finite}\}$ (cf. Problem 1.6) and define $\mu(A) := 0$ if A is finite and $\mu(A) := 1$ if $\mathbb{N} \setminus A$ is finite. Then it is easy to see that μ is additive (observe for two disjoint sets from $\mathcal{A}$ at least one must be finite). Moreover, since every set can be covered by singletons, we get $\mu^*(A) = 0$ for all $A \subseteq \mathbb{N}$. Clearly $0 = \mu^*(\mathbb{N}) < \mu(\mathbb{N}) = 1$. The problem is that μ is not σ-additive. ⋄

Consequently, for any premeasure μ we define its corresponding outer measure $\mu^* : \mathfrak{P}(X) \to [0,\infty]$ (Lemma 1.8) as

$$\mu^*(A) := \inf \Big\{ \sum_{n=1}^\infty \mu(A_n) \Big| A \subseteq \bigcup_{n=1}^\infty A_n,\ A_n \in \mathcal{A} \Big\}, \tag{1.11}$$

where the infimum extends over all countable covers from $\mathcal{A}$. Replacing A_n by $\tilde{A}_n = A_n \setminus \bigcup_{m=1}^{n-1} A_m$ we see that we could even require the covers to be disjoint. In case $A \in \mathcal{A}$ we could even replace A_n by $A_n \cap A \in \mathcal{A}$ such that $\dot\bigcup_n A_n = A$ and hence $\mu^*(A) = \mu(A)$ for $A \in \mathcal{A}$.

Theorem 1.9 (Extensions via outer measures)**.** *Let μ^* be an outer measure. Then the set Σ of all sets A satisfying the Carathéodory condition*

$$\mu^*(E) = \mu^*(A \cap E) + \mu^*(A' \cap E), \quad \forall E \subseteq X, \tag{1.12}$$

(where $A' := X \setminus A$ 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

$$\begin{aligned}\mu^*(E) =& \mu^*(A \cap B \cap E) + \mu^*(A' \cap B \cap E) + \mu^*(A \cap B' \cap E)\\ &+ \mu^*(A' \cap B' \cap E)\\ \geq& \mu^*((A \cup B) \cap E) + \mu^*((A \cup B)' \cap E),\end{aligned}$$

where we have used De Morgan and

$$\mu^*(A \cap B \cap E) + \mu^*(A' \cap B \cap E) + \mu^*(A \cap B' \cap E) \geq \mu^*((A \cup B) \cap E)$$

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 1.1). Abbreviate $\tilde{A}_n = \dot\bigcup_{k \leq n} A_k$, $A = \dot\bigcup_n A_n$. 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

$$\mu^*(E) = \mu^*(\tilde{A}_n \cap E) + \mu^*(\tilde{A}_n' \cap E) \geq \sum_{k=1}^n \mu^*(A_k \cap E) + \mu^*(A' \cap E).$$

Letting $n \to \infty$ and using subadditivity finally gives

$$\mu^*(E) \geq \sum_{k=1}^{\infty} \mu^*(A_k \cap E) + \mu^*(A' \cap E) \geq \mu^*(A \cap E) + \mu^*(A' \cap E) \geq \mu^*(E)$$

and we infer that Σ is a σ-algebra.

Finally, setting $E = A$ in the last equation, 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. In fact, every **null set** A, that is, every set with $\mu^*(A) = 0$, is measurable (Problem 1.11).

The only remaining question is whether there are any nontrivial sets satisfying the Carathéodory condition.

Example 1.21. Let $X := \{0, 1\}$ with the trivial σ-algebra $\mathcal{A} := \{\emptyset, X\}$. Then $\mu(\emptyset) := 0$ and $\mu(X) := 1$ defines a measure. Moreover, one easily checks that the corresponding outer measure satisfies $\mu^*(\{0\}) = \mu^*(\{1\}) = \mu^*(X) = 1$. Hence μ^* is not σ-additive on $\mathfrak{P}(X)$. Moreover, both $\{0\}$ and $\{1\}$ fail the Carathéodory condition. ◇

The following result explains why we have bothered with premeasures rather than starting from an outer measure.

Lemma 1.10. *Let μ be additive on an algebra $\mathcal{A}$ 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

$$\sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n \cap A) + \sum_{n=1}^{\infty} \mu(A_n \cap A') \geq \mu^*(E \cap A) + \mu^*(E \cap A')$$

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. □

Thus the Lebesgue premeasure on $\bar{\mathcal{S}}^n$ gives rise to **Lebesgue measure** λ^n. The corresponding σ-algebra will be larger than the Borel σ-algebra $\mathcal{B}^n$ but we will only use the restriction to the Borel σ-algebra in this book. In fact, with the very same procedure we can obtain a large class of measures in $\mathbb{R}^n$ as will be demonstrated in the next section.

To end this section, I want to emphasize that in our approach we started from a premeasure, which gave rise to an outer measure, which eventually lead to a measure via Carathéodory's theorem. However, while some effort was required to get the premeasure in our case, an outer measure often can be obtained much easier (recall Lemma 1.8). While this immediately leads again to a measure, one is faced with the problem if any nontrivial sets satisfy the Carathéodory condition.

To address this problem let (X, d) be a metric space and call an outer measure μ^* on X a **metric outer measure** if

$$\mu^*(A_1 \cup A_2) = \mu^*(A_1) + \mu^*(A_2)$$

whenever

$$\operatorname{dist}(A_1, A_2) := \inf_{(x_1,x_2)\in A_1\times A_2} d(x_1, x_2) > 0.$$

Lemma 1.11. *Let X be a metric space. An outer measure is metric if and only if all Borel sets satisfy the Carathéodory condition* (1.12).

Proof. To show that all Borel sets satisfy the Carathéodory condition it suffices to show this is true for all closed sets. First of all note that we have $G_n := \{x \in F' \cap E | d(x, F) \geq \frac{1}{n}\} \nearrow F' \cap E$ since F is closed. Moreover, $d(G_n, F) \geq \frac{1}{n}$ and hence $\mu^*(F \cap E) + \mu^*(G_n) = \mu^*((E \cap F) \cup G_n) \leq \mu^*(E)$ by the definition of a metric outer measure. Hence it suffices to show $\mu^*(G_n) \to \mu^*(F' \cap E)$. Moreover, we can also assume $\mu^*(E) < \infty$ since otherwise there is noting to show. Now consider $\tilde{G}_n = G_{n+1} \backslash G_n$. Then $d(\tilde{G}_{n+2}, \tilde{G}_n) > 0$ and hence $\sum_{j=1}^m \mu^*(\tilde{G}_{2j}) = \mu^*(\cup_{j=1}^m \tilde{G}_{2j}) \leq \mu^*(E)$ as well as $\sum_{j=1}^m \mu^*(\tilde{G}_{2j-1}) = \mu^*(\cup_{j=1}^m \tilde{G}_{2j-1}) \leq \mu^*(E)$ and consequently $\sum_{j=1}^\infty \mu^*(\tilde{G}_j) \leq 2\mu^*(E) < \infty$. Now subadditivity implies $\mu^*(F' \cap E) \leq \mu(G_n) + \sum_{j\geq n} \mu^*(\tilde{G}_n)$ and thus

$$\mu^*(F' \cap E) \leq \liminf_{n\to\infty} \mu^*(G_n) \leq \limsup_{n\to\infty} \mu^*(G_n) \leq \mu^*(F' \cap E)$$

as required.

Conversely, suppose $\varepsilon := \operatorname{dist}(A_1, A_2) > 0$ and consider ε neighborhood of A_1 given by $O_\varepsilon = \bigcup_{x\in A_1} B_\varepsilon(x)$. Then $\mu^*(A_1 \cup A_2) = \mu^*(O_\varepsilon \cap (A_1 \cup A_2)) + \mu^*(O'_\varepsilon \cap (A_1 \cup A_2)) = \mu^*(A_1) + \mu^*(A_2)$ as required. □

Example 1.22. For example the Lebesgue outer measure (1.4) is a metric outer measure. In fact, given two sets A_1, A_2 as in the definition note that any cover of rectangles can be assumed to have only rectangles of side length smaller than $n^{-1/2} \operatorname{dist}(A_1, A_2)$ by taking subdivisions (here we use that $|.|$ is additive on rectangles). Hence every rectangle in the cover can intersect at most one of the two sets and we can partition the cover into two covers, one covering A_1 and the other covering A_2. From this we get $\mu^*(A_1 \cup A_2) \geq \mu^*(A_1) + \mu^*(A_2)$ and the other direction follows from subadditivity.

Note that at first sight it might look like this approach avoids the intermediate step of showing that $|.|$ is a premeasure. However, this is not quite true since you need this in order to show $\lambda^{n,*}(R) = |R|$ for rectangles $R \in \mathcal{S}^n$. ⋄

Problem* 1.10. *Suppose $\mathcal{S}_1$, $\mathcal{S}_2$ are semialgebras in X_1, X_2. Then $\mathcal{S} := \mathcal{S}_1 \otimes \mathcal{S}_2 := \{A_1 \times A_2 | A_j \in \mathcal{S}_j\}$ is a semialgebra in $X := X_1 \times X_2$.*

Problem* 1.11. *Show that every null set satisfies the Carathéodory condition* (1.12)*. Hence the measure constructed in Theorem 1.9 is complete.*

Problem* 1.12 (Completion of a measure)**.** *Show that every measure has an extension which is complete as follows:*

Denote by $\mathcal{N}$ the collection of subsets of X which are subsets of sets of measure zero. Define $\bar{\Sigma} := \{A \cup N | A \in \Sigma, N \in \mathcal{N}\}$ and $\bar{\mu}(A \cup N) := \mu(A)$ for $A \cup N \in \bar{\Sigma}$.

Show that $\bar{\Sigma}$ is a σ-algebra and that $\bar{\mu}$ is a well-defined complete measure. Moreover, show $\mathcal{N} = \{N \in \bar{\Sigma} | \bar{\mu}(N) = 0\}$ and $\bar{\Sigma} = \{B \subseteq X | \exists A_1, A_2 \in \Sigma$ with $A_1 \subseteq B \subseteq A_2$ and $\mu(A_2 \setminus A_1) = 0\}$.

Problem 1.13. *Let μ be a finite measure. Show that*

$$d(A,B) := \mu(A \Delta B), \qquad A \Delta B := (A \cup B) \setminus (A \cap B) \tag{1.13}$$

is a metric on Σ if we identify sets differing by 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.)

Problem 1.14. *Let μ^* be an outer measure. Show that*

$$|\mu^*(B) - \mu^*(A)| \le \mu^*(A \Delta B)$$

provided at least one set has finite outer measure.

Problem 1.15. *Let μ be a premeasure and $\bar{\mu}$ the measure obtained from μ^* via Theorem 1.9. Show that $\bar{\mu}^* = \mu^*$ and conclude that iterating the Carathéodory procedure will not produce any new measurable sets.*

1.4. Borel measures

In this section we want to construct a large class of important measures on $\mathbb{R}^n$. We begin with a few abstract definitions.

Let X be a topological space. 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 1.23. 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

$$\mu(A) = \inf_{O \supseteq A, O \text{ open}} \mu(O) \tag{1.14}$$

and **inner regular** if

$$\mu(A) = \sup_{K \subseteq A, K \text{ compact}} \mu(K). \tag{1.15}$$

It is called **regular** if it is both outer and inner regular.

Example 1.24. 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. ⋄

Example 1.25. Let $X := \mathbb{R}^n$ and $\Sigma := \mathfrak{B}^n$. Then Lebesgue measure λ^n is outer regular as can be seen from (1.4) by replacing the rectangles by slightly larger open ones. In fact we will show that it is regular below. ⋄

A set $A \in \Sigma$ is called a **support** for μ if $\mu(X \setminus A) = 0$. Note that a support is not unique (see the examples below). If X is a topological space and $\Sigma = \mathfrak{B}(X)$, one defines **the support** (also **topological support**) of μ via

$$\operatorname{supp}(\mu) := \{x \in X | \mu(O) > 0 \text{ for every open neighborhood } O \text{ of } x\}. \tag{1.16}$$

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[13] theorem (Lemma B.11 from [**22**]) there is a countable subcover, which shows that $X \setminus \operatorname{supp}(\mu)$ has measure zero.

Example 1.26. 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. ⋄

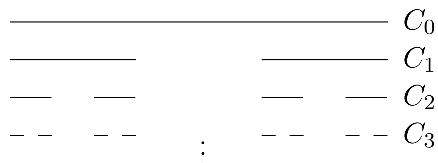
A property is said to hold **μ-almost everywhere** (a.e.) if it holds on a support for μ or, equivalently, if the set where it does not hold is contained in a set of measure zero.

[13]Ernst Leonard Lindelöf (1870–1946), Finnish mathematician

Example 1.27. 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 1.28. 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). Concerning the first note that the length of the subintervals forming C_n go to zero, and concerning the second note that a given point $x \in C$ is in a unique subinterval of C_n and hence we can find a sequence by choosing one of the boundary points of the corresponding subinterval.

Finally, note that if we change the construction by removing pieces of length $\frac{1}{4^n}$ from the middle of the intervals in the nth step almost everything works as before. We get a sequence of closed sets V_n consisting of 2^n subintervals of length $2^{-2n-1}(1+2^n)$. The limiting set V is totally disconnected, perfect and compact but now has Lebesgue measure $\frac{1}{2}$. It is known as **Smith–Volterra–Cantor set**[14] or **fat Cantor set**. ⋄

But how can we obtain some more interesting Borel measures? We will restrict ourselves to the case of $X = \mathbb{R}^n$ and begin with the case of Borel measures on $X = \mathbb{R}$ which are also known as **Lebesgue–Stieltjes measures**.[15] By what we have seen so far it is clear that our strategy is as

[14]Henry John Stephen Smith (1826–1883), Irish mathematician
[14]Vito Volterra (1860–1940), Italian mathematician
[15]Thomas Joannes Stieltjes (1856–1894), Dutch mathematician

follows: Start with some simple sets and then work your way up to all Borel sets. Hence let us first show how we should define μ for intervals: To every Borel measure on $\mathfrak{B}$ we can assign its **distribution function**

$$\mu(x) := \begin{cases} -\mu((x,0]), & x < 0, \\ 0, & x = 0, \\ \mu((0,x]), & x > 0, \end{cases} \tag{1.17}$$

which is right continuous and nondecreasing as can be easily checked.

Example 1.29. The distribution function of the Dirac measure centered at 0 is

$$\mu(x) := \begin{cases} 0, & x \geq 0, \\ -1, & x < 0. \end{cases}$$ ◇

Example 1.30. The support of a Borel measure on $\mathbb{R}$ is given in terms of the distribution function 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 $\mu(x)$. ◇

For a finite measure the alternate normalization $\tilde{\mu}(x) = \mu((-\infty,x])$ can be used. The resulting distribution function differs from our above definition by a constant $\mu(x) = \tilde{\mu}(x) - \mu((-\infty,0])$. In particular, this is the normalization used for probability measures.

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

$$I = (a,b], \quad I = [a,b], \quad I = (a,b), \quad \text{or} \quad I = [a,b), \tag{1.18}$$

with $a \leq b$, $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$. 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 (i.e. $a < b$) we can define its measure to be

$$\mu(I) := \begin{cases} m(b+) - m(a+), & I = (a,b], \\ m(b+) - m(a-), & I = [a,b], \\ m(b-) - m(a+), & I = (a,b), \\ m(b-) - m(a-), & I = [a,b), \end{cases} \tag{1.19}$$

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

$$\mu(\{a\}) := m(a+) - m(a-) \tag{1.20}$$

(which agrees with (1.19) 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 $m(x-) \le \tilde{m}(x) \le m(x+)$ 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 up to the values at the jumps.

Once we have defined μ on $\mathcal{S}^1$, we can now show that (1.19) gives a premeasure.

Lemma 1.12. *Let $m : \mathbb{R} \to \mathbb{R}$ be right continuous and nondecreasing. Then the set function defined via*

$$\mu((a, b]) := m(b) - m(a) \tag{1.21}$$

on $\mathcal{S}^1$ gives rise to a unique σ-finite premeasure on the algebra $\bar{\mathcal{S}}^1$ of finite unions of disjoint half-open intervals.

Proof. If $(a, b] = \dot\bigcup_{j=1}^n (a_j, b_j]$ then we can assume that the a_j's are ordered. Moreover, in this case we must have $b_j = a_{j+1}$ for $1 \le j < n$ and hence our set function is additive on $\mathcal{S}^1$: $\sum_{j=1}^n \mu((a_j, b_j]) = \sum_{j=1}^n (\mu(b_j) - \mu(a_j)) = \mu(b_n) - \mu(a_1) = \mu(b) - \mu(a) = \mu((a, b])$.

So by Lemma 1.6 it remains to verify (1.10). By right continuity we can cover each $A_j := (a_j, b_j]$ by some slightly larger interval $B_j := (a_j, b_j + \delta_j]$ such that $\mu(B_j) \le \mu(A_j) + \frac{\varepsilon}{2^j}$. Then for any $x > 0$ we can find an n such that the open intervals $\{(a_j, b_j + \delta_j)\}_{j=1}^n$ cover the compact set $\overline{A} \cap [-x, x]$ and hence

$$\mu(A \cap (-x, x]) \le \mu(\bigcup_{j=1}^n B_j) = \sum_{j=1}^n \mu(B_j) \le \sum_{j=1}^\infty \mu(A_j) + \varepsilon.$$

Letting $x \to \infty$ and since $\varepsilon > 0$ is arbitrary, we are done. □

And extending this premeasure to a measure we finally obtain:

Theorem 1.13 (Lebesgue–Stieltjes measures)**.** *For every nondecreasing function $m : \mathbb{R} \to \mathbb{R}$ there exists a unique regular Borel measure μ which extends* (1.19). *Two different functions generate the same measure if and only if the difference is a constant away from the discontinuities.*

Proof. Except for regularity existence follows from Theorem 1.9 together with Lemma 1.10 and uniqueness follows from Corollary 1.5. Regularity will be postponed until Section 1.6. □

We remark, that in the previous theorem we could as well consider $m : (a, b) \to \mathbb{R}$ to obtain regular Borel measures on (a, b).

Example 1.31. Suppose $\Theta(x) := 0$ for $x < 0$ and $\Theta(x) := 1$ for $x \geq 0$. 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 $0 \notin A$. $\diamond$

Example 1.32. 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 $\lambda(A) = |A|$. $\diamond$

Finally, we show how to extend Theorem 1.13 to $\mathbb{R}^n$. We will write $x \leq y$ if $x_j \leq y_j$ for $1 \leq j \leq n$ and $(a, b) = (a_1, b_1) \times \cdots \times (a_n, b_n)$, $(a, b] = (a_1, b_1] \times \cdots \times (a_n, b_n]$, etc.

The analog of (1.17) is given by

$$\mu(x) := \operatorname{sign}(x)\mu\Big(\bigtimes_{j=1}^{n} \big(\min(0, x_j), \max(0, x_j)\big]\Big), \tag{1.22}$$

where $\operatorname{sign}(x) = \prod_{j=1}^n \operatorname{sign}(x_j)$. Again, for a finite measure the alternative normalization $\tilde{\mu}(x) = \mu((-\infty, x])$ can be used.

Example 1.33. The distribution function of the Dirac measure $\mu = \delta_0$ centered at 0 is

$$\mu(x) := \begin{cases} 0, & x \geq 0, \\ -1, & \text{else.} \end{cases}$$

$\diamond$

To recover a measure μ from its distribution function we consider the difference with respect to the j'th coordinate

$$\begin{aligned} \Delta^j_{a^1,a^2} m(x) &:= m(x_1, \dots, x_{j-1}, a^2, x_{j+1}, \dots, x_n) \\ &\quad - m(x_1, \dots, x_{j-1}, a^1, x_{j+1}, \dots, x_n) \\ &= \sum_{k \in \{1,2\}} (-1)^k m(x_1, \dots, x_{j-1}, a^k, x_{j+1}, \dots, x_n) \end{aligned} \tag{1.23}$$

and define

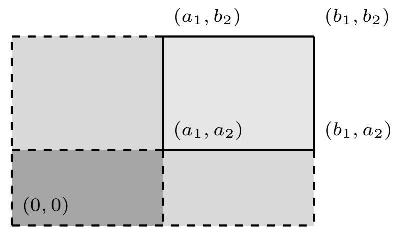
$$\begin{aligned} \Delta_{a^1,a^2} m &:= \Delta^1_{a^1_1,a^2_1} \cdots \Delta^n_{a^1_n,a^2_n} m(x) \\ &= \sum_{k \in \{1,2\}^n} (-1)^{k_1} \cdots (-1)^{k_n} m(a^{k_1}_1, \dots, a^{k_n}_n). \end{aligned} \tag{1.24}$$

Note that the above sum is taken over all vertices of the rectangle $(a^1, a^2]$ weighted with $+1$ if the vertex contains an even number of left endpoints and weighted with -1 if the vertex contains an odd number of left endpoints.

Then

$$\mu((a, b]) = \Delta_{a,b}\mu, \qquad a \leq b. \tag{1.25}$$

Of course in the case $n = 1$ this reduces to $\mu((a, b]) = \mu(b) - \mu(a)$. In the



case $n = 2$ we have $\mu((a, b]) = \mu(b_1, b_2) - \mu(b_1, a_2) - \mu(a_1, b_2) + \mu(a_1, a_2)$ which (for $0 \le a \le b$) is the measure of the rectangle with corners 0, b, minus the measure of the rectangle on the left with corners 0, (a_1, b_2), minus the measure of the rectangle below with corners 0, (b_1, a_2), plus the measure of the rectangle with corners 0, a which has been subtracted twice. The general case can be handled recursively (Problem 1.16).

Hence we will again assume that a nondecreasing function $m : \mathbb{R}^n \to \mathbb{R}^n$ is given (i.e. $m(a) \le m(b)$ for $a \le b$). However, this time monotonicity is not enough as the following example shows.

Example 1.34. Let $\mu := \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(1,0)} - \frac{1}{2}\delta_{(1,1)}$. Then the corresponding distribution function is increasing in each coordinate direction as the decrease due to the last term is compensated by the other two. However, (1.22) will give $-\frac{1}{2}$ for any rectangle containing $(1, 1)$ but not the other two points $(1, 0)$ and $(0, 1)$. ⋄

Now we can show how to get a premeasure on $\bar{\mathcal{S}}^n$.

Lemma 1.14. *Let $m : \mathbb{R}^n \to \mathbb{R}^n$ be right continuous such that μ defined via $\mu((a, b]) := \Delta_{a,b} m$ is a nonnegative set function. Then μ gives rise to a unique σ-finite premeasure on the algebra $\bar{\mathcal{S}}^n$ of finite unions of disjoint half-open rectangles.*

Proof. We first need to show finite additivity. We call $A = \dot\bigcup_k A_k$ a regular partition of $A = (a, b]$ if there are sequences $a_j = c_{j,0} < c_{j,1} < \cdots < c_{j,m_j} = b_j$ such that each rectangle A_k is of the form

$$(c_{1,i-1}, c_{1,i}] \times \cdots \times (c_{n,i-1}, c_{n,i}].$$

That is, the sets A_k are obtained by intersecting A with the hyperplanes

Figure 1.1. A partition and its regular refinement

$x_j = c_{j,i}$, $1 \le i \le m_j - 1$, $1 \le j \le n$. Now let A be bounded. Then additivity holds when partitioning A into two sets by one hyperplane $x_j = c_{j,i}$ (note that in the sum over all vertices, the one containing $c_{j,i}$ instead of a_j cancels with the one containing $c_{j,i}$ instead of b_j as both have opposite signs by the very definition of $\Delta_{a,b}m$). Hence applying this case recursively shows that additivity holds for regular partitions. Finally, for every partition we have a corresponding regular subpartition. Moreover, the sets in this subpartions can be lumped together into regular subpartions for each of the sets in the original partition. Hence the general case follows from the regular case. Finally, the case of unbounded sets A follows by taking limits.

The rest follows verbatim as in the previous lemma. □

Again this premeasure gives rise to a measure.

Theorem 1.15. *For every right continuous function $m : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\mu((a,b]) := \Delta_{a,b}m \geq 0, \quad \forall a \leq b, \tag{1.26}$$

there exists a unique regular Borel measure μ which extends the above definition.

Proof. As in the one-dimensional case existence follows from Theorem 1.9 together with Lemma 1.10 and uniqueness follows from Corollary 1.5. Again regularity will be postponed until Section 1.6. □

Example 1.35. Choosing $m(x) := \prod_{j=1}^n x_j$ we obtain the Lebesgue measure λ^n in $\mathbb{R}^n$, which is the unique Borel measure satisfying

$$\lambda^n((a,b]) = \prod_{j=1}^n (b_j - a_j).$$

We collect some simple properties below.

(i) λ^n is regular.

(ii) λ^n is uniquely defined by its values on $\mathcal{S}^n$.

(iii) For every measurable set we have

$$\lambda^n(A) = \inf\Big\{ \sum_{m=1}^\infty \lambda^n(A_m) \Big| A \subseteq \bigcup_{m=1}^\infty A_m,\ A_m \in \mathcal{S}^n \Big\},$$

where the infimum extends over all countable disjoint covers.

(iv) λ^n is translation invariant and up to normalization the only Borel measure with this property.

(i) and (ii) are part of Theorem 1.15. (iii). This follows from the construction of λ^n via its outer measure. (iv). The previous item implies that λ^n is translation invariant. Moreover, 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 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 $\mathcal{S}^n$ and equality follows from item (ii). $\diamond$

Example 1.36. If $m_j : \mathbb{R} \to \mathbb{R}$ are nondecreasing right continuous functions, then $m(x) := \prod_{j=1}^n m_j(x_j)$ satisfies

$$\Delta_{a,b} m = \prod_{j=1}^n \big(m_j(b_j) - m_j(a_j)\big)$$

and hence the requirements of Theorem 1.15 are fulfilled. $\diamond$

Problem* 1.16. *Let μ be a Borel measure on $\mathbb{R}^n$. For $a, b \in \mathbb{R}^n$ set*

$$m(a,b) := \operatorname{sign}(b-a)\mu\Big(\bigtimes_{j=1}^n \big(\min(a_j,b_j), \max(a_j,b_j)\big]\Big)$$

and $m(x) := m(0,x)$. In particular, for $a \le b$ we have $m(a,b) = \mu((a,b])$. Show that

$$m(a,b) = \Delta_{a,b} m(c,\cdot)$$

for arbitrary $c \in \mathbb{R}^n$. (Hint: Start with evaluating $\Delta^j_{a_j,b_j} m(c,\cdot)$.)

Problem* 1.17. *Let μ be a premeasure such that outer regularity* (1.14) *holds for every set in the algebra. Then the corresponding measure μ from Theorem 1.9 is outer regular.*

1.5. 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 $A \in \Sigma_Y$. When checking this condition it is useful to note that the collection of sets for which it holds, $\{A \subseteq Y | f^{-1}(A) \in \Sigma_X\}$, forms a σ-algebra on Y by $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$ and $f^{-1}(\bigcup_j A_j) = \bigcup_j f^{-1}(A_j)$. Hence it suffices to check this condition for every set A in a collection of sets which generates Σ_Y.

We will be mainly interested in the case where $(Y, \Sigma_Y) = (\mathbb{R}^n, \mathfrak{B}^n)$.

Lemma 1.16. *Let (X, Σ) be a measurable space. A function $f : X \to \mathbb{R}^n$ is measurable if and only if*

$$f^{-1}((a, \infty)) \in \Sigma \qquad \forall\, a \in \mathbb{R}^n, \tag{1.27}$$

where $(a, \infty) := \times_{j=1}^{n}(a_j, \infty)$. In particular, a function $f : X \to \mathbb{R}^n$ is measurable if and only if every component is measurable. Regarding $\mathbb{C} \cong \mathbb{R}^2$ we also get that a complex-valued function $f : X \to \mathbb{C}$ is measurable if and only if both its real and imaginary parts are and similarly for $\mathbb{C}^n$-valued functions.

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 $(a, b) := \times_{j=1}^{n}(a_j, b_j)$, which form a base for the topology. $\square$

Clearly the intervals (a, ∞) can also be replaced by $[a, \infty)$, $(-\infty, a)$, or $(-\infty, a]$.

As an immediate consequence of the definition we get that compositions of measurable functions are measurable:

Lemma 1.17. *Let (X, Σ_X), (Y, Σ_Y), (Z, Σ_Z) be measurable spaces. If $f : X \to Y$ and $g : Y \to Z$ are measurable functions, then the composition $g \circ f$ is again measurable.*

Warning: Some authors call a function $f : \mathbb{R} \to \mathbb{R}$ Lebesgue measurable if $f^{-1}((a, \infty))$ is a Lebesgue measurable set for every $a \in \mathbb{R}$ (with respect to the complete Lebesgue measure; cf. Problem 1.11). While this class is larger than the Borel functions (consider the characteristic function of a Lebesgue measurable set which is not Borel), it has the disadvantage that the composition of Lebesgue measurable functions might not be Lebesgue measurable. We will avoid such a nuisance by always chosing the Borel algebra.

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 1.18. *Let (X, Σ_X), (Y, Σ_Y) be topological spaces with their corresponding Borel σ-algebras. Any continuous function $f : X \to Y$ is measurable.*

The set of all measurable functions forms an algebra.

Corollary 1.19. *Let (X, Σ_X) be a measurable space. Suppose $f, g : X \to \mathbb{R}$ are measurable functions. Then the sum $f + g$ and the product fg are* *measurable. If $f(x) \neq 0$ for all x, then $\frac{1}{f}$ is measurable. Similarly for complex-valued functions.*

Proof. Since $(f, g) : X \to \mathbb{R}^2$ is measurable and addition as well as multiplication are continuous functions from $\mathbb{R}^2 \to \mathbb{R}$, the claim follows from the previous two lemmas. Analogously for the reciprocal function, which is continuous $\mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}$. □

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 $f : X \setminus f^{-1}(\{-\infty, \infty\}) \to \mathbb{R}$ is Borel. Since

$$\{+\infty\} = \bigcap_n (n, +\infty], \qquad \{-\infty\} = \overline{\mathbb{R}} \setminus \bigcup_n (-n, +\infty], \tag{1.28}$$

we see that $f : X \to \overline{\mathbb{R}}$ is measurable if and only if

$$f^{-1}((a, \infty]) \in \Sigma \qquad \forall\, a \in \mathbb{R}. \tag{1.29}$$

Again the intervals $(a, \infty]$ can also be replaced by $[a, \infty]$, $[-\infty, a)$, or $[-\infty, a]$. Moreover, we can generate a corresponding topology on $\overline{\mathbb{R}}$ by intervals of the form $[-\infty, b)$, (a, b), and $(a, \infty]$ with $a, b \in \mathbb{R}$.

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 1.20. *Suppose $f_n : X \to \overline{\mathbb{R}}$ is a sequence of measurable functions. Then*

$$\inf_{n \in \mathbb{N}} f_n, \quad \sup_{n \in \mathbb{N}} f_n, \quad \liminf_{n \to \infty} f_n, \quad \limsup_{n \to \infty} f_n \tag{1.30}$$

are measurable as well.

Proof. It suffices to prove that $\sup f_n$ is measurable since the rest follows from $\inf f_n = -\sup(-f_n)$, $\liminf f_n := \sup_n \inf_{k \ge n} f_k$, and $\limsup f_n := \inf_n \sup_{k \ge n} f_k$. But $(\sup f_n)^{-1}((a, \infty]) = \bigcup_n f_n^{-1}((a, \infty])$ are measurable 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. Moreover, the set where the limit exists,

$$\{x \in X | \lim_{n\to\infty} f(x) \text{ exists}\} = \{x \in X | \limsup_{n\to\infty} f(x) - \liminf_{n\to\infty} f(x) = 0\}, \tag{1.31}$$

is measurable.

Sometimes the case of arbitrary suprema and infima is also of interest. In this respect the following observation is useful: Let X be a topological space. 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

$$\overline{f}(x) := \sup_\alpha f_\alpha(x) \tag{1.32}$$

is again lower semicontinuous (and hence measurable). Similarly, f is **upper semicontinuous** if the set $f^{-1}([-\infty,a))$ is open for every $a \in \mathbb{R}$. In this case the infimum

$$\underline{f}(x) := \inf_\alpha f_\alpha(x) \tag{1.33}$$

is again upper semicontinuous. Note that f is lower semicontinuous if and only if $-f$ is upper semicontinuous.

Problem 1.18 (**preimage σ-algebra**). *Let $\mathcal{S} \subseteq \mathfrak{P}(Y)$. Show that $f^{-1}(\mathcal{S}) := \{f^{-1}(A) | A \in \mathcal{S}\}$ is a σ-algebra if $\mathcal{S}$ is. Conclude that $f^{-1}(\Sigma_Y)$ is the smallest σ-algebra on X for which f is measurable.*

Problem* 1.19. *Let (X,Σ_X), (Y,Σ_Y) be two measurable spaces and $X = \dot\bigcup_{n\in\mathbb{N}} X_n$ a partition into measurable sets $X_n \in \Sigma_X$. Show that $f : X \to Y$ is measurable if and only if $f_n := f|_{X_n}$ is measurable with respect to the trace algebra $\Sigma_n := \Sigma_X|_{X_n}$ for all $n \in \mathbb{N}$.*

Problem 1.20. *Let $\{A_n\}_{n\in\mathbb{N}}$ be a partition for X, $X = \dot\bigcup_{n\in\mathbb{N}} A_n$. Let $\Sigma = \Sigma(\{A_n\}_{n\in\mathbb{N}})$ be the σ-algebra generated by these sets. Show that $f : X \to \mathbb{R}$ is measurable if and only if it is constant on the sets A_n.*

Problem 1.21. *Let $I \subseteq \mathbb{R}$ be some interval. Show that a monotone function $f : I \to \mathbb{R}$ is Borel.*

Problem 1.22. *Show that if $f \in C(\mathbb{R}^n,\mathbb{R}^n)$ is injective, then f^{-1} is Borel. (Hint: Show that $S := \{A \subseteq \mathbb{R}^n | f(A) \in \mathfrak{B}^n\}$ is a σ-algebra containing $\mathfrak{B}^n$.)*

Problem* 1.23. *Show that the supremum over lower semicontinuous functions is again lower semicontinuous.*

1.6. How wild are measurable objects

In this section we want to investigate how far measurable objects are away from well-understood ones. The situation is intuitively summarized in what is known as **Littlewood's[16] three principles of real analysis**:

- Every (measurable) set is nearly a finite union of intervals.
- Every (measurable) function is nearly continuous.
- Every convergent sequence of (measurable) functions is nearly uniformly convergent.

As our first task we want to look at the first item and 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 1.21. *Let X be a metric space and μ a finite Borel measure. Then for every $A \in \mathfrak{B}(X)$ and any given $\varepsilon > 0$ there exists an open set O and a closed set C such that*

$$C \subseteq A \subseteq O \quad \text{and} \quad \mu(O \setminus C) \leq \varepsilon. \tag{1.34}$$

The same conclusion holds for arbitrary Borel measures if there is a sequence of open sets $U_n \nearrow X$ such that $\overline{U}_n \subseteq U_{n+1}$ and $\mu(U_n) < \infty$ (note that μ is also σ-finite in this case).

Proof. To see that (1.34) holds we begin with the case when μ is finite. Denote by $\mathcal{A}$ the set of all Borel sets satisfying (1.34). Then $\mathcal{A}$ contains every closed set C: Given C define $O_n := \{x \in X | d(x, C) < 1/n\}$ and note that O_n are open sets which satisfy $O_n \searrow C$. Thus by Theorem 1.1 (iii) $\mu(O_n \setminus C) \to 0$ and hence $C \in \mathcal{A}$.

Moreover, $\mathcal{A}$ is even a σ-algebra. That it is closed under complements is easy to see (note that $\tilde{O} := X \setminus C$ and $\tilde{C} := X \setminus O$ are the required sets for $\tilde{A} = X \setminus A$). 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 C_n, O_n such that $\mu(O_n \setminus C_n) \leq \varepsilon 2^{-n-1}$. Now $O := \bigcup_{n=1}^\infty O_n$ is open and $C := \bigcup_{n=1}^N C_n$ is closed for any finite N. Since $\mu(A)$ is finite we can choose N sufficiently large such that $\mu(\bigcup_{N+1}^\infty C_n \setminus C) \leq \varepsilon/2$. Then we have found two sets of the required type: $\mu(O \setminus C) \leq \sum_{n=1}^\infty \mu(O_n \setminus C_n) + \mu(\bigcup_{n=N+1}^\infty C_n \setminus C) \leq \varepsilon$. Thus $\mathcal{A}$ is a σ-algebra containing the closed sets, hence it is the entire Borel σ-algebra.

Now suppose μ is not finite. Set $X_1 := U_2$ and $X_n := U_{n+1} \setminus \overline{U_{n-1}}$, $n \geq 2$. Note that $X_{n+1} \cap X_n = U_{n+1} \setminus \overline{U_n}$ and $X_n \cap X_m = \emptyset$ for $|n - m| > 1$. Let $A_n = A \cap X_n$ and observe $A = \bigcup_{n=1}^\infty A_n$. By the finite case we can choose $C_n \subseteq A_n \subseteq O_n \subseteq X_n$ such that $\mu(O_n \setminus C_n) \leq \varepsilon 2^{-n}$. Now set

[16] John Edensor Littlewood (1885–1977), British mathematician

$C := \bigcup_n C_n$, $O := \bigcup_n O_n$ and note that C is closed. Indeed, let $x \in \overline{C}$ and let x_j be some sequence from C converging to x. Then $x \in U_n$ for some n and hence the sequence must eventually lie in $C \cap U_n \subseteq \bigcup_{m \le n} C_m$. Thus $x \in \overline{\bigcup_{m\le n} C_m} = \bigcup_{m \le n} C_m \subseteq C$. Finally, $\mu(O \setminus C) \le \sum_{n=1}^\infty \mu(O_n \setminus C_n) \le \varepsilon$ as required. □

This result immediately gives us outer regularity.

Corollary 1.22. *Under the assumptions of the previous lemma*

$$\mu(A) = \inf_{O \supseteq A, O \text{ open}} \mu(O) = \sup_{C \subseteq A, C \text{ closed}} \mu(C) \tag{1.35}$$

and μ is outer regular.

Proof. This follows from $\mu(A) = \mu(O) - \mu(O \setminus A) = \mu(C) + \mu(A \setminus C)$. □

If we strengthen our assumptions, we also get inner regularity. In fact, if we assume the sets U_n to be relatively compact, then the assumptions for the second case are equivalent to X being locally compact and separable (Lemma B.22 from [**22**]).

Corollary 1.23. *If X is a σ-compact metric space, then every finite Borel measure is regular. If X is a locally compact separable metric space, then every Borel measure is regular.*

Proof. By assumption there is a sequence of compact sets $K_n \nearrow X$ and for every increasing sequence of closed sets C_n with $\mu(C_n) \to \mu(A)$ we also have compact sets $C_n \cap K_n$ with $\mu(C_n \cap K_n) \to \mu(A)$. In the second case we can choose relatively compact open sets U_n (cf. item (iv) of Lemma B.22 from [**22**]) such that the assumptions of the previous theorem hold. Now argue as before using $K_n = \overline{U_n}$. □

In particular, on a locally compact and separable space every Borel measure is automatically regular and σ-finite. For example this holds for $X = \mathbb{R}^n$ (or $X = \mathbb{C}^n$).

An inner regular measure on a Hausdorff[17] space which is locally finite (every point has a neighborhood of finite measure) is called a **Radon measure**.[18] Accordingly every Borel measure on $\mathbb{R}^n$ (or $\mathbb{C}^n$) is automatically a Radon measure.

Example 1.37. 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 a 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

[17] Felix Hausdorff (1868–1942), German mathematician
[18] Johann Radon (1887–1956), Austrian mathematician

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). Of course the complement of a G_δ set is an F_σ set and vice versa.

Example 1.38. The irrational numbers are a G_δ set in $\mathbb{R}$ and the rational numbers are an F_σ set. To see this, let x_n be an enumeration of the rational numbers and consider the intersection of the open sets $O_n := \mathbb{R} \setminus \{x_n\}$. ⋄

Corollary 1.24. *Suppose X is locally compact and separable and μ a Borel measure. A set in X is Borel if and only if it differs from a G_δ set by a Borel set of measure zero. Similarly, a set in X is Borel if and only if it differs from an F_σ set by a Borel set of measure zero.*

Proof. Since G_δ sets are Borel, only the converse direction is nontrivial. By Lemma 1.21 we can find open sets O_n such that $\mu(O_n \setminus A) \le 1/n$. Now let $G := \bigcap_n O_n$. Then $\mu(G \setminus A) \le \mu(O_n \setminus A) \le 1/n$ for any n and thus $\mu(G \setminus A) = 0$. The second claim is analogous. □

A similar result holds for convergence.

Theorem 1.25 (Egorov[19])**.** *Let μ be a finite measure and f_n be a sequence of complex-valued measurable functions converging pointwise to a function f for a.e. $x \in X$. Then for every $\varepsilon > 0$ there is a set A of size $\mu(A) < \varepsilon$ such that f_n converges uniformly on $X \setminus A$.*

Proof. Let A_0 be the set where f_n fails to converge. Set

$$A_{N,k} := \bigcup_{n \ge N} \{x \in X \,|\, |f_n(x) - f(x)| \ge \frac{1}{k}\}, \qquad A_k := \bigcap_{N \in \mathbb{N}} A_{N,k}$$

and note that $A_{N,k} \searrow A_k \subseteq A_0$ as $N \to \infty$ (with k fixed). Hence $\mu(A_{N,k}) \to \mu(A_k) = 0$ by continuity from above. So for every k there is some N_k such that $\mu(A_{N_k,k}) < \frac{\varepsilon}{2^k}$. Then $A := \bigcup_{k \in \mathbb{N}} A_{N_k,k}$ satisfies $\mu(A) < \varepsilon$. Now note that $x \notin A$ implies that $x \notin A_{N_k,k}$ for every k and thus $|f_n(x) - f(x)| < \frac{1}{k}$ for $n \ge N_k$. Thus f_n converges uniformly away from A. □

Example 1.39. The example $f_n := \chi_{[n,n+1]} \to 0$ on $X = \mathbb{R}$ with Lebesgue measure shows that the finiteness assumption is important. In fact, suppose there is a set A of size less than 1 (say). Then every interval $[m, m+1]$ contains a point x_m not in A and thus $|f_m(x_m) - 0| = 1$. ⋄

[19] Dmitri Egorov (1869–1931), Russian mathematician

To end this section let us briefly discuss the third principle, namely that bounded measurable functions can be well approximated by continuous functions (under suitable assumptions on the measure). We will discuss this in detail in Section 3.4. At this point we only mention that in such a situation Egorov's theorem implies that the convergence is uniform away from a small set and hence our original function will be continuous restricted to the complement of this small set. This is known as Luzin's theorem (cf. Theorem 3.19). Note however that this does not imply that measurable functions are continuous at every point of this complement! The characteristic function of the irrational numbers is continuous when restricted to the irrational numbers but it is not continuous at any point when regarded as a function of $\mathbb{R}$.

Problem 1.24. *Show directly that for every $\varepsilon > 0$ there is an open set O of Lebesgue measure $|O| < \varepsilon$ which covers the rational numbers. Show that $\mathbb{R}$ contains a dense G_δ set of Lebesgue measure zero. In particular, $\mathbb{R}$ can be split into two disjoint sets, one which is of Lebesgue measure zero and one which is meager. (Hint for the second part: Take intersections of covers from the first part.)*

Problem* 1.25. *A finite Borel measure is regular if and only if for every Borel set A and every $\varepsilon > 0$ there is an open set O and a compact set K such that $K \subseteq A \subseteq O$ and $\mu(O \setminus K) < \varepsilon$.*

1.7. Appendix: Jordan measurable sets

In this short appendix we want to establish the criterion for Jordan measurability alluded to in Section 1.1. We begin with a useful geometric fact.

Lemma 1.26. *Every open set $O \subseteq \mathbb{R}^n$ can be partitioned into a countable number of half-open cubes from $\mathcal{S}^n$.*

Proof. Partition $\mathbb{R}^n$ into cubes of side length one with vertices from $\mathbb{Z}^n$. Start by selecting all cubes which are fully inside O and discard all those which do not intersect O. Subdivide the remaining cubes into 2^n cubes of half the side length and repeat this procedure. This gives a countable set of cubes contained in O. To see that we have covered all of O, let $x \in O$. Since x is an interior point there it will be a δ such that every cube containing x with smaller side length will be fully covered by O. Hence x will be covered at the latest when the side length of the subdivisions drops below δ. □

Now we can establish the connection between the Jordan content and the Lebesgue measure λ^n in $\mathbb{R}^n$.

Theorem 1.27. *Let $A \subseteq \mathbb{R}^n$. We have $J_*(A) = \lambda^n(A^\circ)$ and for bouneded A also $J^*(A) = \lambda^n(\overline{A})$. Hence a bounded set A is Jordan measurable if and only if its boundary $\partial A = \overline{A} \setminus A^\circ$ has Lebesgue measure zero.*

Proof. First of all note, that for the computation of $J_*(A)$ it makes no difference when we take open rectangles instead of half-open ones. But then every R with $R \subseteq A$ will also satisfy $R \subseteq A^\circ$ implying $J_*(A) = J_*(A^\circ)$. Moreover, from Lemma 1.26 we get $J_*(A^\circ) = \lambda^n(A^\circ)$. Similarly, for the computation of $J^*(A)$ it makes no difference when we take closed rectangles and thus $J^*(A) = J^*(\overline{A})$. Next, if $\overline{A}$ is compact, then given R, a finite number of slightly larger but open rectangles will give us the same volume up to an arbitrarily small error. Hence $J^*(\overline{A}) = \lambda^n(\overline{A})$ for bounded sets A. □

Note that for an unbounded set $A \subseteq \mathbb{R}^n$ we always have $J^*(A) = \infty$.

1.8. Appendix: Equivalent definitions for the outer Lebesgue measure

In this appendix we want to show that the type of sets used in the definition of the outer Lebesgue measure $\lambda^{*,n}$ on $\mathbb{R}^n$ play no role. You can cover the set by half-closed, closed, open rectangles (which is easy to see) or even replace rectangles by balls (which follows from the following lemma). To this end observe that by (1.5)

$$\lambda^n(B_r(x)) = V_n r^n \tag{1.36}$$

where $V_n := \lambda^n(B_1(0))$ is the volume of the unit ball which is computed explicitly in Section 2.3. Will will write $|A| := \lambda^n(A)$ for the Lebesgue measure of a Borel set for brevity. We first establish a covering lemma which is of independent interest.

Lemma 1.28 (Vitali covering lemma). *Let $O \subseteq \mathbb{R}^n$ be an open set and $\delta > 0$ fixed. Let $\mathcal{C}$ be a collection of balls such that every open subset of O contains at least one ball from $\mathcal{C}$. Then there exists a countable set of disjoint open balls from $\mathcal{C}$ of radius at most δ such that $O = N \cup \bigcup_j B_j$ with N a Lebesgue null set.*

Proof. Let O have finite outer measure. Start with all balls which are contained in O and have radius at most δ. Let R be the supremum of the radii of all these balls and take a ball B_1 of radius more than $\frac{R}{2}$. Now consider $O \setminus \overline{B_1}$ and proceed recursively. If this procedure terminates we are done (the missing points must be contained in the boundary of the chosen balls which has measure zero). Otherwise we obtain a sequence of balls B_j whose radii must converge to zero since $\sum_{j=1}^\infty |B_j| \le |O|$. Now fix m and let $x \in O \setminus \bigcup_{j=1}^m \bar{B}_j$. Then there must be a ball $B_0 = B_r(x) \subseteq O \setminus \bigcup_{j=1}^m \bar{B}_j$.

Moreover, there must be a first ball B_k with $B_0 \cap B_k \neq \emptyset$ (otherwise all B_k for $k > m$ must have radius larger than $\frac{r}{2}$ violating the fact that they converge to zero). By assumption $k > m$ and hence r must be smaller than two times the radius of B_k (since both balls are available in the k'th step). So the distance of x to the center of B_k must be less then three times the radius of B_k. Now if $\tilde{B}_k$ is a ball with the same center but three times the radius of B_k, then x is contained in $\tilde{B}_k$ and hence all missing points from $\bigcup_{j=1}^m B_j$ are either boundary points (which are of measure zero) or contained in $\bigcup_{k>m} \tilde{B}_k$ whose measure $|\bigcup_{k>m} \tilde{B}_k| \leq 3^n \sum_{k>m} |B_k| \to 0$ as $m \to \infty$.

If $|O| = \infty$ consider $O_m = O \cap (B_{m+1}(0) \setminus \bar{B}_m(0))$ and note that $O = N \cup \bigcup_m O_m$ where N is a set of measure zero. □

Note that in the one-dimensional case open balls are open intervals and we have the stronger result that every open set can be written as a countable union of disjoint intervals (Problem B.42 from [**22**]).

Now observe that in the definition of outer Lebesgue measure we could replace half-open rectangles by open rectangles (show this). Moreover, every open rectangle can be replaced by a disjoint union of open balls up to a set of measure zero by the Vitali covering lemma. Consequently, the Lebesgue outer measure can be written as

$$\lambda^{n,*}(A) = \inf \Big\{ \sum_{k=1}^{\infty} |A_k| \Big| A \subseteq \bigcup_{k=1}^{\infty} A_k,\ A_k \in \mathcal{C} \Big\}, \tag{1.37}$$

where $\mathcal{C}$ could be the collection of all closed rectangles, half-open rectangles, open rectangles, closed balls, or open balls.

Chapter 2

Integration

Now that we know how to measure sets, we are able to introduce the Lebesgue integral. As already mentioned, in the case of the Riemann integral, the domain of the function is split into intervals leading to an approximation by step functions, that is, linear combinations of characteristic functions of intervals. In the case of the Lebesgue integral we split the range into intervals and consider their preimages. This leads to an approximation by simple functions, that is, linear combinations of characteristic functions of arbitrary (measurable) sets.

2.1. 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 image is finite; that is, if

$$s = \sum_{j=1}^{p} \alpha_j \, \chi_{A_j}, \qquad \mathrm{Ran}(s) =: \{\alpha_j\}_{j=1}^{p}, \quad A_j := s^{-1}(\{\alpha_j\}) \in \Sigma. \tag{2.1}$$

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 $\dot{\bigcup}_{j=1}^{p} A_j = X$. Moreover, the set of simple functions $S(X, \mu)$ is a vector space and while there are different ways of writing a simple function as a linear combination of characteristic functions, the representation (2.1) is unique.

For a nonnegative simple function s as in (2.1) we define its **integral** as

$$\int_A s \, d\mu := \sum_{j=1}^{p} \alpha_j \, \mu(A_j \cap A). \tag{2.2}$$

Here we use the convention $0 \cdot \infty = 0$.

Lemma 2.1. *The integral has the following properties:*

(i) $\int_A s\,d\mu = \int_X \chi_A\, s\,d\mu$.

(ii) $\int_{\dot\bigcup_{n=1}^\infty A_n} s\,d\mu = \sum_{n=1}^\infty \int_{A_n} s\,d\mu$.

(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 $s = \sum_j \alpha_j\,\chi_{A_j}$, $t = \sum_k \beta_k\,\chi_{B_k}$ as in (2.1) and abbreviate $C_{jk} = (A_j \cap B_k) \cap A$. Note $\dot\bigcup_{j,k} C_{jk} = A$. 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, $\alpha_j \le \beta_j$. $\square$

Next we define the **Lebesgue integral** of a nonnegative measurable function $f : X \to [0,\infty]$ by

$$\int_A f\,d\mu := \sup_{\text{simple functions } s \le f} \int_A s\,d\mu, \tag{2.3}$$

where the supremum is taken over all simple functions $s \le f$. By item (vi) from our previous lemma this agrees with (2.2) if f is simple. Note that, except for possibly (ii) and (iv), Lemma 2.1 still holds for arbitrary nonnegative measurable functions s, t.

Theorem 2.2 (Monotone convergence, Beppo Levi's theorem[1]). *Let f_n be a monotone nondecreasing sequence of nonnegative measurable functions, $f_n \nearrow f$. Then f is measurable and*

$$\int_A f_n\,d\mu \to \int_A f\,d\mu. \tag{2.4}$$

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

$$\alpha \le \int_A f\,d\mu.$$

[1] Beppo Levi (1875–1961)), Italian mathematicia

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 $n \to \infty$ and adapting (ii) to the present situation, 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

$$\int_A f \, d\mu = \lim_{n\to\infty} \int_A s_n \, d\mu, \tag{2.5}$$

for every monotone sequence $s_n \nearrow f$ of simple functions. Note that there is always such a sequence, for example,

$$s_n(x) := \sum_{k=0}^{n2^n} \frac{k}{2^n} \chi_{f^{-1}(A_k)}(x), \quad A_k := [\frac{k}{2^n}, \frac{k+1}{2^n}),\ A_{n2^n} := [n, \infty). \tag{2.6}$$

By construction s_n converges uniformly if f is bounded, since $0 \le f(x) - s_n(x) < \frac{1}{2^n}$ if $f(x) \le n$.

Now what about the missing items (ii) and (iv) from Lemma 2.1? Since limits can be spread over sums, item (iv) holds, and (ii) also follows directly from the monotone convergence theorem. We even have the following result:

Lemma 2.3. *If $f \ge 0$ is measurable, then $d\nu = f \, d\mu$ defined via*

$$\nu(A) := \int_A f \, d\mu \tag{2.7}$$

is a measure such that

$$\int_A g \, d\nu = \int_A g f \, d\mu \tag{2.8}$$

for every measurable function g.

Proof. As already mentioned, additivity of ν is equivalent to linearity of the integral and σ-additivity follows from Lemma 2.1 (ii):

$$\nu(\dot\bigcup_{n=1}^{\infty} A_n) = \int_{\dot\bigcup_{n=1}^{\infty} A_n} f \, d\mu = \sum_{n=1}^{\infty} \int_{A_n} f \, d\mu = \sum_{n=1}^{\infty} \nu(A_n).$$

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 2.4 (Fatou's lemma)**.** *If f_n is a sequence of nonnegative measurable functions, then*

$$\int_A \liminf_{n\to\infty} f_n\, d\mu \le \liminf_{n\to\infty} \int_A f_n\, d\mu. \tag{2.9}$$

Proof. Set $g_n := \inf_{k\ge n} f_k$ such that $g_n \nearrow \liminf_n f_n$. Then $g_n \le f_n$ implying

$$\int_A g_n\, d\mu \le \int_A f_n\, d\mu.$$

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. □

Example 2.1. Consider $f_n := \chi_{[n,n+1]}$. Then $\lim_{n\to\infty} f_n(x) = 0$ for every $x \in \mathbb{R}$. However, $\int_{\mathbb{R}} f_n(x)dx = 1$. This shows that the inequality in Fatou's lemma cannot be replaced by equality in general. Another example is $f_n := \frac{1}{n}\chi_{[0,n]}$ which even converges to 0 uniformly. Note also that the same problem occurs on a finite interval. Consider e.g., $f_{2m} = \chi_{[0,1/2)}$, $f_{2m+1} = \chi_{(1/2,1]}$. Then again $\liminf_{n\to\infty} f_n = 0$ while $\int_0^1 f_n(x)dx = \frac{1}{2}$. ⋄

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

$$\int_A f\, d\mu := \int_A f^+ d\mu - \int_A f^- d\mu. \tag{2.10}$$

Similarly, we handle the case where f is complex-valued by calling f integrable if both the real and imaginary part are and setting

$$\int_A f\, d\mu := \int_A \mathrm{Re}(f) d\mu + \mathrm{i} \int_A \mathrm{Im}(f) d\mu. \tag{2.11}$$

Clearly a measurable function f is integrable if and only if $|f|$ is. The set of all integrable functions is denoted by $\mathcal{L}^1(X, d\mu)$.

Lemma 2.5. *The integral is linear and Lemma 2.1 holds for integrable functions s, t.*

Furthermore, for all integrable functions f, g we have

$$|\int_A f\, d\mu| \le \int_A |f|\, d\mu \tag{2.12}$$

and (triangle inequality)

$$\int_A |f + g|\, d\mu \le \int_A |f|\, d\mu + \int_A |g|\, d\mu. \tag{2.13}$$

In the first case we have equality if and only if $f(x) = \mathrm{e}^{\mathrm{i}\theta}|f(x)|$ *for a.e.* x *and some real number* θ*. In the second case we have equality if and only if* $f(x) = \mathrm{e}^{\mathrm{i}\theta(x)}|f(x)|$, $g(x) = \mathrm{e}^{\mathrm{i}\theta(x)}|g(x)|$ *for a.e.* x *and for some real-valued function* θ*.*

Proof. Linearity and Lemma 2.1 are straightforward to check. To see (2.12) put $\alpha := \frac{z^*}{|z|}$, where $z := \int_A f\, d\mu$ (without restriction $z \neq 0$). Then

$$|\int_A f\, d\mu| = \alpha \int_A f\, d\mu = \int_A \alpha\, f\, d\mu = \int_A \mathrm{Re}(\alpha\, f)\, d\mu \leq \int_A |f|\, d\mu,$$

proving (2.12). The second claim follows from $|f+g| \leq |f| + |g|$. The cases of equality are straightforward to check. □

Lemma 2.6. *Let* f *be measurable. Then*

$$\int_X |f|\, d\mu = 0 \qquad \Leftrightarrow \qquad f(x) = 0 \qquad \mu - a.e. \tag{2.14}$$

Moreover, suppose f *is nonnegative or integrable. Then*

$$\mu(A) = 0 \qquad \Rightarrow \qquad \int_A f\, d\mu = 0. \tag{2.15}$$

Proof. Observe that we have $A := \{x | f(x) \neq 0\} = \bigcup_n A_n$, where $A_n := \{x|\, |f(x)| \geq \frac{1}{n}\}$. If $\int_X |f| d\mu = 0$ we must have $\mu(A_n) \leq n \int_{A_n} |f| d\mu = 0$ for every n and hence $\mu(A) = \lim_{n\to\infty} \mu(A_n) = 0$.

The converse will follow from (2.15) since $\mu(A) = 0$ (with A as before) implies $\int_X |f| d\mu = \int_A |f| d\mu = 0$.

Finally, to see (2.15) note that by our convention $0 \cdot \infty = 0$ it holds for any simple function and hence for any nonnegative f by definition of the integral (2.3). Since any function can be written as a linear combination of four nonnegative functions this also implies the case when f is integrable. □

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.

Example 2.2. 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 2.2)

$$\int_{\mathbb{R}} f(x) d\mu(x) = \sum_n \alpha_n f(x_n).$$

Hence our integral contains sums as special cases. ◇

Finally, our integral is well behaved with respect to limiting operations. We first state a simple generalization of Fatou's lemma.

Lemma 2.7 (generalized Fatou lemma[2]). *If f_n is a sequence of real-valued measurable functions and g some integrable function. Then*

$$\int_A \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int_A f_n \, d\mu \tag{2.16}$$

if $g \le f_n$ and

$$\limsup_{n\to\infty} \int_A f_n \, d\mu \le \int_A \limsup_{n\to\infty} f_n \, d\mu \tag{2.17}$$

if $f_n \le g$.

Proof. To see the first apply Fatou's lemma to $f_n - g$ and add $\int_A g \, d\mu$ on both sides of the result. The second follows from the first using $\liminf(-f_n) = -\limsup f_n$. □

If in the last lemma we even have $|f_n| \le g$, we can combine both estimates to obtain

$$\int_A \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int_A f_n \, d\mu \le \limsup_{n\to\infty} \int_A f_n \, d\mu \le \int_A \limsup_{n\to\infty} f_n \, d\mu, \tag{2.18}$$

which is known as **Fatou–Lebesgue theorem**. In particular, in the special case where f_n converges we obtain

Theorem 2.8 (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 $|f_n| \le g$. Then f is integrable and*

$$\lim_{n\to\infty} \int_A f_n d\mu = \int_A f d\mu. \tag{2.19}$$

Proof. The real and imaginary parts satisfy the same assumptions and hence it suffices to prove the case where f_n and f are real-valued. Moreover, since $\liminf f_n = \limsup f_n = f$ equation (2.18) establishes the claim. □

[2]Pierre Fatou (1878–1929), French mathematician and astronomer

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). See Problem 2.6 for another straightforward extension.

Example 2.3. 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 $\int_{\mathbb{R}} f_n(x)dx = 1$. $\diamond$

In calculus one frequently uses the notation $\int_a^b f(x)dx$. In case of general Borel measures on $\mathbb{R}$ this is ambiguous and one needs to mention to what extend the boundary points contribute to the integral. Hence we define

$$\int_a^b f\,d\mu := \begin{cases} \int_{(a,b]} f\,d\mu, & a < b, \\ 0, & a = b, \\ -\int_{(b,a]} f\,d\mu, & b < a. \end{cases} \tag{2.20}$$

such that the usual formulas

$$\int_a^b f\,d\mu = \int_a^c f\,d\mu + \int_c^b f\,d\mu \tag{2.21}$$

remain true. Note that this is also consistent with $\mu(x) = \int_0^x d\mu$.

Example 2.4. Let $f \in C[a,b]$, then the sequence of simple functions

$$s_n(x) := \sum_{j=1}^n f(x_j)\chi_{(x_{j-1},x_j]}(x), \qquad x_j = a + \frac{b-a}{n}j$$

converges to $f(x)$ and hence the integral coincides with the limit of the Riemann–Stieltjes sums:

$$\int_a^b f\,d\mu = \lim_{n\to\infty} \sum_{j=1}^n f(x_j)\big(\mu(x_j) - \mu(x_{j-1})\big).$$

Moreover, the equidistant partition could of course be replaced by an arbitrary partition $\{x_0 = a < x_1 < \cdots < x_n = b\}$ whose length $\max_{1\le j\le n}(x_j - x_{j-1})$ tends to 0. In particular, for $\mu(x) = x$ we get the usual Riemann sums and hence the Lebesgue integral coincides with the Riemann integral at least for (piecewise) continuous functions. Further details on the connection with the Riemann integral will be given in Section 2.6. $\diamond$

Even without referring to the Riemann integral, one can easily identify the Lebesgue integral as an antiderivative: Given a continuous function $f \in C(a,b)$ which is integrable over (a,b) we can introduce

$$F(x) := \int_a^x f(y)dy, \qquad x \in (a,b). \tag{2.22}$$

Then one has

$$\frac{F(x+\varepsilon)-F(x)}{\varepsilon} = f(x) + \frac{1}{\varepsilon}\int_x^{x+\varepsilon} \big(f(y)-f(x)\big)dy$$

and

$$\limsup_{\varepsilon\to 0} \frac{1}{\varepsilon}\int_x^{x+\varepsilon} |f(y)-f(x)|dy \le \limsup_{\varepsilon\to 0} \sup_{y\in(x,x+\varepsilon]} |f(y)-f(x)| = 0$$

by the continuity of f at x. Thus $F \in C^1(a,b)$ and

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

which is a variant of the **fundamental theorem of calculus**. This tells us that the integral of a continuous function f can be computed in terms of its antiderivative and, in particular, all tools from calculus like integration by parts or integration by substitution are readily available for the Lebesgue integral on $\mathbb{R}$. A generalization of the fundamental theorem of calculus will be given in Theorem 4.29.

Example 2.5. Another fact worthwhile mentioning is that integrals with respect to Borel measures μ on $\mathbb{R}$ can be easily computed if the distribution function is continuously differentiable. In this case $\mu([a,b)) = \mu(b) - \mu(a) = \int_a^b \mu'(x)dx$ implying that $d\mu(x) = \mu'(x)dx$ in the sense of Lemma 2.3. Moreover, it even suffices that the distribution function is piecewise continuously differentiable such that the fundamental theorem of calculus holds. $\diamond$

Up to this point we have only considered real- and complex-valued functions, but one could also look at the case of functions with values in $\mathbb{R}^n$ or $\mathbb{C}^n$. In this case a function is called measurable or integrable if all components are and the integral is defined componentwise. Of course linearity of the integral or dominated convergence continue to hold. In particular, note that a measurable function f is integrable if and only if $|f|$ is integrable, where the absolute value is is now understood as Euclidean norm (in fact, any other norm would work as well).

Lemma 2.9. *Let f, g be measurable vector-valued functions, then* (2.12) *and* (2.13) *hold.*

Proof. Without loss of generality we can assume both functions to be integrable since otherwise the claims are trivial. To see (2.12), observe that it holds for simple functions by the triangle inequality (for the Euclidean norm) and hence for integrable functions by dominated convergence choosing a sequence of simple functions as in Problem 2.3. The second equation (2.13) is immediate from the triangle inequality. $\square$

Problem 2.1. *Show the **inclusion exclusion principle**:*

$$\mu(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le n} \mu(A_{i_1} \cap \cdots \cap A_{i_k}).$$

(Hint: $\chi_{A_1 \cup \cdots \cup A_n} = 1 - \prod_{i=1}^n (1 - \chi_{A_i})$*.)*

Problem* 2.2. *Consider a countable set of measures* μ_n *and numbers* $\alpha_n \ge 0$*. Let* $\mu := \sum_n \alpha_n \mu_n$ *and show*

$$\int_A f\, d\mu = \sum_n \alpha_n \int_A f\, d\mu_n \tag{2.23}$$

for any measurable function which is either nonnegative or integrable.

Problem* 2.3. *Show that for any measurable function* f *there exists a sequence of simple functions* s_n *such that* $|s_n| \le |f|$ *and* $s_n \to f$ *pointwise. If* f *is bounded, then the convergence will be uniform. (Hint: Split* f *into a linear combination of four nonnegative functions and use* (2.6)*.)*

Problem 2.4. *Let* (X, Σ) *be a measurable space. 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 the fact that a densely defined bounded linear function has a unique extension to the entire space (Theorem 1.16 from* **[22]***) could be used to extend the integral from simple to bounded measurable functions.)*

Problem 2.5. *Show that the monotone convergence holds for nondecreasing sequences of real-valued measurable functions* $f_n \nearrow f$ *provided* f_1 *is integrable.*

Problem 2.6 (Pratt[3]). *Suppose* $g_n \le f_n \le h_n$ *are sequences of real-valued measurable functions converging to* $g \le f \le h$*, respectively. Show that if* g, h *are integrable with* $\lim_{n\to\infty} \int g_n d\mu = \int g\, d\mu$ *and* $\lim_{n\to\infty} \int h_n d\mu = \int h\, d\mu$*, then* f *is integrable and* $\lim_{n\to\infty} \int f_n d\mu = \int f\, d\mu$*.*

Moreover, suppose f_n*,* g_n *are sequences of measurable functions with* $|f_n| \le g_n$ *and converging to* f*,* g*, respectively. Show that if* g *is integrable with* $\lim_{n\to\infty} \int g_n d\mu = \int g\, d\mu$*, then* f *is integrable and* $\lim_{n\to\infty} \int f_n d\mu = \int f\, d\mu$*.*

Problem 2.7. *Show that the dominated convergence theorem implies (under the same assumptions)*

$$\lim_{n\to\infty} \int |f_n - f| d\mu = 0.$$

[3] John W. Pratt (1931), American mathematician and statistician

Problem 2.8. *Consider*

$$m(x) = \begin{cases} 0, & x < 0, \\ \frac{x}{2}, & 0 \le x < 1, \\ 1 & 1 \le x, \end{cases}$$

and let μ be the associated measure. Compute $\int_{\mathbb{R}} x\, d\mu(x)$.

Problem 2.9. *Let $\mu(A) < \infty$ and f be a real-valued integrable function satisfying $f(x) \le M$. Show that*

$$\int_A f\, d\mu \le M \mu(A)$$

with equality if and only if $f(x) = M$ for a.e. $x \in A$.

Problem 2.10. *Let μ be a nontrivial measure, $\mu(X) > 0$. Suppose f, g are integrable functions with $f < g$. Then $\int_X f\, d\mu < \int_X g\, d\mu$.*

Problem 2.11. *Let $f : [0,1] \to \mathbb{R}$ be an integrable function and let m_0 be a number such that $0 < m_0 < 1$. Show that $\int_A f(x)dx = 0$ for all Borel sets A with $|A| = m_0$ implies $f = 0$ a.e.*

Problem* 2.12. *Let $X \subseteq \mathbb{R}$, Y be some measure space, and $f : X \times Y \to \mathbb{C}$. Suppose $y \mapsto f(x,y)$ is measurable for every x and $x \mapsto f(x,y)$ is continuous for every y. Show that*

$$F(x) := \int_Y f(x,y)\, d\mu(y) \tag{2.24}$$

is continuous if there is an integrable function $g(y)$ such that $|f(x,y)| \le g(y)$.

Problem* 2.13. *Let $X \subseteq \mathbb{R}$, Y be some measure space, and $f : X \times Y \to \mathbb{C}$. 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*

$$F(x) := \int_Y f(x,y)\, d\mu(y) \tag{2.25}$$

is differentiable if there is an integrable function $g(y)$ such that $|\frac{\partial}{\partial x} f(x,y)| \le g(y)$. Moreover, $y \mapsto \frac{\partial}{\partial x} f(x,y)$ is measurable and

$$F'(x) = \int_Y \frac{\partial}{\partial x} f(x,y)\, d\mu(y) \tag{2.26}$$

in this case. (See Problem 4.42 for an extension.)

2.2. 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 $A_1 \times A_2$ with $A_j \in \Sigma_j$, $j = 1, 2$..

Example 2.6. 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 2.10. *Suppose $A \in \Sigma_1 \otimes \Sigma_2$. Then its sections*

$$A_1(x_2) := \{x_1 | (x_1, x_2) \in A\} \quad \text{and} \quad A_2(x_1) := \{x_2 | (x_1, x_2) \in A\} \tag{2.27}$$

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

$$\mu_1 \otimes \mu_2(A_1 \times A_2) := \mu_1(A_1)\mu_2(A_2), \quad A_j \in \Sigma_j,\ j = 1, 2. \tag{2.28}$$

Since the rectangles are closed under intersection, Theorem 1.3 implies that there is at most one measure on $\Sigma_1 \otimes \Sigma_2$ provided μ_1 and μ_2 are σ-finite.

Theorem 2.11. *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*

$$\int_{X_1} \mu_2(A_2(x_1)) d\mu_1(x_1) = \int_{X_2} \mu_1(A_1(x_2)) d\mu_2(x_2). \tag{2.29}$$

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 1.2. 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

$$\int_{X_1} \mu_2(A_2 \cap X_{2,j})\chi_{X_{1,j}} d\mu_1 = \int_{X_2} \mu_1(A_1 \cap X_{1,j})\chi_{X_{2,j}} d\mu_2$$

and the σ-finite case follows from the monotone convergence theorem. $\square$

Hence for given $A \in \Sigma_1 \otimes \Sigma_2$ we can define

$$\mu_1 \otimes \mu_2(A) := \int_{X_1} \mu_2(A_2(x_1)) d\mu_1(x_1) = \int_{X_2} \mu_1(A_1(x_2)) d\mu_2(x_2) \tag{2.30}$$

or equivalently, since $\chi_{A_1(x_2)}(x_1) = \chi_{A_2(x_1)}(x_2) = \chi_A(x_1, x_2)$,

$$\begin{aligned} \mu_1 \otimes \mu_2(A) &= \int_{X_1} \left(\int_{X_2} \chi_A(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1) \\ &= \int_{X_2} \left(\int_{X_1} \chi_A(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2). \end{aligned} \tag{2.31}$$

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.

Example 2.7. Let $X_1 = X_2 = [0, 1]$ with μ_1 Lebesgue measure and μ_2 the counting measure. Let $A = \{(x, x) | x \in [0, 1]\}$ such that $\mu_2(A_2(x_1)) = 1$ and $\mu_1(A_1(x_2)) = 0$ implying

$$1 = \int_{X_1} \mu_2(A_2(x_1)) d\mu_1(x_1) \neq \int_{X_2} \mu_1(A_1(x_2)) d\mu_2(x_2) = 0.$$

Hence the theorem can fail if one of the measures is not σ-finite. Note that it is still possible to define a product measure without σ-finiteness (Problem 2.16), but, as the example shows, it will lack some nice properties. $\diamond$

Finally we have

Theorem 2.12 (Fubini[4]). *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*

$$\begin{aligned} \iint_{X_1 \times X_2} f(x_1, x_2) d\mu_1 \otimes \mu_2(x_1, x_2) &= \int_{X_2} \left(\int_{X_1} f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2) \\ &= \int_{X_1} \left(\int_{X_2} f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1). \end{aligned} \tag{2.32}$$

(ii) *If f is complex-valued, then*

$$\int_{X_1} |f(x_1, x_2)| d\mu_1(x_1) \in \mathcal{L}^1(X_2, d\mu_2), \tag{2.33}$$

respectively,

$$\int_{X_2} |f(x_1, x_2)| d\mu_2(x_2) \in \mathcal{L}^1(X_1, d\mu_1), \tag{2.34}$$

if and only if $f \in \mathcal{L}^1(X_1 \times X_2, d\mu_1 \otimes d\mu_2)$. In this case (2.32) *holds.*

[4]Guido Fubini (1879–1943), Italian mathematician

Proof. By Theorem 2.11 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. The case of nonnegative functions is also called **Tonelli's theorem**.[5] In the general case the integrability condition is crucial, as the following example shows.

Example 2.8. Let $X := [0,1] \times [0,1]$ with Lebesgue measure and consider

$$f(x,y) = \frac{x-y}{(x+y)^3}.$$

Then

$$\int_0^1 \int_0^1 f(x,y) dx\, dy = -\int_0^1 \frac{1}{(1+y)^2} dy = -\frac{1}{2}$$

but (by symmetry)

$$\int_0^1 \int_0^1 f(x,y) dy\, dx = \int_0^1 \frac{1}{(1+x)^2} dx = \frac{1}{2}.$$

Consequently f cannot be integrable over X (verify this directly). ⋄

Lemma 2.13. *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 2.14). Thus the claim follows from Problem 1.17. □

In connection with Theorem 1.3 the following observation is of interest:

Lemma 2.14. *If S_j generates Σ_j and $X_j \in S_j$ for $j = 1,2$, 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.

[5] Leonida Tonelli (1885–1946), Italian mathematician

Lemma 2.15. *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*

$$(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3). \tag{2.35}$$

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 1.3. □

Hence we can take the product of finitely many measures. The case of infinitely many measures requires a bit more effort and will be discussed in Section 5.2.

Example 2.9. If λ is Lebesgue measure on $\mathbb{R}$, then $\lambda^n = \lambda \otimes \cdots \otimes \lambda$ is Lebesgue measure on $\mathbb{R}^n$. In fact, it satisfies $\lambda^n((a, b]) = \prod_{j=1}^n (b_j - a_j)$ and hence must be equal to Lebesgue measure which is the unique Borel measure with this property. ⋄

Example 2.10. If X_1, X_2 are second countable topological spaces, then $\mathfrak{B}(X_1 \times X_2) = \mathfrak{B}(X_1) \otimes \mathfrak{B}(X_2)$ since open rectangles are a base for the product topology. Moreover, if μ_1, μ_2 are Borel measures and both X_1 and X_1 are locally compact, then $\mu_1 \otimes \mu_2$ is also a Borel measure. Indeed, let $K \subseteq X_1 \times X_2$ be compact. Then for every point in K there is a relatively compact open rectangle containing this point. By compactness finitely many of them suffice to cover K, that is $K \subseteq \bigcup_{j=1}^n K_{1,j} \times K_{2,j}$ implying $\mu_1 \otimes \mu_2(K) \leq \sum_{j=1}^n \mu(K_{1,j})\mu(K_{2,j}) < \infty$. ⋄

Problem* 2.14. *Show that the set of all finite unions of measurable rectangles $A_1 \times A_2$ forms an algebra. Moreover, every set in this algebra can be written as a finite union of disjoint rectangles.*

Problem* 2.15. *Let (X_j, Σ_j), $j = 0, 1, 2$ be measurable spaces and $f_j : X_0 \to X_j$, $j = 1, 2$ be functions. Show that $(f_1, f_2) : X_0 \to X_1 \times X_2$ is measurable (with respect to the product algebra $\Sigma_1 \otimes \Sigma_2$) if and only if f_1 and f_2 are measurable.*

Problem 2.16. *Given two measure spaces (X_1, Σ_1, μ_1) and (X_2, Σ_2, μ_2) let $\mathcal{R} = \{A_1 \times A_2 | A_j \in \Sigma_j,\ j = 1, 2\}$ be the collection of measurable rectangles. Define $\rho : \mathcal{R} \to [0, \infty]$, $\rho(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$. Then*

$$(\mu_1 \otimes \mu_2)^*(A) := \inf\Big\{ \sum_{j=1}^\infty \rho(A_j) \Big| A \subseteq \bigcup_{j=1}^\infty A_j,\ A_j \in \mathcal{R} \Big\}$$

is an outer measure on $X_1 \times X_2$. *Show that this constructions coincides with* (2.30) *for* $A \in \Sigma_1 \otimes \Sigma_2$ *in case* μ_1 *and* μ_2 *are* σ*-finite.*

Problem 2.17. *Let* $P : \mathbb{R}^n \to \mathbb{C}$ *be a nonzero polynomial. Show that* $N := \{x \in \mathbb{R}^n | P(x) = 0\}$ *is a Borel set of Lebesgue zero. (Hint: Induction using Fubini.)*

Problem* 2.18. *Let* $U \subseteq \mathbb{C}$ *be a domain,* Y *be some measure space, and* $f : U \times Y \to \mathbb{C}$. *Suppose* $y \mapsto f(z,y)$ *is measurable for every* z *and* $z \mapsto f(z,y)$ *is holomorphic for every* y. *Show that*

$$F(z) := \int_Y f(z,y)\, d\mu(y)$$

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's theorem from complex analysis.)*

Problem 2.19. *Let* $f : X \to \mathbb{R}$ *be measurable. Show that the sublevel sets* $S_f(t) := \{(x,t) | f(x) < t\} \subseteq X \times \mathbb{R}$ *are measurable.*

Problem* 2.20. *Suppose* $\phi : [0,\infty) \to [0,\infty)$ *is integrable over every compact interval and set* $\Phi(r) = \int_0^r \phi(s) ds$. *Let* $f : X \to \mathbb{C}$ *be measurable and introduce its distribution function*

$$E_f(r) := \mu\big(\{x \in X | \, |f(x)| > r\}\big).$$

Show that

$$\int_X \Phi(|f|) d\mu = \int_0^\infty \phi(r) E_f(r) dr.$$

Moreover, show that if f *is integrable, then the set of all* $\alpha \in \mathbb{C}$ *for which* $\mu\big(\{x \in X | \, f(x) = \alpha\}\big) > 0$ *is countable.*

2.3. 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 the **pushforward measure** (also image measure) $f_\star \mu$ on Y via

$$(f_\star \mu)(A) := \mu(f^{-1}(A)). \tag{2.36}$$

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 2.16. *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*

$$\int_Y g \, d(f_\star \mu) = \int_X g \circ f \, d\mu. \tag{2.37}$$

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 2.11. Suppose $f: X \to Y$ and $g: Y \to Z$. Then

$$(g \circ f)_\star \mu = g_\star(f_\star \mu)$$

since $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. ◇

Example 2.12. 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

$$f_\star \lambda^n = \frac{1}{|\det(M)|} \lambda^n.$$

To see this, note that $f_\star \lambda^n$ is translation invariant and hence must be a multiple of λ^n. Moreover, for an orthogonal matrix this multiple is one (since an orthogonal matrix leaves the unit ball invariant) and for a diagonal matrix it must be the absolute value of the product of the diagonal elements (consider a rectangle). Finally, since every invertible matrix can be written as $M = O_1 D O_2$, where O_j are orthogonal and D is diagonal (Problem 2.22), the claim follows.

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 $f(x) = \alpha x$. ◇

This result can be generalized to diffeomorphisms (one-to-one C^1 maps with inverse again C^1):

Theorem 2.17 (change of variables)**.** *Let $U, V \subseteq \mathbb{R}^n$ and suppose $f \in C^1(U, V)$ is a diffeomorphism. Then*

$$(f^{-1})_\star d^n x = |J_f(x)| d^n x, \tag{2.38}$$

where $J_f = \det(\frac{\partial f}{\partial x})$ is the Jacobi determinant of f. In particular,

$$\int_U g(f(x)) |J_f(x)| d^n x = \int_V g(y) d^n y \tag{2.39}$$

whenever g is nonnegative or integrable over V.

Proof. It suffices to show

$$\int_{f(R)} d^n y = \int_R |J_f(x)| d^n x$$

for every bounded open rectangle $R \subseteq U$. By Theorem 1.3 it will then follow for characteristic functions and thus for arbitrary functions by the very definition of the integral.

To this end we consider the integral

$$I_\varepsilon := \int_{f(R)} \int_R |J_f(f^{-1}(y))| \varphi_\varepsilon(f(z) - y) d^n z \, d^n y$$

Here $\varphi := V_n^{-1} \chi_{B_1(0)}$ and $\varphi_\varepsilon(y) := \varepsilon^{-n} \varphi(\varepsilon^{-1} y)$, where V_n is the volume of the unit ball (cf. below), such that $\int \varphi_\varepsilon(x) d^n x = 1$.

We will evaluate this integral in two ways. To begin with we consider the inner integral

$$h_\varepsilon(y) := \int_R \varphi_\varepsilon(f(z) - y) d^n z.$$

For $\varepsilon < \varepsilon_0$ the integrand is nonzero only for $z \in K := f^{-1}(\overline{B_{\varepsilon_0}(y)})$, where K is some compact set containing $x = f^{-1}(y)$. Using the affine change of coordinates $z = x + \varepsilon w$ we obtain

$$h_\varepsilon(y) = \int_{W_\varepsilon(x)} \varphi\left(\frac{f(x + \varepsilon w) - f(x)}{\varepsilon}\right) d^n w, \qquad W_\varepsilon(x) = \frac{1}{\varepsilon}(K - x).$$

By

$$\left|\frac{f(x + \varepsilon w) - f(x)}{\varepsilon}\right| \geq \frac{1}{C}|w|, \qquad C := \sup_K \|df^{-1}\|$$

the integrand is nonzero only for $w \in B_C(0)$. Hence, as $\varepsilon \to 0$, the domain $W_\varepsilon(x)$ will eventually cover all of $B_C(0)$ and dominated convergence implies

$$\lim_{\varepsilon \downarrow 0} h_\varepsilon(y) = \int_{B_C(0)} \varphi(df(x) w) d^n w = |J_f(x)|^{-1}.$$

Consequently, $\lim_{\varepsilon \downarrow 0} I_\varepsilon = |f(R)|$ again by dominated convergence. Now we use Fubini to interchange the order of integration

$$I_\varepsilon = \int_R \int_{f(R)} |J_f(f^{-1}(y))| \varphi_\varepsilon(f(z) - y) d^n y \, d^n z.$$

Since $f(z)$ is an interior point of $f(R)$, continuity of $|J_f(f^{-1}(y))|$ implies

$$\lim_{\varepsilon \downarrow 0} \int_{f(R)} |J_f(f^{-1}(y))| \varphi_\varepsilon(f(z) - y) d^n y = |J_f(f^{-1}(f(z)))| = |J_f(z)|$$

and hence dominated convergence shows $\lim_{\varepsilon \downarrow 0} I_\varepsilon = \int_R |J_f(z)| d^n z$. □

Example 2.13. 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

$$\det \frac{\partial T_2}{\partial(\rho, \varphi)} = \det \begin{vmatrix} \cos(\varphi) & -\rho \sin(\varphi) \\ \sin(\varphi) & \rho \cos(\varphi) \end{vmatrix} = \rho$$

and one has

$$\int_U f(\rho\cos(\varphi), \rho\sin(\varphi))\rho\, d(\rho,\varphi) = \int_{T_2(U)} f(x) d^2x.$$

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 2.14. 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 $\rho \geq 0$. 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

$$\int_U f(T_3(r,\varphi,\theta)) r^2\sin(\theta) d(r,\varphi,\theta) = \int_{T_3(U)} f(x) d^3x.$$

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 to arbitrary dimensions $T_n : [0,\infty)\times[0,2\pi)\times[0,\pi]^{n-2} \to \mathbb{R}^n$ is given by

$$x = T_n(r,\varphi,\theta_1,\ldots,\theta_{n-2})$$

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

$$\det \frac{\partial T_n}{\partial(r,\varphi,\theta_1,\dots,\theta_{n-2})} = r^{n-1}\sin(\theta_1)\sin(\theta_2)^2\cdots\sin(\theta_{n-2})^{n-2}. \qquad \diamond$$

Another useful consequence of Theorem 2.16 is the following rule for integrating radial functions.

Lemma 2.18. *There is a measure σ^{n-1} on the **unit sphere** $S^{n-1} := \partial B_1(0) = \{x \in \mathbb{R}^n | \, |x| = 1\}$, which is rotation invariant and satisfies*

$$\int_{\mathbb{R}^n} g(x)d^n x = \int_0^\infty \int_{S^{n-1}} g(r\omega) r^{n-1} d\sigma^{n-1}(\omega) dr, \tag{2.40}$$

for every integrable function g.

Moreover, the surface area of S^{n-1} is given by

$$S_n := \sigma^{n-1}(S^{n-1}) = nV_n, \tag{2.41}$$

where $V_n := \lambda^n(B_1(0))$ is the volume of the unit ball in $\mathbb{R}^n$, and if $g(x) = \tilde{g}(|x|)$ is radial we have

$$\int_{\mathbb{R}^n} g(x)d^n x = S_n \int_0^\infty \tilde{g}(r) r^{n-1} dr. \tag{2.42}$$

Proof. Consider the measurable transformation $f : \mathbb{R}^n \to [0,\infty) \times S^{n-1}$, $x \mapsto (|x|, \frac{x}{|x|})$ (with $\frac{0}{|0|} = 1$). Let $d\mu(r) := r^{n-1}dr$ and

$$\sigma^{n-1}(A) := n\lambda^n(f^{-1}([0,1) \times A)) \tag{2.43}$$

for every $A \in \mathfrak{B}(S^{n-1}) = \mathfrak{B}^n \cap S^{n-1}$. Note that σ^{n-1} inherits the rotation invariance from λ^n. By Theorem 2.16 it suffices to show $f_\star\lambda^n = \mu \otimes \sigma^{n-1}$. This follows from

$$
\begin{aligned}
(f_\star\lambda^n)([0,r) \times A) &= \lambda^n(f^{-1}([0,r) \times A)) = r^n\lambda^n(f^{-1}([0,1) \times A))\\
&= \mu([0,r))\sigma^{n-1}(A).
\end{aligned}
$$

since these sets determine the measure uniquely. □

Clearly in spherical coordinates the surface measure is given by

$$d\sigma^{n-1} = \sin(\theta_1)\sin(\theta_2)^2\cdots\sin(\theta_{n-2})^{n-2} d\varphi\, d\theta_1 \cdots d\theta_{n-2}. \tag{2.44}$$

Example 2.15. 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 $V_n := V_n(1)$.

To this end we use (Problem 2.23)

$$\begin{aligned}\pi^{n/2} &= \int_{\mathbb{R}^n} \mathrm{e}^{-|x|^2} d^n x = n V_n \int_0^\infty \mathrm{e}^{-r^2} r^{n-1} dr = \frac{n V_n}{2} \int_0^\infty \mathrm{e}^{-s} s^{n/2-1} ds \\ &= \frac{n V_n}{2} \Gamma(\frac{n}{2}) = V_n \Gamma(\frac{n}{2} + 1),\end{aligned}$$

where Γ is the gamma function (Problem 2.24). Hence

$$V_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)}. \tag{2.45}$$

By $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ (see Problem 2.25) this coincides with the well-known values $V_1 = 2$, $V_2 = \pi$, $V_3 = \frac{4\pi}{3}$. ◇

Example 2.16. The above lemma can be used to determine when a radial function is integrable. For example, we obtain

$$|x|^\alpha \in L^1(B_1(0)) \;\Leftrightarrow\; \alpha > -n, \qquad |x|^\alpha \in L^1(\mathbb{R}^n \setminus B_1(0)) \;\Leftrightarrow\; \alpha < -n. \quad \diamond$$

Problem 2.21. *Let λ be Lebesgue measure on $\mathbb{R}$, and let f be a strictly increasing function with $\lim_{x\to\pm\infty} f(x) = \pm\infty$. Show that*

$$f_\star \lambda = d(f^{-1}),$$

where f^{-1} is the inverse of f extended to all of $\mathbb{R}$ by setting $f^{-1}(y) = x$ for $y \in [f(x-), f(x+)]$ (note that f^{-1} is continuous).

Moreover, if $f \in C^1(\mathbb{R})$ with $f' > 0$, then

$$f_\star \lambda = \frac{1}{f'(f^{-1})} d\lambda.$$

Problem* 2.22. *Show that every invertible matrix M can be written as $M = O_1 D O_2$, where D is diagonal and O_j are orthogonal. (Hint: The matrix $M^* M$ is nonnegative and hence there is an orthogonal matrix U which diagonalizes $M^* M = U D^2 U^*$. Then one can choose $O_1 = M U D^{-1}$ and $O_2 = U^*$.)*

Problem* 2.23. *Show*

$$I_n := \int_{\mathbb{R}^n} \mathrm{e}^{-|x|^2} d^n x = \pi^{n/2}.$$

(Hint: Use Fubini to show $I_n = I_1^n$ and compute I_2 using polar coordinates.)

Problem* 2.24. *The* ***gamma function*** *is defined via*

$$\Gamma(z) := \int_0^\infty x^{z-1} \mathrm{e}^{-x} dx, \qquad \mathrm{Re}(z) > 0. \tag{2.46}$$

Verify that the integral converges and defines an analytic function in the indicated half-plane (cf. Problem 2.18). Use integration by parts to show

$$\Gamma(z+1) = z\Gamma(z), \qquad \Gamma(1) = 1. \tag{2.47}$$

Conclude $\Gamma(n) = (n-1)!$ *for* $n \in \mathbb{N}$. *Show that the relation* $\Gamma(z) = \Gamma(z+1)/z$ *can be used to define* $\Gamma(z)$ *for all* $z \in \mathbb{C}\backslash\{0, -1, -2, \dots\}$. *Show that near* $z = -n$, $n \in \mathbb{N}_0$, *the Gamma functions behaves like* $\Gamma(z) = \frac{(-1)^n}{n!(z+n)} + O(1)$.

Problem* 2.25. *Show that* $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. *Moreover, show*

$$\Gamma(n + \frac{1}{2}) = \frac{(2n)!}{4^n n!}\sqrt{\pi}$$

(Hint: Use the change of coordinates $x = t^2$ *and then use Problem 2.23.)*

Problem 2.26. *Establish*

$$\left(\frac{d}{dz}\right)^j \Gamma(z) = \int_0^\infty \log(x)^j x^{z-1} \mathrm{e}^{-x} dx, \qquad \mathrm{Re}(z) > 0.$$

Conclude that Γ *is strictly convex and has a unique minimum between* 1 *and* 2 *on* $(0, \infty)$. *Show that* Γ *is even log-convex, that is,* $\log(\Gamma(x))$ *is convex. (Hint: Regard the first derivative as a scalar product and apply Cauchy–Schwarz.)*

Problem 2.27. *Show that the* ***Beta function*** *satisfies*

$$B(u,v) := \int_0^1 t^{u-1}(1-t)^{v-1} dt = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}, \quad \mathrm{Re}(u) > 0,\ \mathrm{Re}(v) > 0.$$

A few other common forms are

$$\begin{aligned} B(u,v) &= 2\int_0^{\pi/2} \sin(\theta)^{2u-1} \cos(\theta)^{2v-1} d\theta \\ &= \int_0^\infty \frac{s^{u-1}}{(1+s)^{u+v}} ds = 2^{-u-v+1} \int_{-1}^1 (1+s)^{u-1}(1-s)^{v-1} ds \\ &= n\int_0^1 s^{nu-1}(1-s^n)^{v-1} ds, \qquad n > 0. \end{aligned}$$

Use this to establish ***Euler's reflection formula***[6]

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$$

[6] Leonhard Euler (1707–1783), Swiss mathematician, physicist, astronomer, geographer, logician and engineer

and ***Legendre's duplication formula***[7]

$$\Gamma(2z) = \frac{2^{2z-1}}{\sqrt{\pi}}\Gamma(z)\Gamma(z+\frac{1}{2}).$$

Conclude that the Gamma function has no zeros on $\mathbb{C}$.

(Hint: Start with $\Gamma(u)\Gamma(v)$ *and make a change of variables* $x = ts$, $y = t(1-s)$*. For the reflection formula evaluate* $B(z, 1-z)$ *using Problem 2.28. For the duplication formula relate* $B(z,z)$ *and* $B(\frac{1}{2}, z)$*.)*

Problem 2.28. *Show*

$$\int_{-\infty}^{\infty} \frac{\mathrm{e}^{zx}}{a+\mathrm{e}^x} dx = \int_0^{\infty} \frac{y^{z-1}}{a+y} dy = \frac{\pi a^{z-1}}{\sin(z\pi)}, \quad 0 < \mathrm{Re}(z) < 1,\, a \in \mathbb{C}\setminus(-\infty, 0].$$

(Hint: First reduce it to the case $a = 1$*. Then, use a contour consisting of the straight lines connecting the points* $-R$, R, $R + 2\pi\mathrm{i}$, $-R + 2\pi\mathrm{i}$*. Evaluate the contour integral using the residue theorem and let* $R \to \infty$*. Show that the contributions from the vertical lines vanish in the limit and relate the integrals along the horizontal lines.)*

Problem 2.29. *Show* ***Stirling's formula***[8]

$$\Gamma(x) = \frac{x^x}{\mathrm{e}^x}\left(\sqrt{\frac{2\pi}{x}} + O(x^{-3/2})\right), \qquad x \to \infty,$$

for $x > 0$*. (Hint: The maximum of* $t^x \mathrm{e}^{-t}$ *occurs at* $t = x$ *and this suggests a change of variables* $s = tx$ *which gives*

$$\Gamma(x) = \frac{x^x}{\mathrm{e}^x}\int_0^{\infty} \left(s\mathrm{e}^{1-s}\right)^x ds.$$

Now observe that the main contribution is from the maximum at $s = 1$*. Split the integral into a neighborhood of* $s = 1$ *and the rest. The rest does not contribute asymptotically. To determine the contribution from the neighborhood make a change of variables* $s = r(s)$ *such that* $s\mathrm{e}^{1-s} = \mathrm{e}^{-r^2/2}$*.)*

Problem 2.30. *Let* $U \subseteq \mathbb{R}^m$ *be open and let* $f : U \to \mathbb{R}^n$ *be locally Lipschitz (i.e., for every compact set* $K \subset U$ *there is some constant* L *such that* $|f(x) - f(y)| \le L|x - y|$ *for all* $x, y \in K$*). Show that if* $A \subset U$ *has Lebesgue measure zero, then* $f(A)$ *is contained in a set of Lebesgue measure zero. (Hint: By Lindelöf it is no restriction to assume that* A *is contained in a compact ball contained in* U*. Now approximate* A *by a union of rectangles.)*

---

[7] Adrien-Marie Legendre (1752–1833), French mathematician

[8] James Stirling (1692–1770), Scottish mathematician

2.4. Surface measure and the Gauss–Green theorem

We begin by recalling the definition of an m-dimensional submanifold: We will call a subset $\Sigma \subseteq \mathbb{R}^n$ an m-dimensional **submanifold** if there is a parametrization $\varphi \in C^1(U, \mathbb{R}^n)$, where $U \subseteq \mathbb{R}^m$ is open, $\Sigma = \varphi(U)$, and φ is an immersion (i.e., the Jacobian is injective at every point). Somewhat more general one extends this definition to the case where a parametrization only exists locally in the sense that every point $z \in \Sigma$ has a neighborhood W such that there is a parametrization for $W \cap \Sigma$.

Moreover, given a parametrization near a point $z^0 \in \Sigma$, the assumption that the Jacobian $\frac{\partial \varphi}{\partial x}$ is injective implies that, after a permutation of the coordinates, the first m vectors of the Jacobian are linearly independent. Hence, after restricting U, we can assume that $(\varphi_1, \dots, \varphi_m)$ is invertible and hence there is a parametrization of the form

$$\phi(x) = (x_1, \dots, x_m, \phi_{m+1}(x), \dots, \phi_n(x)) \tag{2.48}$$

(up to a permutation of the coordinates in $\mathbb{R}^n$). We will require for all parameterizations U to be so small that this is possible.

Given a submanifold Σ and a parametrization $\varphi : U \subseteq \mathbb{R}^m \to \Sigma \subseteq \mathbb{R}^n$ let

$$\Gamma(\partial\varphi) := \Gamma(\partial_1\varphi, \dots, \partial_m\varphi) = \det\left(\partial_j\varphi \cdot \partial_k\varphi\right)_{1 \le j,k \le m} \tag{2.49}$$

be the **Gram determinant**[9] of the tangent vectors (here the dot indicates a scalar product in $\mathbb{R}^n$) and define the **submanifold measure** dS via

$$\int_\Sigma g\, dS := \int_U g(\varphi(x))\sqrt{\Gamma(\partial\varphi(x))}d^m x. \tag{2.50}$$

Note that this can be geometrically motivated since the Gram determinant can be interpreted as the square of the volume of the parallelotope $\{\sum_{j=1}^m \alpha_j \partial_j\varphi | 0 \le \alpha_j \le 1\}$ formed by the tangent vectors, which is just the linear approximation to the surface at the given point. Indeed, defining the volume recursively by the volume of the base $\{\sum_{j=1}^{m-1} \alpha_j \partial_j\varphi | 0 \le \alpha_j \le 1\}$ times the height (i.e., the distance of $\partial_j\varphi_m$ from $\operatorname{span}\{\partial_1\varphi, \dots, \partial_{m-1}\varphi\}$), this follows from Problem 2.31.

If $\phi : V \subseteq \mathbb{R}^m \to \Sigma \subseteq \mathbb{R}^n$ is another parametrization, and hence $f = \phi^{-1} \circ \varphi \in C^1(U, V)$ is a diffeomorphism, the change of variables formula gives

$$\begin{aligned}\int_V g(\phi(x))\sqrt{\Gamma(\partial\phi(y))}d^m y &= \int_U g(\phi(f(x)))\sqrt{\Gamma(\partial\phi(f(x)))}|J_f(x)|d^n x \\ &= \int_U g(\varphi(x))\sqrt{\Gamma(\partial\varphi(x))}d^m x,\end{aligned}$$

[9] Jørgen Pedersen Gram (1850–1916), Danish actuary and mathematician

where we have used the chain rule $\frac{\partial\varphi}{\partial x}(x) = \frac{\partial(\phi\circ f)}{\partial x}(f(x)) = \frac{\partial\phi}{\partial y}(f(x))\frac{\partial f}{\partial x}(x)$ in the last step. Hence our definition is independent of the parametrization chosen. If our submanifold cannot be covered by a single parametrization, we choose a partition into countably many measurable subsets A_j such that for each A_j there is a parametrization (U_j, φ_j) such that $A_j \subseteq \varphi_j(U_j)$. Then we set

$$\int_\Sigma g\, dS := \sum_j \int_{\varphi_j^{-1}(A_j)} g(\varphi_j(x))\sqrt{\Gamma(\partial\varphi_j(x))}d^m x. \tag{2.51}$$

Note that given a different splitting B_k with parameterizations (V_k, ϕ_k) we can first change to a common refinement $A_j \cap B_k$ and then conclude that the individual integrals are equal by our above calculation. Hence again our definition is independent of the splitting and the parametrization chosen.

Example 2.17. Let T_n be spherical coordinates from Example 2.14 and let $S_n(\varphi, \theta_1, \dots, \theta_{n-2}) = T_n(1, \varphi, \theta_1, \dots, \theta_{n-2})$ be the corresponding parametrization of the unit sphere. Then one computes

$$\det(\partial T_n)^2 = \Gamma(\partial T_n) = r^{2n-2}\,\Gamma(\partial S_n)$$

since $\partial_r T^n \cdot \partial_r T^n = 1$, $\partial_r T^n \cdot \partial_\varphi T^n = \partial_r T^n \cdot \partial_{\theta_j} T^n = 0$. Hence dS_n coincides with $d\sigma^{n-1}$ from (2.44). ⋄

In the case $m = n-1$ a submanifold is also known as a **(hyper-)surface**. Given a surface and a parametrization, a **normal vector** is given by

$$\tilde\nu := \big(\det(\partial_1\varphi, \dots, \partial_{n-1}\varphi, \delta_1), \dots, \det(\partial_1\varphi, \dots, \partial_{n-1}\varphi, \delta_n)\big), \tag{2.52}$$

where δ_j are the canonical basis vectors in $\mathbb{R}^n$ (it is straightforward to check that $\tilde\nu \cdot \partial_j\varphi = 0$ for $1 \le j \le n-1$). Its length is given by (Problem 2.32)

$$|\tilde\nu|^2 = \Gamma(\partial\varphi) \tag{2.53}$$

and the unit normal is given by

$$\nu := \frac{1}{\sqrt{\Gamma(\partial\varphi)}}\tilde\nu. \tag{2.54}$$

It is uniquely defined up to orientation. Moreover, given a vector field $u : \Sigma \to \mathbb{R}^n$ (or $\mathbb{C}^n$) we have

$$\int_\Sigma u \cdot \nu\, dS = \int_U \det(\partial_1\varphi, \dots, \partial_{n-1}\varphi, u\circ\varphi)d^{n-1}x. \tag{2.55}$$

Here we will mainly be interested in the case of a surface arising as the boundary of some open domain $\Omega \subset \mathbb{R}^n$. To this end we recall that $\Omega \subseteq \mathbb{R}^n$ is said to have a C^1 boundary if around any point $x^0 \in \partial\Omega$ we can find a small neighborhood $O(x^0)$ so that, after a possible permutation of the coordinates, we can write

$$\Omega \cap O(x^0) = \{x \in O(x^0) | x_n > \gamma(x_1, \dots, x_{n-1})\} \tag{2.56}$$

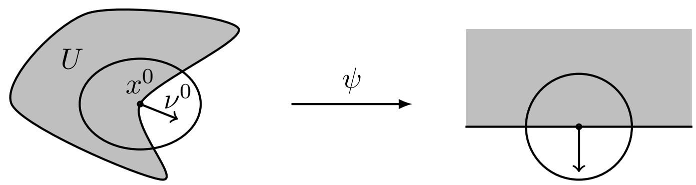


Figure 2.1. Straightening out the boundary

with $\gamma \in C^1$. Similarly we could define C^k or $C^{k,\theta}$ domains. According to our definition above, $\partial\Omega$ is then a surface in $\mathbb{R}^n$ and we have

$$\partial\Omega \cap O(x^0) = \{x \in O(x^0) | x_n = \gamma(x_1, \dots, x_{n-1})\}. \tag{2.57}$$

In this case our coordinate patch reads

$$\varphi(x_1, \dots, x_{n-1}) = (x_1, \dots, x_{n-1}, \gamma(x_1, \dots, x_{n-1})), \tag{2.58}$$

the **outward pointing unit normal vector** is

$$\nu := \frac{1}{\sqrt{1 + (\partial_1\gamma)^2 + \cdots + (\partial_{n-1}\gamma)^2}} (\partial_1\gamma, \dots, \partial_{n-1}\gamma, -1), \tag{2.59}$$

and hence the surface integral reads

$$\int_\Sigma u \cdot \nu \, dS = \int_U u(x_1, \dots, x_{n-1}, \gamma) \cdot (\partial_1\gamma, \dots, \partial_{n-1}\gamma, -1) d^{n-1}x. \tag{2.60}$$

The surface measure dS follows by choosing $u = \nu$:

$$dS = \sqrt{1 + (\partial_1\gamma)^2 + \cdots + (\partial_{n-1}\gamma)^2} d^{n-1}x. \tag{2.61}$$

Moreover, we have a change of coordinates $y = \psi(x)$ such that in these coordinates the boundary is given by (part of) the hyperplane $y_n = 0$. Explicitly we have $\psi \in C^1_b(U \cap O(x^0), V_+(y^0))$ given by

$$\psi(x) = (x_1, \dots, x_{n-1}, x_n - \gamma(x_1, \dots, x_{n-1})) \tag{2.62}$$

with inverse $\psi^{-1} \in C^1_b(V_+(y^0), U \cap O(x^0))$ given by

$$\psi^{-1}(y) = (y_1, \dots, y_{n-1}, y_n + \gamma(y_1, \dots, y_{n-1})). \tag{2.63}$$

Clearly, $\nu = (0, \dots, 0, -1)$ and $dS = d^{n-1}y$ in the new coordinates. This is known as straightening out the boundary (see Figure 2.1). Moreover, at every point of the boundary we have the **outward pointing unit normal vector** $\nu(x^0)$ which, in the above setting, is given as

$$\nu(x^0) := \frac{1}{\sqrt{1 + (\partial_1\gamma)^2 + \cdots + (\partial_{n-1}\gamma)^2}} (\partial_1\gamma, \dots, \partial_{n-1}\gamma, -1). \tag{2.64}$$

If we straighten out the boundary, then clearly, $\nu(y^0) = (0, \dots, 0, -1)$.

Theorem 2.19 (Gauss–Green[10]). *If Ω is a bounded C^1 domain in $\mathbb{R}^n$ and $u \in C^1(\overline{\Omega}, \mathbb{R}^n)$ is a vector field, then*

$$\int_\Omega (\operatorname{div} u) d^n x = \int_{\partial\Omega} u \cdot \nu \, dS. \tag{2.65}$$

Here $\operatorname{div} u = \sum_{j=1}^n \partial_j u_j$ *is the* ***divergence*** *of a vector field.*

This theorem is also known as the **divergence theorem** or as **Ostrogradski formula**. In the one-dimensional case it is just the fundamental theorem of calculus.

Proof. By linearity it suffices to prove

$$\int_\Omega (\partial_j f) d^n x = \int_{\partial\Omega} f \nu_j dS, \qquad 1 \le j \le n, \tag{2.66}$$

for $f \in C^1(\overline{\Omega})$. We first suppose that f is supported in a neighborhood $O(x^0)$ as in (2.56). We also assume that $O(x^0)$ is a rectangle. Let $O = O(x^0) \cap \partial\Omega$. Then for $j = n$ we have

$$\begin{aligned}\int_\Omega (\partial_n f) d^n x &= \int_O \left(\int_{x_n \ge \gamma(x')} \partial_n f(x', x_n) dx_n \right) d^{n-1} x' \\ &= -\int_O f(x', \gamma(x')) d^{n-1} x' = \int_{\partial\Omega} f \nu_n dS,\end{aligned}$$

where we have used Fubini and the fundamental theorem of calculus. For $j < n$ let us assume that we have just $n = 2$ to simplify notation (as the other coordinates will not affect the calculation). Then $O(x^0) = (a_1, b_1) \times (a_2, b_2)$ and we have (by the fundamental theorem of calculus and the Leibniz integral rule — Problem 2.35)

$$\begin{aligned}0 &= \int_{a_1}^{b_1} \partial_1 \int_{\gamma(x_1)}^{b_2} f(x_1, x_2) dx_2 dx_1 \\ &= \int_{a_1}^{b_1} \int_{\gamma(x_1)}^{b_2} \big(\partial_1 f(x_1, x_2)\big) dx_2 dx_1 - \int_{a_1}^{b_1} f(x_1, \gamma(x_1)) \partial_1 \gamma(x_1) dx_1\end{aligned}$$

from which the claim follows.

For the general case cover $\overline{\Omega}$ by rectangles which either contain no boundary points or otherwise are as in (2.56). By compactness there is a finite subcover. Choose a smooth partition of unity ζ_j subordinate to this cover (Lemma B.30 from [**22**]) and consider $f = \sum_j \zeta_j f$. Then for each summand having support in a rectangle intersecting the boundary, the claim holds by the above computation. Similarly, for each summand having support in

[10] Carl Friedrich Gauss (1777–1855), German mathematician and physicist
[10] George Green (1793–1841), British mathematical physicist

an interior rectangle, Fubini and the fundamental theorem of calculus show $\int_\Omega (\partial_n \zeta_j f) d^n x = 0$. $\square$

The formulation of the theorem suggests that it should hold under the weaker assumption $u \in C^1(U) \cap C(\overline{U})$. However, the problem is that under this assumption the derivatives $\partial_j u$ might no longer be integrable (cf. Problem 4.32) and hence it is no longer clear how the integral on the left-hand should be understood. If one adds an extra assumption ensuring integrability of the derivatives (e.g. $u \in C_b^1(U) \cap C(\overline{U})$) this will indeed hold as we will show in Lemma 7.24.

Example 2.18. Let us verify the Gauss–Green theorem for the unit ball in $\mathbb{R}^3$. By linearity it suffices to consider the case where the vector field u is parallel to one of the coordinate axes, say $u = (0, 0, u_3)$ such that $\operatorname{div} u = \partial_3 u_3$. Abbreviating $\rho := \sqrt{x_1^2 + x_2^2}$ we obtain

$$\int_{B_1} (\operatorname{div} u) d^3 x = \int_{\rho \le 1} \int_{-\sqrt{1-\rho^2}}^{\sqrt{1-\rho^2}} \frac{\partial u_3}{\partial x_3}(x) dx_3 \, d(x_1, x_2)$$
$$= \int_{\rho \le 1} \Big(u_3(x_1, x_2, \sqrt{1-\rho^2}) - u_3(x_1, x_2, -\sqrt{1-\rho^2}) \Big) d(x_1, x_2).$$

Parametrizing the upper/lower hemisphere $S_\pm^2 := \{x | \pm x_3 > 0, |x| = 1\}$ using $x_3 = \pm\sqrt{1-\rho^2}$ we obtain $dS = \frac{1}{\sqrt{1-\rho^2}} d(x_1, x_2)$. Since $\nu = \frac{x}{|x|}$ (remember that ν needs to point outwards) this gives

$$\int_{S_\pm^2} u \cdot \nu \, dS = \pm \int_{\rho \le 1} u_3(x_1, x_2, \pm\sqrt{1-\rho^2}) d(x_1, x_2)$$

and verifies the Gauss–Green theorem for the unit ball in $\mathbb{R}^3$. Of course the calculation easily generalizes to $\mathbb{R}^n$. $\diamond$

Applying the Gauss–Green theorem to a product fg we obtain

Corollary 2.20 (Integration by parts). *We have*

$$\int_\Omega (\partial_j f) g \, d^n x = \int_{\partial\Omega} f g \nu_j dS - \int_\Omega f (\partial_j g) d^n x, \qquad 1 \le j \le n, \tag{2.67}$$

for $f, g \in C^1(\overline{\Omega})$.

Problem* 2.31. *Given some vectors* $a_1, \dots, a_m \in \mathbb{R}^n$ *let* $A = (a_1, \cdots, a_m)$ *be the matrix formed from these vectors. We define their* ***Gram determinant*** *as*

$$\Gamma(a_1, \dots, a_m) := \det(A^T A) = \det\left(a_j \cdot a_k\right)_{1 \le j,k \le m}.$$

Note that in the case of $m = n$ *vectors we have*

$$\Gamma(a_1, \dots, a_n) = \det(A)^2.$$

Show that the Gram determinant is nonzero if and only if the vectors are linearly independent. Moreover, show that in this case

$$\operatorname{dist}(b, \operatorname{span}\{a_1, \dots, a_m\})^2 = \frac{\Gamma(a_1, \dots, a_m, b)}{\Gamma(a_1, \dots, a_m)}$$

and

$$\Gamma(a_1, \dots, a_m) \le \prod_{j=1}^m |a_j|^2.$$

with equality if the vectors are orthogonal. (Hint: First establish $\Gamma(f_1, \dots, f_j + \alpha f_k, \dots, f_n) = \Gamma(f_1, \dots, f_n)$ *for* $j \neq k$ *and use it to investigate how* Γ *changes when you apply the Gram–Schmidt procedure?)*

Problem* 2.32. *Show* (2.53). *(Hint: Problem 2.31)*

Problem 2.33. *Verify the Gauss–Green theorem (by computing both integrals) in the case* $u(x) = x$ *and* $U = B_1(0) \subset \mathbb{R}^n$.

Problem 2.34. *Let* Ω *be a bounded* C^1 *domain in* $\mathbb{R}^n$ *and set* $\frac{\partial g}{\partial \nu} := \nu \cdot \partial g$. *Verify* ***Green's first identity***

$$\int_\Omega (f \Delta g + \partial f \cdot \partial g) d^n x = \int_{\partial\Omega} f \frac{\partial g}{\partial \nu} dS$$

for $f \in C^1(\overline{\Omega})$, $g \in C^2(\overline{\Omega})$ *and* ***Green's second identity***

$$\int_\Omega (f \Delta g - g \Delta f) d^n x = \int_{\partial\Omega} \left(f \frac{\partial g}{\partial \nu} - g \frac{\partial f}{\partial \nu} \right) dS$$

for $f, g \in C^2(\overline{\Omega})$.

Problem* 2.35 (Leibniz integral rule)**.** *Suppose* $f \in C(R)$ *with* $\frac{\partial f}{\partial x}(x,y) \in C(R)$, *where* $R = [a_1, b_1] \times [a_2, b_2]$ *is some rectangle, and* $g \in C^1([a_1, b_1], [a_2, b_2])$. *Show*

$$\frac{d}{dx} \int_{a_2}^{g(x)} f(x,y) dy = f(x, g(x)) g'(x) + \int_{a_2}^{g(x)} \frac{\partial f}{\partial x}(x,y) dy.$$

2.5. 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

$$f(x\pm) := \lim_{\varepsilon\downarrow 0} f(x \pm \varepsilon). \tag{2.68}$$

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 $f(\pm\infty\pm) = f(\pm\infty)$.

If we want to define an inverse, problems will occur at points where f jumps and on intervals where f is constant. Informally 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

$$\Gamma(f) := \{(x, y) | y \in [f(x-), f(x+)]\}. \tag{2.69}$$

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 $\Gamma(f)$ is nondecreasing in the sense that $x_1 < x_2$ implies $y_1 \le y_2$ for $(x_1, y_1), (x_2, y_2) \in \Gamma(f)$ (just note $y_1 \le f(x_1+) \le f(x_2-) \le y_2$). 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

$$\Gamma(f)^{-1} := \{(y, x) | (x, y) \in \Gamma(f)\} \tag{2.70}$$

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

$$\Gamma(f^{-1}) = \Gamma(f)^{-1} \tag{2.71}$$

since both are maximal. In particular, it follows that if f^{-1} is 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

$$\begin{aligned} f_-^{-1}(y) &:= \inf f^{-1}([y,\infty)) = \sup f^{-1}((-\infty,y)), \\ f_+^{-1}(y) &:= \inf f^{-1}((y,\infty)) = \sup f^{-1}((-\infty,y]). \end{aligned} \tag{2.72}$$

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 f^{-1} extends the usual inverse by setting it constant on the gaps missing in the range of f. In particular we have $f^{-1}(f(x)) = x$ and $f(f^{-1}(y)) = y$ 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 2.21. *Let f 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) *$f_-^{-1}(f(x)) \le x \le f_+^{-1}(f(x))$ with equality on the left, right iff f is not constant to the right, left of x, respectively.*

(iii) *$f(f^{-1}(y)-) \le y \le f(f^{-1}(y)+)$ with equality on the left, right iff f^{-1} is not constant to the right, left of y, respectively.*

Proof. Item (i) follows since both claims are equivalent to $y \le f(\tilde{x})$ for all $\tilde{x} > x$. Similarly for (i'). Item (ii) follows from $f_-^{-1}(f(x)) = \inf f^{-1}([f(x),\infty)) = \inf\{\tilde{x}|f(\tilde{x}) \ge f(x)\} \le x$ with equality iff $f(\tilde{x}) < f(x)$ for $\tilde{x} < x$. Similarly for the other inequality. Item (iii) follows by reversing the roles of f and f^{-1} in (ii). □

In particular, $f(f^{-1}(y)) = y$ if f is continuous. We will also need the set

$$L(f) := \{y | f^{-1}((y,\infty)) = (f_+^{-1}(y), \infty)\}. \tag{2.73}$$

Note that $y \notin L(f)$ if and only if there is some x such that $y \in [f(x-), f(x))$.

Lemma 2.22. *Let $m : \mathbb{R} \to \mathbb{R}$ be a nondecreasing function on $\mathbb{R}$ and μ its associated measure via* (1.19). *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 2.19. 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 2.23. *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*

$$\int g\, d(m_\pm \circ f^{-1}) = \int (g \circ f_\mp)\, dm. \tag{2.74}$$

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

$$\int g\, dh = \int g(h^{-1}(y))dy. \tag{2.75}$$

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

$$i_m(y) := m(m_-^{-1}(y)). \tag{2.76}$$

Note that $i_m(y) = y$ if m is continuous. By $\mathrm{conv}(\mathrm{Ran}(m))$ we denote the convex hull of the range of m.

Corollary 2.24. *Suppose m, n are two nondecreasing functions on $\mathbb{R}$ with n right continuous. Then we have*

$$\int_{\mathbb{R}} (g \circ m)\, d(n \circ m) = \int_{\mathrm{conv}(\mathrm{Ran}(m))} (g \circ i_m) dn \tag{2.77}$$

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 2.25. *Suppose m, n are nondecreasing functions on $\mathbb{R}$ and g is monotone. Then we have*

$$\int_{\mathbb{R}} (g \circ m)\, d(n \circ m) \le \int_{\mathrm{conv}(\mathrm{Ran}(m))} g\, dn \tag{2.78}$$

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 2.21. □

Problem* 2.36. *Show* (2.72).

Problem 2.37. *Show that* $\Gamma(f) \circ \Gamma(f^{-1}) = \{(y,z) | y, z \in [f(x-), f(x+)]$ *for some* $x\}$ *and* $\Gamma(f^{-1}) \circ \Gamma(f) = \{(y,z) | f(y+) > f(z-)$ *or* $f(y-) < f(z_+)\}$.

Problem 2.38. *Let* $d\mu(\lambda) := \chi_{[0,1]}(\lambda) d\lambda$ *and* $f(\lambda) := \chi_{(-\infty,t]}(\lambda)$, $t \in \mathbb{R}$. *Compute* $f_\star \mu$.

2.6. 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

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b. \tag{2.79}$$

The number

$$\|P\| := \max_{1 \le j \le n} x_j - x_{j-1} \tag{2.80}$$

is called the norm of P. Given a partition P and a bounded real-valued function f we can define

$$s_{P,f,-}(x) := \sum_{j=1}^n m_j \chi_{[x_{j-1},x_j)}(x), \qquad m_j := \inf_{x \in [x_{j-1},x_j]} f(x), \tag{2.81}$$

$$s_{P,f,+}(x) := \sum_{j=1}^n M_j \chi_{[x_{j-1},x_j)}(x), \qquad M_j := \sup_{x \in [x_{j-1},x_j]} f(x), \tag{2.82}$$

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,

$$m \le s_{P,f,-}(x) \le f(x) \le s_{P,f,+}(x) \le M, \quad m := \inf_{x \in [a,b]} f(x), M := \sup_{x \in [a,b]} f(x). \tag{2.83}$$

Moreover, we can define the upper and lower Riemann sum associated with P as

$$L(P,f) := \sum_{j=1}^n m_j (x_j - x_{j-1}), \qquad U(P,f) := \sum_{j=1}^n M_j (x_j - x_{j-1}). \tag{2.84}$$

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

$$m\,(b-a) \le L(P,f) \le U(P,f) \le M\,(b-a). \tag{2.85}$$

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

$$s_{P_1,f,-}(x) \le s_{P_2,f,-}(x) \le f(x) \le s_{P_2,f,+}(x) \le s_{P_1,f,+}(x) \tag{2.86}$$

as well as

$$L(P_1, f) \le L(P_2, f) \le U(P_2, f) \le U(P_1, f). \tag{2.87}$$

Hence we define the **lower, upper Riemann integral** of f as

$$\underline{\int} f(x)dx := \sup_P L(P, f), \qquad \overline{\int} f(x)dx := \inf_P U(P, f), \tag{2.88}$$

respectively. Since for arbitrary partitions P and Q we have

$$L(P, f) \le L(P \cup Q, f) \le U(P \cup Q, f) \le U(Q, f). \tag{2.89}$$

we obtain

$$m\,(b - a) \le \underline{\int} f(x)dx \le \overline{\int} f(x)dx \le M\,(b - a). \tag{2.90}$$

We will call f **Riemann integrable** if both values coincide and the common value will be called the **Riemann integral** of f.

Example 2.20. 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 2.39). ◇

Example 2.21. If we enumerate the rational numbers in $[0, 1]$, $\mathbb{Q} \cap [0, 1] =: \{x_n\}_{n\in\mathbb{N}}$, we can define $f_n := \chi_{\{x_1,\dots,x_n\}}$. Then it is straightforward to check that f_n is Riemann integrable (cf. also Theorem 2.27 below). Since $f_n(x) \to \chi_{\mathbb{Q}}(x)$ pointwise, this example shows that the pointwise limit of Riemann integrable functions is not Riemann integrable in general and demonstrates the drawbacks of the Riemann integral with respect to limit operations. ◇

Example 2.22. Let f nondecreasing, then f is integrable. In fact, since $m_j = f(x_{j-1})$ and $M_j = f(x_j)$ we obtain

$$U(f, P) - L(f, P) \le \|P\| \sum_{j=1}^{n} (f(x_j) - f(x_{j-1})) = \|P\|(f(b) - f(a))$$

and the claim follows (cf. also the next lemma). Similarly nonincreasing functions are integrable. ◇

Lemma 2.26. *A function f is Riemann integrable if and only if there exists a sequence of partitions P_n such that*

$$\lim_{n\to\infty} L(P_n, f) = \lim_{n\to\infty} U(P_n, f). \tag{2.91}$$

In this case the above limits equal the Riemann integral of f and P_n can be chosen such that $P_n \subseteq P_{n+1}$ and $\|P_n\| \to 0$.

Proof. If there is such a sequence of partitions then f is integrable by $\lim_n L(P_n, f) \le \sup_P L(P, f) \le \inf_P U(P, f) \le \lim_n U(P_n, f)$.

Conversely, given an integrable f, there is a sequence of partitions $P_{L,n}$ such that $\underline{\int} f(x)dx = \lim_n L(P_{L,n}, f)$ and a sequence $P_{U,n}$ such that $\overline{\int} f(x)dx = \lim_n U(P_{U,n}, f)$. By (2.87) the common refinement $P_n = P_{L,n} \cup P_{U,n}$ is the partition we are looking for. Since, again by (2.87), any refinement will also work, the last claim follows. $\square$

Note that when computing the Riemann integral as in the previous lemma one could choose instead of m_j or M_j any value in $[m_j, M_j]$ (e.g. $f(x_{j-1})$ or $f(x_j)$).

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 2.27 (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 2.26. 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 2.6 $s_{f,+}(x) = s_{f,-}(x)$ 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. $\square$

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.

Finally, recall that if one endpoint is unbounded or f is unbounded near one endpoint, one defines the **improper Riemann integral** by taking limits towards this endpoint. More specifically, if f is Riemann integrable for every $(a,c) \subset (a,b)$ one defines

$$\int_a^b f(x)dx := \lim_{c\uparrow b} \int_a^c f(x)dx \tag{2.92}$$

with an analogous definition if f is Riemann integrable for every $(c,b) \subset (a,b)$. Note that in this case improper integrability no longer implies Lebesgue integrability unless $|f(x)|$ has a finite improper integral.

Example 2.23. The prototypical example being the **Dirichlet integral**[11]

$$\int_0^\infty \frac{\sin(x)}{x}dx = \lim_{c\to\infty} \int_0^c \frac{\sin(x)}{x}dx = \frac{\pi}{2}$$

(cf. Problem 8.31) which does not exist as a Lebesgue integral since

$$\int_0^\infty \frac{|\sin(x)|}{x}dx \ge \sum_{k=0}^\infty \int_{\pi/4}^{3\pi/4} \frac{|\sin(k\pi + x)|}{k\pi + 3/4}dx \ge \frac{1}{2\sqrt{2}} \sum_{k=1}^\infty \frac{1}{k} = \infty. \qquad \diamond$$

Problem* 2.39. *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.

Problem 2.40. *Prove that the Riemann integral is linear: If f, g are Riemann integrable and $\alpha \in \mathbb{R}$, then αf and $f+g$ are Riemann integrable with $\int(f+g)dx = \int f\,dx + \int g\,dx$ and $\int \alpha f\,dx = \alpha \int f\,dx$.*

Problem 2.41. *Suppose f is Riemann integrable and ϕ is Lipschitz continuous on the range of f, then $\phi \circ f$ is Riemann integrable. Moreover, show that if f, g are Riemann integrable, so is fg. (Hint: The second claim can be reduced to the first using $\phi(x) = x^2$.)*

Problem 2.42. *Show that the uniform limit of Riemann integrable functions is again Riemann integrable. Conclude that in the previous problem it suffices to assume that ψ is continuous.*

[11]Peter Gustav Lejeune Dirichlet (1805 –1859), German mathematician

Problem 2.43. *Let* $\{q_n\}_{n\in\mathbb{N}}$ *be an enumeration of the rational numbers in* $[0,1)$. *Show that*

$$f(x) := \sum_{n\in\mathbb{N}:q_n<x} \frac{1}{2^n}$$

is discontinuous at every q_n *but still Riemann integrable.*

Chapter 3

The Lebesgue spaces L^p

3.1. Functions almost everywhere

We fix some measure space (X, Σ, μ) and define the L^p norm by

$$\|f\|_p := \left(\int_X |f|^p \, d\mu \right)^{1/p}, \qquad 1 \le p, \tag{3.1}$$

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 2.6 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 3.1. *Let f be measurable. Then*

$$\int_X |f|^p \, d\mu = 0 \tag{3.2}$$

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

$$\mathcal{N}(X, d\mu) := \{f | f(x) = 0 \ \mu\text{-almost everywhere}\}. \tag{3.3}$$

Then $\mathcal{N}(X, d\mu)$ is a linear subspace of $\mathcal{L}^p(X, d\mu)$ and we can consider the quotient space

$$L^p(X, d\mu) := \mathcal{L}^p(X, d\mu)/\mathcal{N}(X, d\mu). \tag{3.4}$$

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 treat them as functions for notational convenience. However, if we do so, it is important to ensure that every statement made does not depend on the representative in the equivalence classes. In particular, note that for $f \in L^p(X, d\mu)$ the value $f(x)$ is not well defined. However, there are situations where a well-defined value $f(x)$ can be assigned. For example, if f has a continuous representative (and continuous functions with different values 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[1] space with scalar product given by

$$\langle f, g \rangle := \int_X f(x)^* g(x) d\mu(x). \tag{3.5}$$

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**

$$\|f\|_\infty := \inf\{C \,|\, \mu(\{x| \, |f(x)| > C\}) = 0\}. \tag{3.6}$$

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 3.1. 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). $\diamond$

As before we set

$$L^\infty(X, d\mu) := B(X)/\mathcal{N}(X, d\mu) \tag{3.7}$$

and observe that $\|f\|_\infty$ is independent of the representative from the equivalence class.

If you wonder where the ∞ comes from, have a look at Problem 3.2.

Since the support of a *function* in L^p is also not well defined one uses the **essential support** in this case:

$$\operatorname{supp}(f) = X \setminus \bigcup \{O | f = 0 \ \mu\text{-almost everywhere on } O \subseteq X \text{ open}\}. \tag{3.8}$$

[1] David Hilbert (1862–1943), German mathematician

In other words, x is in the essential support if for every neighborhood the set of points where f does not vanish has positive measure. Here we use the same notation as for functions and it should be understood from the context which one is meant. Note that the essential support is always smaller than the support (since we get the latter if we require f to vanish everywhere on O in the above definition).

Example 3.2. The support of $\chi_{\mathbb{Q}}$ is $\overline{\mathbb{Q}} = \mathbb{R}$ but the essential support with respect to Lebesgue measure is $\emptyset$ since the function is 0 a.e. $\diamond$

If X is a locally compact Hausdorff space (together with the Borel sigma algebra), a function is called **locally integrable** if it is integrable when restricted to any compact subset $K \subseteq X$. The set of all (equivalence classes of) locally integrable functions will be denoted by $L^1_{loc}(X, d\mu)$. We will say that $f_n \to f$ in $L^1_{loc}(X, d\mu)$ if this holds on $L^1(K, d\mu)$ for all compact subsets $K \subseteq X$. Of course this definition extends to L^p for any $1 \le p \le \infty$.

Problem* 3.1. *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* 3.2. *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 3.3. *Consider $X := (0, 1)$ with Lebesgue measure. Is it true that*

$$\bigcap_{1\le p<\infty} L^p(0, 1) = L^\infty(0, 1)?$$

Problem 3.4. *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.)*

Problem 3.5. *Show that $\mu(\{x| \, |f(x)| > \|f\|_\infty\}) = 0$.*

Problem 3.6. *Show that for a continuous function on $\mathbb{R}^n$ the support and the essential support with respect to Lebesgue measure coincide.*

3.2. Jensen $\le$ Hölder $\le$ Minkowski

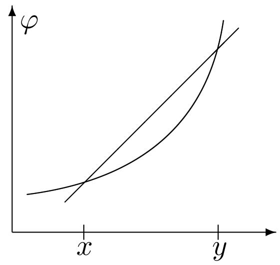
As a preparation for proving that L^p is a Banach space, we will need Hölder's inequality, which plays a central role in the theory of L^p spaces. In particular, Hölder's inequality will imply Minkowski's inequality, which is just the triangle inequality for L^p. Our proof is based on Jensen's inequality and emphasizes the connection with convexity. In fact, the triangle inequality just states that a norm is convex:

$$\|(1-\lambda)f + \lambda\, g\| \le (1-\lambda)\|f\| + \lambda\|g\|, \quad \lambda \in (0,1). \tag{3.9}$$

Recall that a real function φ defined on an open interval (a,b) is called **convex** if

$$\varphi((1-\lambda)x + \lambda y) \le (1-\lambda)\varphi(x) + \lambda\varphi(y), \quad \lambda \in (0,1), \tag{3.10}$$

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**. A function φ is **concave** if $-\varphi$ is convex.

Lemma 3.2. *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 (use $z = (1-\lambda)x + \lambda y$) that the definition implies

$$D(x,z) \le D(x,y) \le D(y,z), \quad x < z < y,$$

where the inequalities are strict if φ is strictly convex. 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). $\square$

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., $\varphi'' \geq 0$ if $\varphi \in C^2$) — Problem 3.7.

With these preparations out of the way we can show

Theorem 3.3 (Jensen's inequality[2]). *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*

$$\varphi\Big(\int_X f\, d\mu\Big) \leq \int_X (\varphi \circ f)\, d\mu. \tag{3.11}$$

For $f \geq 0$ the requirement that f is integrable can be dropped if $\varphi(b)$ is understood as $\lim_{x\to b} \varphi(x)$. Similarly, if $\varphi(x)$ depends only on the absolute value of x, finiteness of the right-hand side will imply integrability of f.

Proof. By (iii) of the previous lemma we have

$$\varphi(f(x)) \geq \varphi(I) + \alpha(f(x) - I), \qquad I = \int_X f\, d\mu \in (a,b).$$

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 $f \geq 0$ we note that for $X_n = \{x \in X | f(x) \leq n\}$ the first part implies

$$\varphi\Big(\frac{1}{\mu(X_n)} \int_{X_n} f\, d\mu\Big) \leq \frac{1}{\mu(X_n)} \int_{X_n} \varphi(f)\, d\mu.$$

Taking $n \to \infty$ the claim follows from $X_n \nearrow X$ and the monotone convergence theorem. If $\varphi(x)$ depends only on the absolute value, we can replace f by $|f|$ to conclude that f is integrable. □

Observe that if φ is strictly convex, then equality can only occur if f is constant.

Now we are ready to prove

Theorem 3.4 (Hölder's inequality[3]). *Let p and q be **dual** indices; that is,*

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{3.12}$$

with $1 \leq p \leq \infty$. 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

$$\|f\, g\|_1 \leq \|f\|_p \|g\|_q. \tag{3.13}$$

[2] Johan Jensen (1859–1925)), Danish mathematician and engineer

[3] Otto Hölder (1859–1937), German mathematician

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. For a generalization see Problem 3.10. Moreover, in the case $1 < p < \infty$ the function x^p is strictly convex and equality will occur precisely if $|f|$ is a multiple of $|g|^{q-1}$ or g is trivial. This gives us a

Corollary 3.5. *Consider $f \in L^p(X, d\mu)$ with $1 \le p < \infty$ and let q be the corresponding dual index, $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\|f\|_p = \sup_{\|g\|_q = 1} \left| \int_X f g \, d\mu \right|. \tag{3.14}$$

If every set of infinite measure has a subset of finite positive measure (e.g. if μ is σ-finite), then the claim also holds for $p = \infty$.

Proof. In the case $1 < p < \infty$ equality is attained for $g = c^{-1} \operatorname{sign}(f^*) |f|^{p-1}$, where $c = \| |f|^{p-1} \|_q = \|f\|_p^{p-1}$ (assuming $c > 0$ w.l.o.g.). In the case $p = 1$ equality is attained for $g = \operatorname{sign}(f^*)$. 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, by assumption on μ we can choose a subset $B_\varepsilon \subseteq A_\varepsilon$ with finite positive measure. Then $g_\varepsilon = \operatorname{sign}(f^*) \chi_{B_\varepsilon} / \mu(B_\varepsilon)$ satisfies $\int_X f g_\varepsilon \, d\mu \ge \|f\|_\infty - \varepsilon$. □

Of course it suffices to take the sup in (3.14) over a dense set of L^q (e.g. integrable simple functions — see Problem 3.23). Moreover, note that the extra assumption for $p = \infty$ is crucial since if there is a set of infinite measure which has no subset with finite positive measure, then every integrable function must vanish on this subset.

If it is not a priori known that $f \in L^p$, the following generalization will be useful.

Lemma 3.6. *Suppose μ is σ-finite. Let $1 \le p \le \infty$ with q the corresponding dual index, $\frac{1}{p} + \frac{1}{q} = 1$. If $f \notin L^p(X, d\mu)$ is measurable then there exists a* *sequence of simple functions* s_n *with* $\|s_n\|_q = 1$ *such that* $\mathrm{Re}(fs_n) \ge 0$ *and*

$$\lim_{n\to\infty} \int_X \mathrm{Re}(fs_n) d\mu = \infty.$$

In particular, for any measurable f *we have*

$$\|f\|_p = \sup_{s \text{ simple}, \; \|s\|_q=1} \left| \int_X f s \, d\mu \right| = \sup_{\|g\|_q=1} \left| \int_X f g \, d\mu \right|,$$

where we set $\int_X fg\, d\mu := \infty$ *if* fg *is not integrable.*

Proof. If $f \not\in L^p$ we can split f into nonnegative functions $f = f_1 - f_2 + \mathrm{i}(f_3 - f_4)$ as usual and assume $f_1 \not\in L^p$ without loss of generality. Now choose $\hat{s}_n \nearrow f_1$ as in (2.6). Moreover, since μ is σ-finite we can find $X_n \nearrow X$ with $\mu(X_n) < \infty$. Then $\tilde{s}_n = \chi_{X_n} \hat{s}_n$ will be in L^p and will still satisfy $\tilde{s}_n \nearrow f_1$. Now if $1 < p < \infty$ choose $s_n = (\int \tilde{s}_n^p d\mu)^{-1/q} \tilde{s}_n^{p-1} \in L^q$. Then

$$\begin{aligned}\int_X \mathrm{Re}(fs_n) d\mu = \int_X f_1 s_n d\mu &= \left(\frac{\int_X \tilde{s}_n^{p-1} f_1 d\mu}{\int_X \tilde{s}_n^p d\mu} \right)^{1/q} \left(\int_X \tilde{s}_n^{p-1} f_1 d\mu \right)^{1/p} \\ &\ge \left(\int_X \tilde{s}_n^{p-1} f_1 d\mu \right)^{1/p} \to \infty\end{aligned}$$

by monotone convergence. Similarly, if $p = 1$ set $s_n = \chi_{X_n \cap \mathrm{supp}(f_1)} \in L^\infty$ and observe $\int_X \mathrm{Re}(fs_n) d\mu = \int_{X_n} f_1 d\mu \to \infty$. If $p = \infty$ then for every n there is some m such that $\mu(A_n) > 0$, where $A_n = \{x \in X_m | f_1(x) \ge n\}$. Now use $s_n = \mu(A_n)^{-1} \chi_{A_n} \in L^1$ proceed as before.

If $\|f\|_p < \infty$ the final claim follows from the previous corollary since simple functions are dense (Problem 3.23). If $\|f\|_p = \infty$ it follows from the first part. □

As another consequence we get

Theorem 3.7 (Minkowski's integral inequality[4]). *Suppose,* μ *and* ν *are two* σ*-finite measures and* f *is* $\mu \otimes \nu$ *measurable. Let* $1 \le p \le \infty$. *Then*

$$\left\| \int_Y f(.,y) d\nu(y) \right\|_p \le \int_Y \|f(.,y)\|_p d\nu(y), \tag{3.15}$$

where the p*-norm is computed with respect to* μ. *In particular, this says that* $f(x,.)$ *is integrable for a.e* x *and* $\int_Y f(.,y) d\nu(y) \in L^p(X, d\mu)$ *if the integral on the right is finite.*

[4]Hermann Minkowski (1864–1909), German mathematician

Proof. Let $g \in L^q(X, d\mu)$ with $g \geq 0$ and $\|g\|_q = 1$. Then using Fubini

$$\begin{aligned}\int_X g(x) \int_Y |f(x,y)| d\nu(y) d\mu(x) &= \int_Y \int_X |f(x,y)| g(x) d\mu(x) d\nu(y) \\ &\leq \int_Y \|f(.,y)\|_p d\nu(y)\end{aligned}$$

and the claim follows from Lemma 3.6. □

In the special case where ν is supported on two points this reduces to the triangle inequality (our proof inherits the assumption that μ is σ-finite, but this can be avoided – Problem 3.9).

Corollary 3.8 (Minkowski's inequality)**.** *Let $f, g \in L^p(X, d\mu)$, $1 \leq p \leq \infty$. Then*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \tag{3.16}$$

This shows that $L^p(X, d\mu)$ is a normed vector space.

Note that Fatou's lemma implies that the norm is lower semi continuous $\|f\|_p \leq \liminf_{n\to\infty} \|f_n\|_p$ with respect to pointwise convergence (a.e.). The next lemma sheds some light on the missing part.

Lemma 3.9 (Brezis–Lieb[5])**.** *Let $1 \leq p < \infty$ and let $f_n \in L^p(X, d\mu)$ be a sequence which converges pointwise a.e. to f such that $\|f_n\|_p \leq C$. Then $f \in L^p(X, d\mu)$ and*

$$\lim_{n\to\infty} \left(\|f_n\|_p^p - \|f_n - f\|_p^p\right) = \|f\|_p^p. \tag{3.17}$$

In the case $p = 1$ we can replace $\|f_n\|_1 \leq C$ by $f \in L^1(X, d\mu)$.

Proof. As pointed out before $\|f\|_p \leq \liminf_{n\to\infty} \|f_n\|_p \leq C$ which shows $f \in L^p(X, d\mu)$. Moreover, one easyly checks the elementary inequality

$$\big||s + t|^p - |t|^p - |s|^p\big| \leq \varepsilon |t|^p + C_\varepsilon |s|^p$$

(note that by scaling it suffices to consider the case $s = 1$ and $|t| \leq 1$). Setting $t = f_n - f$ and $s = f$, bringing everything to the right-hand-side and applying Fatou gives

$$\begin{aligned}C_\varepsilon \|f\|_p^p &\leq \liminf_{n\to\infty} \int_X \left(\varepsilon |f_n - f|^p + C_\varepsilon |f|^p - \big||f_n|^p - |f - f_n|^p - |f|^p\big|\right) d\mu \\ &\leq \varepsilon (2C)^p + C_\varepsilon \|f\|_p^p - \limsup_{n\to\infty} \int_X \big||f_n|^p - |f - f_n|^p - |f|^p\big| d\mu.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary the claim follows. Finally, note that in the case $p = 1$ we can choose $\varepsilon = 0$ and $C_\varepsilon = 2$. □

[5] Haïm Brezis (*1944), French mathematician

[5] Elliott H. Lieb (*1932), American mathematical physicist

It might be more descriptive to write the conclusion of the lemma as

$$\|f_n\|_p^p = \|f\|_p^p + \|f_n - f\|_p^p + o(1), \tag{3.18}$$

which shows an important consequence:

Corollary 3.10. *Let $1 \le p < \infty$ and let $f_n \in L^p(X, d\mu)$ be a sequence which converges pointwise a.e. to f such that either $\|f_n\|_p \le C$ or $f \in L^1(X, d\mu)$ if $p = 1$. Then $\|f_n - f\|_p \to 0$ if and only if $\|f_n\|_p \to \|f\|_p$.*

Note that it even suffices to show $\limsup \|f_n\|_p \le \|f\|_p$ since $\|f\|_p \le \liminf \|f_n\|_p$ comes for free from Fatou as pointed out before.

The L^p spaces have another convenient property from a Banach space point of view if $1 < p < \infty$. To this end recall that in a Banach space B, the unit ball is convex by the triangle inequality. Moreover, B is called **strictly convex** if the unit ball is a strictly convex set, that is, if for any two points on the unit sphere their average is inside the unit ball. A more qualitative notion is to require that if two unit vectors $f, g \in B$ satisfy $\|f - g\| \ge \varepsilon$ for some $\varepsilon > 0$, then there is some $\delta > 0$ such that $\|\frac{f+g}{2}\| \le 1 - \delta$. In this case one calls B **uniformly convex**. We refer to Section 6.5 from **[22]** for further details.

For example, in a uniformly convex space the Radon–Riesz theorem (Theorem 6.19 from **[22]**) states that a weakly convergent sequence $f_n \rightharpoonup f$ converges in norm $f_n \to f$ if and only if $\limsup \|f_n\| \le \|f\|$.

Theorem 3.11 (Clarkson[6]). *Suppose $1 < p < \infty$, then $L^p(X, d\mu)$ is uniformly convex.*

Proof. As a preparation we note that strict convexity of $|.|^p$ implies that $|\frac{t+s}{2}|^p \le |\frac{|t|+|s|}{2}|^p < \frac{|t|^p+|s|^p}{2}$ and hence

$$\rho(\varepsilon) := \min \left\{ \frac{|t|^p + |s|^p}{2} - \left|\frac{t+s}{2}\right|^p \Big| |t|^p + |s|^p = 2, \left|\frac{t-s}{2}\right|^p \ge \varepsilon \right\} > 0.$$

Hence, by scaling,

$$\left|\frac{t-s}{2}\right|^p \ge \varepsilon \frac{|t|^p + |s|^p}{2} \quad \Rightarrow \quad \frac{|t|^p + |s|^p}{2} \rho(\varepsilon) \le \frac{|t|^p + |s|^p}{2} - \left|\frac{t+s}{2}\right|^p.$$

Now given f, g with $\|f\|_p = \|g\|_p = 1$ and $\varepsilon > 0$ we need to find a $\delta > 0$ such that $\|\frac{f+g}{2}\|_p > 1 - \delta$ implies $\|f - g\|_p < 2\varepsilon$. Introduce

$$M := \left\{ x \in X \Big| \left|\frac{f(x) - g(x)}{2}\right|^p \ge \varepsilon \frac{|f(x)|^p + |g(x)|^p}{2} \right\}.$$

[6] James A. Clarkson (1906–1970), American mathematician

Then

$$\begin{aligned}\int_X \Big|\frac{f-g}{2}\Big|^p d\mu &= \int_{X\setminus M} \Big|\frac{f-g}{2}\Big|^p d\mu + \int_M \Big|\frac{f-g}{2}\Big|^p d\mu\\ &\le \varepsilon \int_{X\setminus M} \frac{|f|^p+|g|^p}{2} d\mu + \frac{1}{\rho}\int_M \left(\frac{|f|^p+|g|^p}{2} - \Big|\frac{f+g}{2}\Big|^p\right) d\mu\\ &\le \varepsilon + \frac{1-(1-\delta)^p}{\rho} < 2\varepsilon\end{aligned}$$

provided $\delta < (1-\varepsilon\rho)^{1/p} - 1$. □

Note that this also gives uniform convexity of vector-valued spaces since we can identify $L^p(X_1, d\mu_1) \oplus_p L^p(X_2, d\mu_2)$ with $L^p(X_1 \oplus X_2, d(\mu_1 \oplus \mu_2))$ if we identify (f_1, f_2) with f defined to be $f := f_1$ on X_1 and $f := f_2$ on X_2.

In particular, by the Milman–Pettis theorem (Theorem 6.21 from **[22]**), $L^p(X, d\mu)$ is reflexive for $1 < p < \infty$. We will give a direct proof for this fact in Corollary 6.2.

Problem* 3.7. *Show that a differentiable function $\varphi : (a,b) \to \mathbb{R}$ is (strictly) convex if and only if φ' is (strictly) increasing. Moreover, if φ is twice differentiable it is (strictly) convex if and only if $\varphi'' \ge 0$ ($\varphi'' > 0$ a.e.).*

Problem 3.8. *Prove*

$$\prod_{k=1}^n x_k^{\alpha_k} \le \sum_{k=1}^n \alpha_k x_k, \qquad \text{if} \quad \sum_{k=1}^n \alpha_k = 1,$$

for $\alpha_k > 0$, $x_k > 0$. (Hint: Take a sum of Dirac-measures and use that the exponential function is convex.)

Problem 3.9. *Show Minkowski's inequality directly from Hölder's inequality. Show that $L^p(X, d\mu)$ is strictly convex for $1 < p < \infty$ but not for $p = 1, \infty$ if X contains two disjoint subsets of positive finite measure. (Hint: Start from $|f+g|^p \le |f|\,|f+g|^{p-1} + |g|\,|f+g|^{p-1}$.)*

Problem* 3.10. *Show the* ***generalized Hölder's inequality****:*

$$\|f\,g\|_r \le \|f\|_p \|g\|_q, \qquad \frac{1}{p} + \frac{1}{q} = \frac{1}{r}. \tag{3.19}$$

Here we can allow $p, q, r \in (0, \infty]$ but of course $\|.\|_p$ will only be a norm for $p \ge 1$.

Problem* 3.11. *Show the iterated Hölder's inequality:*

$$\|f_1 \cdots f_m\|_r \le \prod_{j=1}^m \|f_j\|_{p_j}, \qquad \frac{1}{p_1} + \cdots + \frac{1}{p_m} = \frac{1}{r}. \tag{3.20}$$

Again with $p_j, r \in (0, \infty]$ as in the previous problem.

Problem* 3.12. *Suppose μ is finite. Show that $L^p \subseteq L^{p_0}$ and*

$$\|f\|_{p_0} \le \mu(X)^{\frac{1}{p_0}-\frac{1}{p}}\|f\|_p, \qquad 1 \le p_0 \le p.$$

(Hint: Generalized Hölder's inequality.)

Problem* 3.13. *Show that if $f \in L^{p_0} \cap L^{p_1}$ for some $p_0 < p_1$ then $f \in L^p$ for every $p \in [p_0, p_1]$ and we have the* ***Lyapunov inequality***[7]

$$\|f\|_p \le \|f\|_{p_0}^{1-\theta}\|f\|_{p_1}^{\theta},$$

where $\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$, $\theta \in (0,1)$. (Hint: Generalized Hölder inequality from Problem 3.10.)

Problem 3.14. *Let $1 < p < \infty$ and μ σ-finite. Let $f_n \in L^p(X, d\mu)$ be a sequence which converges pointwise a.e. to f such that $\|f_n\|_p \le C$. Then*

$$\int_X f_n g\, d\mu \to \int_X f g\, d\mu$$

for every $g \in L^q(X, d\mu)$. By Theorem 6.1 this implies that f_n converges weakly to f. (Hint: Recall that since f_n is bounded, it suffices to check convergence on a total subset of L^q (cf. Problem 4.48 from **[22]***) and use Theorem 1.25.)*

Problem 3.15. *Show that the Radon–Riesz theorem fails for $p = 1$: Find a sequence $f_n \in L^1(0,1)$ such that $f_n \ge 0$, $\int_0^1 f_n g\, dx \to \int_0^1 f g\, dx$ for every $g \in L^\infty(0,1)$ (i.e. $f_n \rightharpoonup f$ by Theorem 6.1), $\|f_n\|_1 \to \|f\|_1$ but $\|f_n - f\|_1 \not\to 0$. (Hint: Riemann–Lebesgue lemma.)*

Problem 3.16. *Given a function $f \in L^p(\mathbb{R}^n)$ define its* ***spherical average*** *as*

$$\tilde{f}(r) = S_n^{-1} \int_{S^{n-1}} f(r\omega) d\sigma^{n-1}(\omega).$$

Show that $\tilde{f}$ is well-defined (a.e.) and satisfies

$$\left(S_n \int_0^\infty |\tilde{f}(r)|^p r^{n-1} dr\right)^{1/p} \le \|f\|_p, \quad 1 \le p < \infty,$$

with equality for radial functions. In the case $p = \infty$ we have $\|\tilde{f}\|_\infty \le \|f\|_\infty$.

Problem 3.17 (Hardy inequality[8]). *Show that the integral operator*

$$(Kf)(x) := \frac{1}{x}\int_0^x f(y) dy$$

is bounded in $L^p(0,\infty)$ for $1 < p \le \infty$:

$$\|Kf\|_p \le \frac{p}{p-1}\|f\|_p.$$

---

[7] Aleksandr Lyapunov (1857–1918), Russian mathematician, mechanician and physicist

[8] G. H. Hardy (1877–1947), English mathematician

(Hint: Note $(Kf)(x) = \int_0^1 f(sx)ds$ and apply Minkowski's integral inequality.)

3.3. Nothing missing in L^p

Finally it remains to show that $L^p(X, d\mu)$ is complete.

Theorem 3.12 (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 $x \in X \setminus A$. The pointwise limit $f(x) = \lim_{n\to\infty} f_n(x)$, $x \in X \setminus A$, 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 3.13. *If $\|f_n - f\|_p \to 0$, $1 \le p \le \infty$, then there is a subsequence f_{n_j} (of representatives) which converges pointwise almost everywhere and a nonnegative function $G \in L^p(X, d\mu)$ such that $|f_{n_j}(x)| \le G(x)$ almost everywhere.*

Consequently, if $f_n \in L^{p_0} \cap L^{p_1}$ converges in both L^{p_0} and L^{p_1}, then the limits will be equal a.e. Be warned that this corollary is not true in general without passing to a subsequence (Problem 3.18).

It even turns out that L^p is separable.

Lemma 3.14. *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 base the set of characteristic functions $\chi_O(x)$ with O in this base and $\mu(O)$ finite 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 (Problem 3.23). 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 $O_n \supseteq A$ such that $\mu(O_n) \to \mu(A)$. Since $\mu(A) < \infty$, it is no restriction to assume $\mu(O_n) < \infty$, and thus $\|\chi_A - \chi_{O_n}\|_p^p = \mu(O_n \setminus A) = \mu(O_n) - \mu(A) \to 0$. 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 base 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. □

Finally, we give a characterization of relatively compact sets. Our proof is based on the following simple criterion (compare Lemma 1.11 from **[22]** for further background):

Lemma 3.15. *Let X be a Banach space and K some subset. Assume that for every $\varepsilon > 0$ there is a linear map P_ε onto a finite dimensional subspace Y_ε such that $\|P_\varepsilon\| \le C$, $P_\varepsilon K$ is bounded, and $\|(1 - P_\varepsilon)x\| \le \varepsilon$ for $x \in K$. Then K is relatively compact.*

Proof. We use the fact that in a metric space a set is relatively compact if and only if it is totally bounded (i.e. for every $\varepsilon > 0$ it can be covered by a finite number of balls of radius ε) together with the fact that in a finite dimensional Banach space the relatively compact sets are precisely the bounded ones (the Heine–Borel theorem).

Fix $\varepsilon > 0$. Then by total boundedness of $P_\varepsilon(K)$ we can find an ε-cover $\{B_\varepsilon(y_j)\}_{j=1}^m$ for $P_\varepsilon(K)$. Now if we choose $x_j \in P_\varepsilon^{-1}(\{y_j\}) \cap K$, then

$\{B_{3\varepsilon}(x_j)\}_{j=1}^n$ is a 3ε-cover for K. Indeed for $x \in K$ we have $P_\varepsilon x \in B_\varepsilon(y_j)$ for some j and hence $\|x-x_j\| \le \|(1-P_\varepsilon)x\|+\|P_\varepsilon x-y_j\|+\|(1-P_\varepsilon)x_j\| < 3\varepsilon$. □

To formulate our result let $X \subseteq \mathbb{R}^n$, $f \in L^p(X)$ and consider the **translation operator**

$$T_a(f)(x) := \begin{cases} f(x-a), & x-a \in X, \\ 0, & \text{else}, \end{cases} \tag{3.21}$$

for fixed $a \in \mathbb{R}^n$. Then one checks $\|T_a\| = 1$ (unless $|(X-a)\cap X| = 0$ in which case $T_a \equiv 0$) and $T_a f \to f$ as $a \to 0$ for $1 \le p < \infty$ (Problem 3.19).

Theorem 3.16 (Kolmogorov–Riesz–Sudakov[9]). *Let $X \subseteq \mathbb{R}^n$ be open. A subset F of $L^p(X)$, $1 \le p < \infty$, is relatively compact if and only if*

(i) *for every $\varepsilon > 0$ there is some $\delta > 0$ such that $\|T_a f - f\|_p \le \varepsilon$ for all $|a| \le \delta$ and $f \in F$.*

(ii) *for every $\varepsilon > 0$ there is some $r > 0$ such that $\|(1-\chi_{B_r(0)})f\|_p \le \varepsilon$ for all $f \in F$.*

Of course the last condition is void if X is bounded.

Proof. We first show that F is bounded. For this fix $\varepsilon = 1$ and choose δ, r according to (i), (ii), respectively. Then

$$\|f\chi_{B_r(x)}\|_p \le \|(T_y f - f)\chi_{B_r(x)}\|_p + \|T_y f \chi_{B_r(x+y)}\|_p \le 1 + \|f\chi_{B_r(x+y)}\|_p$$

for $f \in F$ and $|y| \le \delta$. Hence by induction $\|f\chi_{B_r(0)}\|_p \le m + \|f\chi_{B_r(my)}\|_p$ and choosing $|y| = \delta$ and $m \ge \frac{2r}{\delta}$ such that $B_r(my) \cap B_r(0) = \emptyset$ we obtain

$$\|f\|_p = \|f\chi_{B_r(0)}\|_p + \|f\chi_{\mathbb{R}^n\setminus B_r(0)}\|_p \le 2 + m.$$

To use our lemma we fix ε and choose a cube Q centered at 0 with side length δ according to (i) and finitely many disjoint cubes $\{Q_j\}_{j=1}^m$ of side length $\delta/2$ such that they cover $B_r(0)$ with r as in (ii). Now let Y be the finite dimensional subspace spanned by the characteristic functions of the cubes Q_j and let

$$P_m f := \sum_{j=1}^m \left(\frac{1}{|Q_j|} \int_{Q_j} f(y) d^n y \right) \chi_{Q_j}$$

be the projection from $L^p(X)$ onto Y. Note that using the triangle inequality and then Hölder's inequality shows

$$\|P_m f\|_p \le \sum_{j=1}^m \left(\frac{1}{|Q_j|} \int_{Q_j} |f(y)| d^n y \right) \|\chi_{Q_j}\|_p \le \sum_{j=1}^m \|f\|_{L^p(Q_j)} \le \|f\|_p$$

[9] Andrey Kolmogorov (1903–1987), Soviet mathematician
[9] Marcel Riesz (1886 –1969), Hungarian mathematician
[9] Vladimir Nikolaevich Sudakov (1934–2016), Soviet mathematician

and since F is bounded, so is $P_m F$. Moreover, for $f \in F$ we have

$$\|(1-P_m)f\|_p^p \le \varepsilon^p + \sum_{j=1}^m \int_{Q_j} |f(x) - P_m f(x)|^p d^n x$$

and using Jensen's inequality we further get

$$\begin{aligned}\|(1-P_m)f\|_p^p &\le \varepsilon^p + \sum_{j=1}^m \int_{Q_j} \frac{1}{|Q_j|}\int_{Q_j} |f(x)-f(y)|^p d^n y\, d^n x\\ &\le \varepsilon^p + \sum_{j=1}^m \int_{Q_j} \frac{2^n}{|Q|}\int_{Q} |f(x)-f(x-y)|^p d^n y\, d^n x\\ &\le \varepsilon^p + \frac{2^n}{|Q|}\int_Q \|f - T_y f\|_p^p d^n y \le (1+2^n)\varepsilon^p,\end{aligned}$$

since $x, y \in Q_j$ implies $x - y \in Q$.

Conversely, suppose F is relatively compact. To see (i) and (ii) pick an ε-cover $\{B_\varepsilon(f_j)\}_{j=1}^m$ and choose δ such that $\|f_j - T_a f_j\|_p \le \varepsilon$ for all $|a| \le \delta$ and $1 \le j \le m$. Then for every f there is some j such that $f \in B_\varepsilon(f_j)$ and hence $\|f - T_a f\|_p \le \|f - f_j\|_p + \|f_j - T_a f_j\|_p + \|T_a(f_j - f)\|_p \le 3\varepsilon$ implying (i). For (ii) choose r such that $\|(1-\chi_{B_r(0)})f_j\|_p \le \varepsilon$ for $1 \le j \le m$ implying $\|(1-\chi_{B_r(0)})f\|_p \le 3\varepsilon$ as before. □

Note that it suffices to require (i) on any given ball since the complement of the ball comes for free from (ii):

Corollary 3.17. *A subset $F \subset L^p(X)$ is relatively compact if and only if for every $\varepsilon > 0$ there is some $\delta > 0$ and some $r > 0$ such that*

(i') $\|(T_a - \mathbb{I})\chi_{B_r(0)} f\|_p \le \varepsilon$ *for all* $|a| \le \delta$ *and*

(ii) $\|(1-\chi_{B_r(0)})f\|_p \le \varepsilon$

for all $f \in F$.

Moreover, item (i') could be replaced by any condition ensuring compactness on finite balls (Problem 3.24).

Example 3.3. Choosing a fixed $f_0 \in L^p(X)$ condition (ii) is for example satisfied if $|f(x)| \le |f_0(x)|$ for all $f \in F$. Similarly, if $f_0(x) \ge 1$ with $\lim_{|x|\to\infty} f_0(x) = \infty$ such that $\|f_0 f\|_p \le C$ for all $f \in F$, then (ii) holds (to see this let ε be given and choose r such that $f_0(x) \ge \frac{C}{\varepsilon}$ for $|x| \ge r$). Condition (i) is for example satisfied if F is equicontinuous. ⋄

Problem* 3.18. *Find a sequence f_n which converges to 0 in $L^p(0,1)$, $1 \le p < \infty$, 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* 3.19. *Let $f \in L^p(X)$, $X \subseteq \mathbb{R}^n$ open, $1 \le p < \infty$ and show that $T_a f \to f$ in L^p as $a \to 0$. (Hint: Start with $f \in C_c(X)$ and use Theorem 3.18 below.)*

Problem 3.20. *Show that $L^p(X, d\mu) \cap L^q(X, d\mu)$ (with $1 \le p, q \le \infty$) together with the norm $\|f\|_{p,q} := \|f\|_p + \|f\|_q$ is a Banach space.*

Problem 3.21. *Consider $X := (0,1)$ with Lebesgue measure. Show that*

$$C := \{f \in L^p(X) | f(x) \ge 0 \text{ a.e.}\} \subset L^p(X)$$

is closed. Compute its interior. (Hint: For the second part distinguish $1 \le p < \infty$ and $p = \infty$.)

Problem 3.22. *Let X_1, X_2 be second countable topological spaces and let μ_1, μ_2 be outer regular σ-finite Borel measures. Let $\mathcal{B}_1$, $\mathcal{B}_2$ be bases for X_1, X_2, respectively. Show that the set of all functions $\chi_{O_1 \times O_2}(x_1, x_2) = \chi_{O_1}(x_1)\chi_{O_2}(x_2)$ for $O_1 \in \mathcal{B}_1$, $O_2 \in \mathcal{B}_2$ is total in $L^2(X_1 \times X_2, \mu_1 \otimes \mu_2)$. (Hint: Lemma 2.13 and 3.14.)*

Problem* 3.23. *Show that for any $f \in L^p(X, d\mu)$, $1 \le p \le \infty$ there exists a sequence of simple functions s_n such that $|s_n| \le |f|$ and $s_n \to f$ in $L^p(X, d\mu)$. If $p < \infty$ then s_n will be integrable. (Hint: Problem 2.3.)*

Problem* 3.24. *Consider $F \subset L^p(X)$, $1 \le p \le \infty$. Then F is relatively compact if $F|_{B_r(0)} \subset L^p(X \cap B_r(0))$ is relatively compact for every $r > 0$ and for every r there is some ε such that $\|(1 - \chi_{B_r(0)})f\|_p \le \varepsilon$ for all $f \in F$. The converse only holds if $p < \infty$.*

3.4. Approximation by nicer functions

Since measurable functions can be quite wild, they are sometimes hard to work with. In fact, in many situations some properties are much easier to prove for a dense set of *nice* functions and the general case can then be reduced to the nice case by an *approximation argument*. But for such a strategy to work one needs to identify suitable sets of nice functions which are dense in L^p.

Theorem 3.18. *Let X be a locally compact topological 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 proof of Lemma 3.14 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 $\mu(O\setminus K) \le \varepsilon$. By Urysohn's lemma (Lemma B.26 from [**22**]) there is a continuous function $f_\varepsilon : X \to [0,1]$ with compact support which is 1 on K and 0 outside O. Since

$$\int_X |\chi_K - f_\varepsilon|^p d\mu = \int_{O\setminus K} |f_\varepsilon|^p d\mu \le \mu(O\setminus K) \le \varepsilon,$$

we have $\|f_\varepsilon - \chi_K\|_p \to 0$ and we are done. □

In other words, the completion of $C_c(X)$ (or any larger set of p-integrable functions) with respect to $\|.\|_p$ gives $L^p(X, d\mu)$, $1 \le p < \infty$, up to isomorphism.

Clearly this result has to fail in the case $p = \infty$ (in general) since the uniform limit of continuous functions is again continuous. In fact, the closure of $C_c(\mathbb{R}^n)$ in the infinity norm is the space $C_0(\mathbb{R}^n)$ of continuous functions vanishing at ∞ (Problem 7.4). Another variant of this result is

Theorem 3.19 (Luzin[10])**.** *Let X be a locally compact metric space and let μ be a finite regular Borel measure. Let f be integrable. Then for every $\varepsilon > 0$ there is an open set O_ε with $\mu(O_\varepsilon) < \varepsilon$ and $X\setminus O_\varepsilon$ compact and a continuous function g which coincides with f on $X \setminus O_\varepsilon$.*

Proof. From the proof of the previous theorem we know that the set of all characteristic functions $\chi_K(x)$ with K compact is total. Hence we can restrict our attention to a sufficiently large compact set K. By Theorem 3.18 we can find a sequence of continuous functions f_n which converges to f in L^1. After passing to a subsequence we can assume that f_n converges a.e. and by Egorov's theorem there is a subset A_ε on which the convergence is uniform. By outer regularity we can replace A_ε by a slightly larger open set O_ε such that $C = K \setminus O_\varepsilon$ is compact. Now Tietze's extension theorem implies that $f|_C$ can be extended to a continuous function g on X. □

If X is some subset of $\mathbb{R}^n$, we can do even better and approximate integrable functions by smooth functions. The idea is to replace the value $f(x)$ by a suitable average computed from the values in a neighborhood. This is done by choosing a nonnegative bump function ϕ, whose area is normalized to 1, and considering the **convolution**

$$(\phi * f)(x) := \int_{\mathbb{R}^n} \phi(x-y) f(y) d^n y = \int_{\mathbb{R}^n} \phi(y) f(x-y) d^n y. \tag{3.22}$$

For example, if we choose $\phi_r = |B_r(0)|^{-1}\chi_{B_r(0)}$ to be the characteristic function of a ball centered at 0, then $(\phi_r * f)(x)$ will be precisely the average of the values of f in the ball $B_r(x)$. In the general case we can think of $(\phi * f)(x)$ as a weighted average. Moreover, if we choose ϕ differentiable, we

[10]Nikolai Luzin (1883–1950), Soviet mathematician

can interchange differentiation and integration to conclude that $\phi * f$ will also be differentiable. Iterating this argument shows that $\phi * f$ will have as many derivatives as ϕ. Finally, if the set over which the average is computed (i.e., the support of ϕ) shrinks, we expect $(\phi * f)(x)$ to get closer and closer to $f(x)$.

To make these ideas precise we begin with a few properties of the convolution.

Lemma 3.20. *The convolution has the following properties:*

(i) $f(x-.)g(.)$ *is integrable if and only if* $f(.)g(x-.)$ *is and*
$$(f*g)(x) = (g*f)(x) \tag{3.23}$$
in this case.

(ii) *Suppose* $\phi \in C_c^k(\mathbb{R}^n)$ *and* $f \in L^1_{loc}(\mathbb{R}^n)$, *then* $\phi * f \in C^k(\mathbb{R}^n)$ *and*
$$\partial_\alpha(\phi * f) = (\partial_\alpha \phi) * f \tag{3.24}$$
for any partial derivative of order at most k.

(iii) *We have* $\operatorname{supp}(f*g) \subseteq \overline{\operatorname{supp}(f)+\operatorname{supp}(g)}$. *In particular, if* $\phi \in C_c^k(\mathbb{R}^n)$ *and* $f \in L^1_c(\mathbb{R}^n)$, *then* $\phi * f \in C_c^k(\mathbb{R}^n)$.

(iv) *Suppose* $\phi \in L^1(\mathbb{R}^n)$ *and* $f \in L^p(\mathbb{R}^n)$, $1 \le p \le \infty$, *then their convolution is in* $L^p(\mathbb{R}^n)$ *and satisfies* ***Young's inequality***[11]
$$\|\phi * f\|_p \le \|\phi\|_1 \|f\|_p. \tag{3.25}$$

(v) *Suppose* $\phi \ge 0$ *with* $\|\phi\|_1 = 1$ *and* $f \in L^\infty(\mathbb{R}^n)$ *real-valued, then*
$$\inf_{x\in\mathbb{R}^n} f(x) \le (\phi * f)(x) \le \sup_{x\in\mathbb{R}^n} f(x). \tag{3.26}$$

Proof. (i) is a simple affine change of coordinates. (ii) follows by interchanging differentiation with the integral using Problems 2.12 and 2.13. (iii) If $x \notin \operatorname{supp}(f)+\operatorname{supp}(g)$, then $x-y \notin \operatorname{supp}(f)$ for $y \in \operatorname{supp}(g)$ and hence $f(x-y)g(y)$ vanishes on $\operatorname{supp}(g)$. This establishes the claim about the support and the rest follows from the previous item. (iv) The case $p=\infty$ follows from Hölder's inequality and we can assume $1 \le p < \infty$. Without loss of generality let $\|\phi\|_1 = 1$. Then
$$\begin{aligned}\|\phi * f\|_p^p &\le \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} |f(y-x)||\phi(y)| d^n y \right|^p d^n x \\ &\le \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |f(y-x)|^p |\phi(y)| d^n y\, d^n x = \|f\|_p^p,\end{aligned}$$
where we have used Jensen's inequality with $\varphi(x) = |x|^p$, $d\mu = |\phi| d^n y$ in the first and Fubini in the second step. (v) Immediate from integrating $\phi(y)\inf_{x\in\mathbb{R}^n} f(x) \le \phi(y) f(x-y) \le \phi(y)\sup_{x\in\mathbb{R}^n} f(x)$. □

[11] William Henry Young (1863–1942), English mathematician

Example 3.4. For $f := \chi_{[0,1]} - \chi_{[-1,0]}$ and $g := \chi_{[-2,2]}$ we have that $f * g$ is the difference of two triangles supported on $[1,3]$ and $[-3,-1]$ whereas $\operatorname{supp}(f) + \operatorname{supp}(g) = [-3,3]$, which shows that the inclusion in (iii) is strict in general. ⋄

Next we turn to approximation of f. To this end we call a family of integrable functions ϕ_ε, $\varepsilon \in (0,1]$, an **approximate identity** if it satisfies the following three requirements:

(i) $\|\phi_\varepsilon\|_1 \le C$ for all $\varepsilon > 0$.

(ii) $\int_{\mathbb{R}^n} \phi_\varepsilon(x) d^n x = 1$ for all $\varepsilon > 0$.

(iii) For every $r > 0$ we have $\lim_{\varepsilon\downarrow 0} \int_{|x|\ge r} \phi_\varepsilon(x) d^n x = 0$.

Moreover, a nonnegative function $\phi \in C_c^\infty(\mathbb{R}^n)$ satisfying $\|\phi\|_1 = 1$ is called a **mollifier**. Note that if the support of ϕ is within a ball of radius s, then the support of $\phi * f$ will be within $\{x \in \mathbb{R}^n | \operatorname{dist}(x, \operatorname{supp}(f)) \le s\}$.

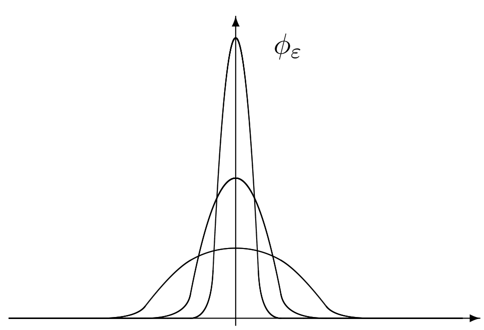
Example 3.5. The standard (also Friedrichs) mollifier is

$$\phi(x) := \begin{cases} \frac{1}{c}\exp(\frac{1}{|x|^2-1}), & |x| < 1, \\ 0, & |x| \ge 1, \end{cases} \qquad c := \int_{B_1(0)} \exp(\frac{1}{|x|^2-1}) d^n x.$$

Here the normalization constant c is chosen such that $\|\phi\|_1 = 1$. To show that this function is indeed smooth it suffices to show that all right derivatives of $f(r) = \exp(\frac{1}{r})$ at $r = 0$ vanish, which can be done using l'Hôpital's rule. ⋄

Example 3.6. If $\phi_1(x)$ is a mollifier on $\mathbb{R}$, then $\phi(x) = \prod_{j=1}^n \phi_1(x_j)$ is a mollifier on $\mathbb{R}^n$ which has a convenient product structure. ⋄

Example 3.7. Scaling a mollifier according to $\phi_\varepsilon(x) = \varepsilon^{-n}\phi(\frac{x}{\varepsilon})$ such that its mass is preserved ($\|\phi_\varepsilon\|_1 = 1$) and it concentrates more and more around the origin as $\varepsilon \downarrow 0$ we obtain an approximate identity:



In fact, (i), (ii) are obvious from $\|\phi_\varepsilon\|_1 = 1$ and the integral in (iii) will be identically zero for $\varepsilon \ge \frac{r}{s}$, where s is chosen such that $\operatorname{supp}\phi \subseteq \overline{B_s(0)}$. ⋄

Now we are ready to show that an approximate identity deserves its name.

Lemma 3.21. *Let ϕ_ε be an approximate identity. If $f \in L^p(\mathbb{R}^n)$ with $1 \leq p < \infty$, then*

$$\lim_{\varepsilon \downarrow 0} \phi_\varepsilon * f = f \tag{3.27}$$

with the limit taken in L^p. In the case $p = \infty$ the claim holds for $f \in C_0(\mathbb{R}^n)$.

Proof. We begin with the case where $f \in C_c(\mathbb{R}^n)$. Fix some small $\delta > 0$. Since f is uniformly continuous we know $|f(x-y) - f(x)| \to 0$ as $y \to x$ uniformly in x. Since the support of f is compact, this remains true when taking the L^p norm and thus we can find some r such that

$$\|f(.-y) - f(.)\|_p \leq \frac{\delta}{2C}, \qquad |y| \leq r.$$

(Here the C is the one for which $\|\phi_\varepsilon\|_1 \leq C$ holds.) Now we use

$$(\phi_\varepsilon * f)(x) - f(x) = \int_{\mathbb{R}^n} \phi_\varepsilon(y)(f(x-y) - f(x)) d^n y.$$

Splitting the domain of integration according to $\mathbb{R}^n = \{y \,|\, |y| \leq r\} \cup \{y \,|\, |y| > r\}$, we can estimate the L^p norms of the individual integrals using Minkowski's integral inequality as follows:

$$\left\| \int_{|y| \leq r} \phi_\varepsilon(y)(f(.-y) - f(.)) d^n y \right\|_p \leq \\ \int_{|y| \leq r} |\phi_\varepsilon(y)| \|f(.-y) - f(.)\|_p d^n y \leq \frac{\delta}{2}$$

and

$$\left\| \int_{|y| > r} \phi_\varepsilon(y)(f(.-y) - f(.)) d^n y \right\|_p \leq \\ 2\|f\|_p \int_{|y| > r} |\phi_\varepsilon(y)| d^n y \leq \frac{\delta}{2}$$

provided ε is so small such that the integral in (iii) is less than $\delta/(2\|f\|_p)$.

This establishes the claim for $f \in C_c(\mathbb{R}^n)$. Since these functions are dense in L^p for $1 \leq p < \infty$ and in $C_0(\mathbb{R}^n)$ for $p = \infty$ the claim follows from Lemma 4.34 from **[22]** and Young's inequality. □

Note that in case of a mollifier with support in $B_r(0)$ this result implies a corresponding local version since the value of $(\phi_\varepsilon * f)(x)$ is only affected by the values of f on $B_{\varepsilon r}(x)$. The question when the pointwise limit exists will be addressed in Problem 3.31.

Example 3.8. The **Fejér kernel**[12]

$$F_n(x) = \frac{1}{n}\left(\frac{\sin(nx/2)}{\sin(x/2)}\right)^2$$

is an approximate identity if we set it equal to 0 outside $[-\pi,\pi]$ (see the proof of Theorem 2.22 from [**22**]). Then taking a 2π periodic function and setting it equal to 0 outside $[-2\pi, 2\pi]$ Lemma 3.21 shows that for $f \in L^p(-\pi,\pi)$ the mean values of the partial sums of a Fourier series $\bar{S}_n(f)$ converge to f in $L^p(-\pi,\pi)$ for every $1 \le p < \infty$. For $p = \infty$ we recover Theorem 2.22 from [**22**]. Note also that this shows that the map $f \mapsto \hat{f}$ is injective on L^p.

Another classical example it the Poisson kernel, see Problem 3.30. Note that the **Dirichlet kernel**

$$D_n(x) = \frac{\sin((n+1/2)x)}{\sin(x/2)}$$

is no approximate identity since $\|D_n\|_1 \to \infty$ — see Example 4.7 from [**22**]. ◇

Now we are ready to prove

Theorem 3.22. *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 Theorem 3.18 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 ϕ and observe that $\phi_\varepsilon * f$ has compact support inside X for ε sufficiently small (since the distance from $\operatorname{supp}(f)$ to the boundary ∂X is positive by compactness). Moreover, $\phi_\varepsilon * f \to f$ uniformly by the previous lemma and hence also in $L^p(X, d\mu)$. □

Our final result is known as the **fundamental lemma of the calculus of variations**.

Lemma 3.23. *Suppose $X \subseteq \mathbb{R}^n$ is open and $f \in L^1_{loc}(X)$. (i) If f is real-valued then*

$$\int_X \varphi(x) f(x) d^n x \ge 0, \qquad \forall \varphi \in C_c^\infty(X),\ \varphi \ge 0, \tag{3.28}$$

if and only if $f(x) \ge 0$ (a.e.). (ii) Moreover,

$$\int_X \varphi(x) f(x) d^n x = 0, \qquad \forall \varphi \in C_c^\infty(X),\ \varphi \ge 0, \tag{3.29}$$

if and only if $f(x) = 0$ (a.e.).

[12] Lipót Fejér (1880–1959), Hungarian mathematician

Proof. (i) Choose a compact set $K \subset X$ and some $\varepsilon_0 > 0$ such that $K_{\varepsilon_0} := K + B_{\varepsilon_0}(0) \subseteq X$. Set $\tilde{f} := f\chi_{K_{\varepsilon_0}}$ and let ϕ be the standard mollifier. Then $(\phi_\varepsilon * \tilde{f})(x) = (\phi_\varepsilon * f)(x) \geq 0$ for $x \in K$, $\varepsilon < \varepsilon_0$ and since $\phi_\varepsilon * \tilde{f} \to \tilde{f}$ in $L^1(X)$ we have $(\phi_\varepsilon * \tilde{f})(x) \to f(x) \geq 0$ for a.e. $x \in K$ for an appropriate subsequence. Since $K \subset X$ is arbitrary the first claim follows. (ii) The first part shows that $\mathrm{Re}(f) \geq 0$ as well as $-\mathrm{Re}(f) \geq 0$ and hence $\mathrm{Re}(f) = 0$. Applying the same argument to $\mathrm{Im}(f)$ establishes the claim. □

The following variant is also often useful

Lemma 3.24 (du Bois-Reymond[13])**.** *Suppose $X \subseteq \mathbb{R}^n$ is open and connected. If $f \in L^1_{loc}(X)$ with*

$$\int_X f(x)\partial_j\varphi(x)d^n x = 0, \qquad \forall \varphi \in C_c^\infty(X),\ 1 \leq j \leq n, \tag{3.30}$$

then f is constant a.e. on X.

Proof. Choose a compact set $K \subset X$ and $\tilde{f}$, ϕ as in the proof of the previous lemma but additionally assume that K is connected. Then by Lemma 3.20 (ii)

$$\partial_j(\phi_\varepsilon * \tilde{f})(x) = ((\partial_j\phi_\varepsilon) * \tilde{f})(x) = ((\partial_j\phi_\varepsilon) * f)(x) = 0, \qquad x \in K,\ \varepsilon \leq \varepsilon_0.$$

Hence $(\phi_\varepsilon * \tilde{f})(x) = c_\varepsilon$ for $x \in K$ and as $\varepsilon \to 0$ there is a subsequence which converges a.e. on K. Clearly this limit function must also be constant: $(\phi_\varepsilon * \tilde{f})(x) = c_\varepsilon \to f(x) = c$ for a.e. $x \in K$. Now write X as a countable union of open balls whose closure is contained in X. If the corresponding constants for these balls were not all the same, we could find a partition into two union of open balls which were disjoint. This contradicts that X is connected. □

Of course the last result can be extended to higher derivatives. For the one-dimensional case this is outlined in Problem 3.36.

Problem* 3.25 (Smooth Urysohn lemma)**.** *Suppose K and C are disjoint closed subsets of $\mathbb{R}^n$ with K compact. Then there is a smooth function $f \in C_c^\infty(\mathbb{R}^n, [0,1])$ such that f is zero on C and one on K.*

Problem 3.26. *Let $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Show that $f * g \in C_0(\mathbb{R}^n)$ with*

$$\|f * g\|_\infty \leq \|f\|_p \|g\|_q.$$

Problem 3.27. *Show that the convolution on $L^1(\mathbb{R}^n)$ is associative. Conclude that $L^1(\mathbb{R}^n)$ together with convolution as a product is a commutative Banach algebra (without identity). (Hint: It suffices to verify associativity for nice functions.)*

[13] Paul du Bois-Reymond (1831–1889), German mathematician

Problem 3.28. *Let μ be a finite measure on $\mathbb{R}$. Then the set of all exponentials $\{e^{itx}\}_{t\in\mathbb{R}}$ is total in $L^p(\mathbb{R}, d\mu)$ for $1 \le p < \infty$.*

Problem 3.29. *Let ϕ be integrable and normalized such that $\int_{\mathbb{R}^n} \phi(x) d^n x = 1$. Show that $\phi_\varepsilon(x) = \varepsilon^{-n}\phi(\frac{x}{\varepsilon})$ is an approximate identity.*

Problem 3.30. *Show that the* ***Poisson kernel***[14]

$$P_\varepsilon(x) := \frac{1}{\pi}\frac{\varepsilon}{x^2+\varepsilon^2}$$

is an approximate identity on $\mathbb{R}$.

Show that the ***Cauchy transform*** *(also* ***Borel transform****)*

$$F(z) := \frac{1}{\pi}\int_{\mathbb{R}} \frac{f(\lambda)}{\lambda - z} d\lambda$$

of a real-valued function $f \in L^p(\mathbb{R})$ is analytic in the upper half-plane with imaginary part given by

$$\mathrm{Im}(F(x + \mathrm{i}y)) = (P_y * f)(x).$$

In particular, by Young's inequality $\|\mathrm{Im}(F(.+\mathrm{i}y))\|_p \le \|f\|_p$ and thus also $\sup_{y>0}\|\mathrm{Im}(F(.+\mathrm{i}y))\|_p = \|f\|_p$. Such harmonic functions are said to be in the ***Hardy space*** *$h^p(\mathbb{C}_+)$.*

(Hint: To see analyticity of F use Problem 2.18 plus the estimate

$$\left|\frac{1}{\lambda - z}\right| \le \frac{1}{1+|\lambda|}\frac{1+|z|}{|\mathrm{Im}(z)|}.)$$

Problem* 3.31. *Let ϕ be bounded with support in $\overline{B_1(0)}$ and normalized such that $\int_{\mathbb{R}^n} \phi(x) d^n x = 1$. Set $\phi_\varepsilon(x) = \varepsilon^{-n}\phi(\frac{x}{\varepsilon})$.*

For f locally integrable show

$$|(\phi_\varepsilon * f)(x) - f(x)| \le \frac{V_n\|\phi\|_\infty}{|B_\varepsilon(x)|}\int_{B_\varepsilon(x)} |f(y) - f(x)| d^n y.$$

Hence at every Lebesgue point (cf. Theorem 4.6) x we have

$$\lim_{\varepsilon\downarrow 0}(\phi_\varepsilon * f)(x) = f(x).$$

If f is uniformly continuous then the above limit will be uniform. See Problem 9.15 for the case when ϕ is not compactly supported.

Problem 3.32. *Let f, g be integrable (or nonnegative). Show that*

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

Problem 3.33. *Let f, g be integrable radial functions. Show that $f * g$ is again radial.*

---

[14] Siméon Denis Poisson (1781–1840), French mathematician, engineer, and physicist

Problem 3.34. *Let μ, ν be two complex measures on $\mathbb{R}^n$ and set $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, $(x, y) \mapsto x + y$. Define the convolution of μ and ν by*

$$\mu * \nu := S_\star(\mu \otimes \nu).$$

Show

- *$\mu * \nu$ is a complex measure given by*

$$(\mu * \nu)(A) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \chi_A(x + y) d\mu(x) d\nu(y) = \int_{\mathbb{R}^n} \mu(A - y) d\nu(y)$$

  and satisfying $|\mu*\nu|(\mathbb{R}^n) \le |\mu|(\mathbb{R}^n)|\nu|(\mathbb{R}^n)$ with equality for positive measures.
- *$\mu * \nu = \nu * \mu$.*
- *If $d\nu(x) = g(x)d^n x$ then $d(\mu * \nu)(x) = h(x)d^n x$ with $h(x) = \int_{\mathbb{R}^n} g(x - y)d\mu(y)$.*

In particular the last item shows that this definition agrees with our definition for functions if both measures have a density.

Problem 3.35. *Show that if $X = X_1 \times X_2 \subseteq \mathbb{R}^{n_1+n_2}$ then it suffices to restrict the requirements in Lemma 3.23 to test functions of the form $\varphi(x) = \varphi_1(x_1)\varphi_2(x_2)$, where $\varphi_j \in C_c^\infty(X_j)$.*

Problem 3.36 (du Bois-Reymond lemma). *Let f be a locally integrable function on the interval (a, b) and let $n \in \mathbb{N}_0$. If*

$$\int_a^b f(x)\varphi^{(n)}(x)dx = 0, \qquad \forall \varphi \in C_c^\infty(a, b),$$

then f is a polynomial of degree at most $n - 1$ a.e. (Hint: Begin by showing that there exists $\{\phi_{n,j}\}_{0 \le j \le n} \subset C_c^\infty(a, b)$ such that

$$\int_a^b x^k \phi_{n,j}(x)dx = \delta_{j,k}, \qquad 0 \le j, k \le n.$$

The case $n = 0$ is easy and for the general case note that one can choose $\phi_{n+1,n+1} = (n + 1)^{-1}\phi'_{n,n}$. Then, for given $\phi \in C_c^\infty(a, b)$, look at $\varphi(x) = \frac{1}{n!}\int_a^x (\phi(y) - \tilde{\phi}(y))(x - y)^n dy$ where $\tilde{\phi}$ is chosen such that this function is in $C_c^\infty(a, b)$.)

Problem 3.37. *Let $X \subseteq \mathbb{R}^n$ be open. Consider $f \in L^p(X)$ with $1 \le p \le \infty$ and let q be the corresponding dual index, $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\|f\|_p = \sup_{\varphi \in C_c^\infty(X), \|\varphi\|_q = 1} \left| \int_X f\varphi \, d^n x \right|.$$

3.5. Integral operators

Using Hölder's inequality, we can also identify a class of bounded operators from $L^p(Y, d\nu)$ to $L^p(X, d\mu)$. We will assume all measures to be σ-finite throughout this section.

Lemma 3.25 (Schur criterion[15]). *Let μ, ν be measures on X, Y, respectively, 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*

$$\|K_1(x, .)\|_{L^q(Y,d\nu)} \le C_1, \qquad \|K_2(., y)\|_{L^p(X,d\mu)} \le C_2 \tag{3.31}$$

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

$$(Kf)(x) := \int_Y K(x, y)f(y)d\nu(y), \tag{3.32}$$

for μ-almost every x is bounded with $\|K\| \le C_1C_2$.

Proof. 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

$$\begin{aligned}\int_Y |K(x, y)f(y)|d\nu(y) &\le \int_Y K_1(x, y)K_2(x, y)|f(y)|d\nu(y)\\ &\le C_1\|K_2(x, .)f(.)\|_{L^p(Y,d\nu)}\end{aligned}$$

for μ a.e. x (if $K_2(x, .)f(.) \notin L^p(Y, d\nu)$, the inequality is trivially true). If $p < \infty$ take this inequality to the p'th power and integrate with respect to x using Fubini

$$\begin{aligned}\int_X \left(\int_Y |K(x, y)f(y)|d\nu(y)\right)^p d\mu(x) &\le C_1^p \int_X \int_Y |K_2(x, y)f(y)|^p d\nu(y)d\mu(x)\\ = C_1^p \int_Y \int_X |K_2(x, y)f(y)|^p d\mu(x)d\nu(y) &\le C_1^p C_2^p \|f\|_p^p.\end{aligned}$$

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. If $p = \infty$ just take the supremum and note that integrability again follows from Fubini by restricting to subsets of finite μ-measure. □

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 $K_2(x, y) = |K(x, y)|^{1/p}$. For related results see also Problems 3.40 and 9.4.

[15] Issai Schur (1875–1941), Russian mathematician

Another case of special importance is the case of integral operators

$$(Kf)(x) := \int_X K(x,y)f(y)d\mu(y), \qquad f \in L^2(X,d\mu), \tag{3.33}$$

where $K(x,y) \in L^2(X \times X, d\mu \otimes d\mu)$. Such an operator is called a **Hilbert–Schmidt operator**.[16] Using Cauchy-Schwarz one sees that K is bounded,

$$\|Kf\|_2 \le \|K\|_2 \|f\|_2 \tag{3.34}$$

where

$$\|K\|_2^2 := \int_X \int_X |K(x,y)|^2 d\mu(x)d\mu(y) \tag{3.35}$$

is known as the **Hilbert–Schmidt norm** of K (in the case of matrices the name **Frobenius norm**[17] is more common). In particular, $\|K\| \le \|K\|_2$.

Lemma 3.26. *Let K be a Hilbert–Schmidt operator in $L^2(X,d\mu)$. Then*

$$\|K\|_2^2 = \sum_{j\in J} \|Ku_j\|^2 \tag{3.36}$$

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

$$\begin{aligned}
\sum_j \|Ku_j\|^2 &= \sum_j \int_X \left| \int_X K(x,y)u_j(y)d\mu(y) \right|^2 d\mu(x) \\
&= \int_X \sum_j \left| \int_X K(x,y)u_j(y)d\mu(y) \right|^2 d\mu(x) \\
&= \int_X \int_X |K(x,y)|^2 d\mu(x)d\mu(y)
\end{aligned}$$

as claimed. □

Note that the right-hand side of (3.36) is frequently used as an abstract way to define Hilbert–Schmidt operators in an arbitrary Hilbert space (cf. Lemma 3.24 from **[22]**). In particular, taking an ONB and restricting K to the subspace spanned by the first n basis vectors gives a sequence of finite

[16] Erhard Schmidt (1876–1959), Baltic German mathematician
[17] Ferdinand Georg Frobenius (1849 –1917), German mathematician

rank operators K_n which will converge to K with respect to the Hilbert–Schmidt norm thanks to (3.36). Hence Hilbert–Schmidt operators are compact and computing the Hilbert–Schmidt norm gives us an easy to check criterion for compactness of an integral operator.

Example 3.9. 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. Note that this result is frequently proven for continuous K using the Arzelà-Ascoli theorem (cf. Lemma 3.4 from [**22**]). Here we could even allow singularities as long as they are square integrable. $\diamond$

In combination with the spectral theorem for compact operators (compare in particular Corollary 3.8 from [**22**]) we obtain the classical Hilbert–Schmidt theorem:

Theorem 3.27 (Hilbert–Schmidt). *Let K be a self-adjoint Hilbert–Schmidt operator in $L^2(X, d\mu)$. Let $\{u_j\}$ be an orthonormal set of eigenfunctions with corresponding nonzero eigenvalues $\{\kappa_j\}$ from the spectral theorem for compact operators (Theorem 3.7 from* [**22**]*). Then*

$$K(x, y) = \sum_j \kappa_j u_j(x) u_j(y)^*, \tag{3.37}$$

where the sum converges in $L^2(X \times X, d\mu \otimes d\mu)$.

Proof. First of all we can extend $\{u_j\}$ to an orthonormal basis (setting the corresponding κ_j equal to 0). Now by Problem 3.41 $\{u_j(x)u_j(y)^*\}$ is an orthogonal basis for $L^2(X \times X, d\mu \otimes d\mu)$ and the expansion coefficients of $K(x, y)$ are given by $\int_X \int_X u_j(x)^* u_k(y) K(x, y) d\mu(y) d\mu(x) = \langle u_j, K u_k\rangle = \kappa_k \delta_{j,k}$. $\square$

In this context the above theorem is known as second Hilbert–Schmidt theorem and the spectral theorem for compact operators is known as first Hilbert–Schmidt theorem.

If an integral operator is positive we can say more. But first we will discuss two equivalent definitions of positivity in this context. First of all recall that an operator $K \in \mathscr{L}(L^2(X, d\mu))$ is positive if $\langle f, Kf\rangle \geq 0$ for all $f \in L^2(X, d\mu)$. Recall that positive operators are in particular self-adjoint. Secondly we call a kernel **symmetric** if it satisfies $K(x, y)^* = K(y, x)$ (cf. Problem 3.38) and a continuous kernel **positive semidefinite** on $U \subseteq X$ if

$$\sum_{j,k=1}^{n} \alpha_j^* \alpha_k K(x_j, x_k) \geq 0 \tag{3.38}$$

for all $(\alpha_1, \dots, \alpha_n) \in \mathbb{C}^n$ and $\{x_j\}_{j=1}^n \subseteq U$. In other words, for any $\{x_j\}_{j=1}^n \subseteq U$ the matrix $\{K(x_j, x_k)\}_{1\le j,k\le n}$ is positive semidefinite. In particular, a positive semidefinite kernel is symmetric since a positive semidefinite matrix is symmetric.

Both conditions have their advantages. For example, note that for a positive semidefinite kernel the case $n = 1$ shows that $K(x, x) \ge 0$ for $x \in U$ and the case $n = 2$ shows (look at the determinant) $|K(x, y)|^2 \le K(x, x)K(y, y)$ for $x, y \in U$. On the other hand, note that for a positive operator all eigenvalues are nonnegative.

It turns out that under mild assumptions both conditions are equivalent:

Lemma 3.28. *Suppose X is a locally compact metric space and μ a Borel measure. Let $K \in \mathscr{L}(L^2(X, d\mu))$ be an integral operator with a continuous kernel. Then, if K is positive, its kernel is positive semidefinite on* $\operatorname{supp}(\mu)$. *If μ is regular, the converse is also true.*

Proof. Let $x_0 \in \operatorname{supp}(\mu)$ and consider $\delta_{x_0,\varepsilon} = \mu(B_\varepsilon(x_0))^{-1}\chi_{B_\varepsilon(x_0)}$. Then for $f_\varepsilon = \sum_{j=1}^n \alpha_j \delta_{x_0,\varepsilon}$

$$0 \le \lim_{\varepsilon\downarrow 0}\langle f_\varepsilon, K f_\varepsilon\rangle = \lim_{\varepsilon\downarrow 0}\sum_{j,k=1}^n \alpha_j^*\alpha_k \int_{B_\varepsilon(x_j)}\int_{B_\varepsilon(x_k)} K(x,y)\frac{d\mu(x)}{\mu(B_\varepsilon(x_j))}\frac{d\mu(y)}{\mu(B_\varepsilon(x_k))}$$

$$= \sum_{j,k=1}^n \alpha_j^*\alpha_k K(x_j, x_k).$$

Conversely, let $f \in C_c(X)$ and let $S := \operatorname{supp}(f)\cap\operatorname{supp}(\mu)$. Then the function $f(x)^*K(x, y)f(y) \in C_c(X \times X)$ is uniformly continuous and for every $\varepsilon > 0$ we can partition the compact set S into a finite number of sets U_j which are contained in a ball $B_\delta(x_j)$ such that

$$|f(x)^*K(x,y)f(y) - \sum_{j,k}\chi_{U_j}(x)f(x_j)^*K(x_j,x_k)f(x_k)\chi_{U_k}(y)| \le \varepsilon, \quad x, y \in S.$$

Hence

$$\Big|\langle f, Kf\rangle - \sum_{j,k}\mu(U_j)f(x_j)^*K(x_j,x_k)f(x_k)\mu(U_k)\Big| \le \varepsilon\mu(S)^2$$

and since $\varepsilon > 0$ is arbitrary we have $\langle f, Kf\rangle \ge 0$. Since $C_c(X)$ is dense in $L^2(X, d\mu)$ if μ is regular by Theorem 3.18, we get this for all $f \in L^2(X, d\mu)$ by taking limits. □

Now we are ready for the following classical result:

Theorem 3.29 (Mercer[18]). *Let K be a positive integral operator with a continuous kernel on $L^2(X, d\mu)$ with X a locally compact metric space. Let*

[18]James Mercer (1883–1932), English mathematician

μ a Borel measure such that the diagonal $K(x,x)$ is integrable. Then K is trace class, all eigenfunctions u_j corresponding to positive eigenvalues κ_j are continuous, and (3.37) *converges uniformly on compact subsets of the support of μ. Moreover,*

$$\operatorname{tr}(K) = \sum_j \kappa_j = \int_X K(x,x) d\mu(x). \tag{3.39}$$

Proof. Define $k(x)^2 := K(x,x)$ such that $|K(x,y)| \le k(x)k(y)$ for $x,y \in \operatorname{supp}(\mu)$. Since by assumption $k \in L^2(X,d\mu)$ we see that K is Hilbert–Schmidt and we have the representation (3.37) with $\kappa_j > 0$. Moreover, dominated convergence shows that $u_j(x) = \kappa_j^{-1} \int_X K(x,y) u_j(y) d\mu(y)$ is continuous.

Now note that the operator corresponding to the kernel

$$K_n(x,y) = K(x,y) - \sum_{j\le n} \kappa_j u_j(x)^* u_j(y) = \sum_{j>n} \kappa_j u_j(x)^* u_j(y)$$

is also positive (its nonzero eigenvalues are κ_j, $j > n$). Hence $K_n(x,x) \ge 0$ implying

$$\sum_{j\le n} \kappa_j |u_j(x)|^2 \le k(x)^2$$

for $x \in \operatorname{supp}(\mu)$ and hence (3.37) converges absolutely for $x,y \in \operatorname{supp}(\mu)$. Denote the limit by $\tilde{K}(x,y)$ and observe

$$\begin{aligned} \left| \sum_{m<j<n} \kappa_j u_j(x)^* u_j(y) \right| &\le \left(\sum_{m<j<n} \kappa_j |u_j(x)|^2 \right)^{1/2} \left(\sum_{m<j<n} \kappa_j |u_j(y)|^2 \right)^{1/2} \\ &\le \left(\sum_{m<j<n} \kappa_j |u_j(x)|^2 \right)^{1/2} k(y) \end{aligned}$$

shows that the series converges uniformly for fixed x with respect to y in compact sets. Hence $\tilde{K}(x,.)$ is continuous for fixed x and satisfies $|\tilde{K}(x,y)| \le k(x)k(y)$. Moreover, for fixed x we have $K(x,.), \tilde{K}(x,.) \in L^2(X,d\mu)$ and

$$\int_X K(x,y) u(y) d\mu(y) = \kappa u(x) = \int_X \tilde{K}(x,y) u(y) d\mu(y)$$

for either $u \in \operatorname{Ker}(K)$ with $\kappa = 0$ or $u = u_j$ with $\kappa = \kappa_j$ (use dominated convergence to interchange the sum and the integral). Since these functions are total we get $K(x,y) = \tilde{K}(x,y)$ for fixed x and a.e. y. But since both functions are continuous for fixed x, we get equality for all y. In particular, we have $\sum_{j\le n} \kappa_j |u_j(x)|^2 \nearrow k(x)^2$ and the convergence is uniform on compact subsets of $\operatorname{supp}(\mu)$ by Dini's theorem (Problem B.64 from **[22]**). By our estimate this also shows uniform convergence of (3.37) on compact sets.

Finally, integrating (3.37) for $x = y$ using $\|u_j\| = 1$ shows the last claim. □

Example 3.10. Let k be a periodic function which is square integrable over $[-\pi, \pi]$. Then the integral operator

$$(Kf)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} k(y-x) f(y) dy$$

has the eigenfunctions $u_j(x) = (2\pi)^{-1/2} \mathrm{e}^{-\mathrm{i}jx}$ with corresponding eigenvalues $\hat{k}_j$, $j \in \mathbb{Z}$, where $\hat{k}_j$ are the Fourier coefficients of k. Since $\{u_j\}_{j\in\mathbb{Z}}$ is an ONB we have found all eigenvalues. Moreover, in this case (3.37) is just the Fourier series

$$k(y-x) = \sum_{j\in\mathbb{Z}} \hat{k}_j \mathrm{e}^{\mathrm{i}j(y-x)}.$$

Choosing a continuous function for which the Fourier series does not converge absolutely shows that the positivity assumption in Mercer's theorem is crucial. ⋄

Of course given a kernel this raises the question if it is positive semi-definite. This can often be done by reverse engineering basend on constructions which produce new positive semi-definite kernels out of old ones.

Lemma 3.30. *Let $K_j(x,y)$ be positive semi-definite kernels on X. Then the following kernels are also positive semi-definite.*

(i) $\alpha_1 K_1 + \alpha_2 K_2$ *provided* $\alpha_j \ge 0$.
(ii) $\lim_j K_j(x,y)$ *provided the limit exists pointwise.*
(iii) $K_1(\phi(x), \phi(y))$ *for* $\phi : X \to X$.
(iv) $f(x)^* K_1(x,y) f(y)$ *for* $f : X \to \mathbb{C}$.
(v) $K_1(x,y) K_2(x,y)$.

Proof. Only the last item is not straightforward. However, it boils down to the Schur product theorem from linear algebra given below. □

Theorem 3.31 (Schur). *Let A and B be two $n \times n$ matrices and let $A \circ B$ be their **Hadamard product**[19] defined by multiplying the entries pointwise: $(A \circ B)_{jk} := A_{jk} B_{jk}$. If A and B are positive (semi-) definite, then so is $A \circ B$.*

Proof. By the spectral theorem we can write $A = \sum_{j=1}^n \alpha_j \langle u_j, .\rangle u_j$, where $\alpha_j > 0$ ($\alpha_j \ge 0$) are the eigenvalues and u_j are a corresponding ONB of eigenvectors. Similarly $B = \sum_{j=1}^n \beta_j \langle v_j, .\rangle v_j$. Then $A \circ B = \sum_{j,k=1}^n \alpha_j \beta_k (\langle u_j, .\rangle u_j) \circ (\langle v_k, .\rangle v_k) = \sum_{j,k=1}^n \alpha_j \beta_k \langle u_j \circ v_k, .\rangle u_j \circ v_k$, which proves the claim. □

[19] Jacques Hadamard (1865–1963), French mathematician

Example 3.11. Let $X = \mathbb{R}^n$. The most basic kernel is $K(x,y) = x^* \cdot y$ which is positive semi-definite since $\sum_{j,k} \alpha_j^* \alpha_k K(x_j, x_k) = \left| \sum_j \alpha_j x_j \right|^2$. Slightly more general is $K(x,y) = \langle x, Ay \rangle$, where A is a positive semi-definite matrix. Indeed writing $A = B^2$ this follows from the previous observation using item (iii) with $\phi = B$. Another famous kernel is the Gaussian kernel

$$K(x,y) = \mathrm{e}^{-|x-y|^2/\sigma}, \qquad \sigma > 0.$$

To see that it is positive semi-definite start by observing that $K_1(x,y) = \exp(2x^* \cdot y/\sigma)$ is by items (i) and (ii) since the Taylor series of the exponential function has positive coefficients. Finally observe that our kernel is of the form (vi) with $f(x) = \exp(-|x|^2/\sigma)$. $\diamond$

Example 3.12. A **reproducing kernel Hilbert space** is a Hilbert space $\mathfrak{H}$ of complex-valued functions $X \to \mathbb{C}$ such that point evaluations are continuous linear functionals. In this case the Riesz lemma implies that for every $x \in X$ there is a corresponding function $K_x \in \mathfrak{H}$ such that $f(x) = \langle K_x, f \rangle$. Applying this to $f = K_y$ we get $K_y(x) = \langle K_x, K_y \rangle$ which suggests the more symmetric notation

$$K(x,y) := \langle K_x, K_y \rangle.$$

The kernel K is called the **reproducing kernel** for $\mathfrak{H}$. A short calculation

$$\sum_{j,k} \alpha_j^* \alpha_k K(x_j, x_k) = \sum_{j,k} \alpha_j^* \alpha_k \langle K_{x_j}, K_{x_k} \rangle = \Big\| \sum_j \alpha_j K_{x_j} \Big\|^2$$

verifies that it is a positive semi-definite kernel. An explicit example is given by the quadratic form domains of positive Sturm–Liouville operators; Problem 3.19 from **[22]**. $\diamond$

Problem* 3.38. *Suppose K is a Hilbert–Schmidt operator in $L^2(X, d\mu)$ with kernel $K(x,y)$. Show that the adjoint operator is given by*

$$(K^* f)(x) = \int_X K(y,x)^* f(y) d\mu(y), \qquad f \in L^2(X, d\mu).$$

In particular, K is self-adjoint if and only if its kernel is symmetric.

Problem 3.39. *Obtain Young's inequality* (3.25) *from Schur's criterion.*

Problem 3.40 (Schur test)**.** *Let $K(x,y)$ be given and suppose there are positive measurable functions $a(x)$ and $b(y)$ such that*

$$\|K(x,.)b(.)\|_{L^1(Y,d\nu)} \le C_1 a(x), \qquad \|a(.)K(.,y)\|_{L^1(X,d\mu)} \le C_2 b(y).$$

Then the operator $K : L^2(Y, d\nu) \to L^2(X, d\mu)$, defined by

$$(Kf)(x) := \int_Y K(x,y) f(y) d\nu(y),$$

for μ-almost every x is bounded with $\|K\| \le \sqrt{C_1 C_2}$. Show that the adjoint operator is given by

$$(K^* f)(x) = \int_X K(y,x)^* f(y) d\mu(y), \qquad f \in L^2(X, d\mu).$$

(Hint: Estimate $|(Kf)(x)|^2 \le |\int_Y |K(x,y)|^{1/2} b(y) |K(x,y)|^{1/2} \frac{|f(y)|}{b(y)} d\nu(y)|^2$ using Cauchy–Schwarz and integrate the result with respect to x.)

Problem* 3.41. *Let $(X, d\mu)$ and $(Y, d\nu)$ be two measure spaces and $\{u_j(x)\}_{j\in J}$, $\{v_k(y)\}_{k\in K}$ be orthonormal bases for $L^2(X, d\mu)$, $L^2(Y, d\nu)$, respectively. Then $\{u_j(x)v_k(y)\}_{(j,k)\in J\times K}$ is an orthonormal base for $L^2(X\times Y, d\mu \otimes d\nu)$.*

Problem 3.42. *Let K be an self-adjoint integral operator with continuous kernel satisfying the estimate $|K(x,y)| \le k(x)k(y)$ with $k \in L^2(X, d\mu) \cap L^\infty(X, d\mu)$. Show that the conclusion from Mercer's theorem still hold if K has only a finite number of negative eigenvalues.*

Problem 3.43. *Show that the Fourier transform of a finite (positive) measure*

$$\hat{\mu}(p) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}px} d\mu(x)$$

gives rise to a positive semi-definite kernel $K(x,y) = \hat{\mu}(x-y)$. (The converse is Bochner's theorem — see Theorem 8.26)

3.6. Rearrangements

Many situations lead to the problem of finding a function which maximizes a certain integral. Moreover, typically the maximizing function will preserve the symmetry of the original problem. In case of radially symmetric problems this suggests to *rearrange* a function into a radially symmetric one thereby increasing the value of the integral under consideration.

We start by defining the **symmetric rearrangement** $A^\star$ of a finite Borel set $A \subset \mathbb{R}^n$ to be the open ball $B_r(0)$ with the same volume, that is, with r chosen such that $|A| = V_n r^n$. In other words,

$$A^\star := \{x \in \mathbb{R}^n | V_n |x|^n < |A|\}. \tag{3.40}$$

Then the symmetric rearrangement of a nonnegative Borel function $f : \mathbb{R}^n \to [0,\infty)$ is defined by rearranging its strict super-level sets

$$S_f(t) := f^{-1}((t,\infty)) = \{x \in \mathbb{R}^n | f(x) > t\}. \tag{3.41}$$

For this to work we assume that f vanishes at ∞ in the sense that $|S_f(t)| < \infty$ for all $t \ge 0$. More precisely, we define $f^\star$ such that

$$S_{f^\star}(t) = S_f(t)^\star \tag{3.42}$$

for all $t \ge 0$. This process is illustrated Figure 3.1. To this end note that

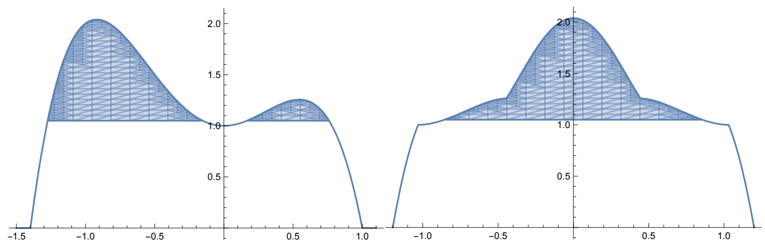


Figure 3.1. A function is rearranged into a decreasing symmetric function such that the areas of the super-level sets agree.

$x \in S_f(t)$ is equivalent to $f(x) > t$ (i.e., $\{t \geq 0 | x \in S_f(t)\} = [0, f(x))$ and hence we can reconstruct f from its super-level sets using $f(x) = \sup\{t \geq 0 | x \in S_f(t)\}$. This motivates to explicitly define

$$f^\star(x) := \sup\{t \geq 0 | x \in S_f(t)^\star\}. \tag{3.43}$$

Note that monotonicity of the super-level sets $S_f(t_2) \subseteq S_f(t_1)$ (and hence also $S_f(t_2)^\star \subseteq S_f(t_1)^\star$) for $t_1 \leq t_2$ implies that $f^\star$ is radial and nonincreasing. Moreover, we have $0 \in S_f(t)$ if and only if $S_f(t)$ is nonempty and hence

$$f^\star(0) = \|f\|_\infty. \tag{3.44}$$

In particular, $f^\star(x)$ is finite. Also by monotonicity of S_f it follows that $t \in I(x) := \{t \geq 0 | x \in S_f(t)^\star\}$ implies $s \in I(x)$ for every $s \leq t$ and hence $I(x)$ is an interval. Moreover, since the set $\{(x,t) | x \in S_f(t)^*\}$ is open, so is the projection for fixed x and thus $I(x) = [0, f^\star(x))$. But this implies $x \in S_f(t)^\star$ if and only if $f^\star(x) > t$ and hence we indeed have (3.42). In particular, the super-level sets of $f^\star$ are open, that is, $f^\star$ is lower semicontinuous and hence measurable. For a complex-valued function f we set $f^\star := |f|^\star$.

As a trivial consequence of (3.42) we see that the distribution functions of f and $f^\star$ agree, $E_f(t) := |S_{|f|}(t)| = |S_{|f|}(t)^\star| =: E_{f^\star}(t)$. This is sometimes known as **equimeasurability** since (cf. Problem 2.20) it implies

$$\int_{\mathbb{R}^n} \Phi(|f|) d^n x = \int_0^\infty \phi(r) E_f(r) dr = \int_0^\infty \phi(r) E_{f^\star}(r) dr = \int_{\mathbb{R}^n} \Phi(f^\star) d^n x \tag{3.45}$$

for an arbitrary absolutely continuous function Φ with $\Phi(0) = 0$, whose derivative is $\Phi' = \phi$. In particular, the choice $\Phi(t) := t^p$ shows

$$\|f\|_p = \|f^\star\|_p. \tag{3.46}$$

The observation $\{t \geq 0 | x \in S_f(t)\} = [0, f(x))$ can also be phrased as

$$f(x) = \int_0^\infty \chi_{S_f(t)}(x) dt \tag{3.47}$$

and this is known as **layer cake representation**. The name alludes to the fact that $f(x)$ is obtained by summing the contributions from all layers (level sets) corresponding to values t below $f(x)$. It is particularly useful in proofs since it frequently allows a reduction to the case of characteristic functions. This is illustrated in the following result:

Theorem 3.32 (Hardy–Littlewood)**.** *Let f and g be nonnegative Borel functions on $\mathbb{R}^n$, vanishing at ∞. Then*

$$\int_{\mathbb{R}^n} f(x)g(x)d^n x \le \int_{\mathbb{R}^n} f^\star(x)g^\star(x)d^n x. \tag{3.48}$$

Proof. We start with the case where $f = \chi_A$ and $g = \chi_B$ are characteristic functions in which case the claim follows from $|A^\star \cap B^\star| = \min(|A|, |B|) \ge |A \cap B|$. For the general case the layer cake representation and Fubini give

$$\begin{aligned}\int_{\mathbb{R}^n} f(x)g(x)d^n x &= \int_0^\infty \int_0^\infty \int_{\mathbb{R}^n} \chi_{S_{f(t)}}(x)\chi_{S_{g(s)}}(x)d^n x\, ds\, dt\\ &\le \int_0^\infty \int_0^\infty \int_{\mathbb{R}^n} \chi_{S_{f^\star(t)}}(x)\chi_{S_{g^\star(s)}}(x)d^n x\, ds\, dt\\ &= \int_{\mathbb{R}^n} f^\star(x)g^\star(x)d^n x\end{aligned}$$

as desired. □

Problem 3.44. *Suppose $f : \mathbb{R} \to [0, \infty)$ vanishes at ∞ and is nondecreasing on $(-\infty, 0]$ and nonincreasing on $[0, \infty)$. Compute $f^\star$.*

Problem 3.45. *Show that the rearrangement is order preserving, that is, if $f(x) \le g(x)$ for all x, then we also have $f^\star(x) \le g^\star(x)$ for all x.*

Problem 3.46. *Suppose $\phi : [0, \infty) \to [0, \infty)$ is nondecreasing, lower semicontinuous, and satisfies $\phi(0) = 0$ as well as continuity at 0. Show that $(\phi \circ f)^\star = \phi \circ f^\star$.*

Chapter 4

More measure theory

4.1. 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 $\nu(X \setminus N) = 0$.

Example 4.1. 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 $\Theta(\mathbb{R} \setminus \{0\}) = 0$. ⋄

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 4.2. The prototypical example is the measure $d\nu := f\, d\mu$ (compare Lemma 2.3). Indeed by Lemma 2.6 $\mu(A) = 0$ implies

$$\nu(A) = \int_A f\, d\mu = 0 \tag{4.1}$$

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 4.1. *Let μ, ν be σ-finite measures. Then there exists a nonnegative function f and a set N of μ measure zero, such that*

$$\nu(A) = \nu(A \cap N) + \int_A f\, d\mu. \tag{4.2}$$

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:

$$\begin{aligned} |\ell(h)|^2 &= \left|\int 1 \cdot h\, d\nu\right|^2 \le \left(\int |1|^2\, d\nu\right)\left(\int |h|^2 d\nu\right) \\ &\le \nu(X)\left(\int |h|^2 d\alpha\right) = \nu(X)\|h\|^2. \end{aligned}$$

Hence by the Riesz lemma (Theorem 2.10 from [**22**]) there exists a $g \in L^2(X, d\alpha)$ such that

$$\ell(h) = \int_X hg\, d\alpha.$$

By construction

$$\nu(A) = \int \chi_A\, d\nu = \int \chi_A g\, d\alpha = \int_A g\, d\alpha. \tag{4.3}$$

In particular, g must be positive a.e. (take A the set where g is negative). Moreover,

$$\mu(A) = \alpha(A) - \nu(A) = \int_A (1-g) d\alpha$$

which shows that $g \le 1$ a.e. Now choose $N := \{x | g(x) = 1\}$ such that $\mu(N) = 0$ and set

$$f := \frac{g}{1-g}\chi_{N'}, \qquad N' = X \setminus N.$$

Then, since (4.3) 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 $\alpha(X_n) < \infty$. Now we set $\tilde{X}_n := X_n \setminus X_{n-1}$ (where $X_0 = \emptyset$) and consider $\mu_n(A) := \mu(A \cap \tilde{X}_n)$ and $\nu_n(A) := \nu(A \cap \tilde{X}_n)$. Then there exist corresponding sets N_n and functions f_n such that

$$\nu_n(A) = \nu_n(A \cap N_n) + \int_A f_n d\mu_n = \nu(A \cap N_n) + \int_A f_n d\mu,$$

where for the last equality we have assumed $N_n \subseteq \tilde{X}_n$ and $f_n(x) = 0$ for $x \in \tilde{X}_n'$ without loss of generality. Now set $N := \bigcup_n N_n$ as well as $f := \sum_n f_n$, then $\mu(N) = 0$ and

$$\nu(A) = \sum_n \nu_n(A) = \sum_n \nu(A \cap N_n) + \sum_n \int_A f_n d\mu = \nu(A \cap N) + \int_A f d\mu,$$

which finishes the proof. □

Note that another set $\tilde{N}$ will give the same decomposition as long as $\mu(\tilde{N}) = 0$ and $\nu(\tilde{N}' \cap N) = 0$ since in this case $\nu(A) = \nu(A \cap \tilde{N}) + \nu(A \cap \tilde{N}') = \nu(A \cap \tilde{N}) + \nu(A \cap \tilde{N}' \cap N) + \int_{A \cap \tilde{N}'} f d\mu = \nu(A \cap \tilde{N}) + \int_A f d\mu$. Hence we can increase N by sets of μ measure zero and decrease N by sets of ν measure zero.

Now the anticipated results follow with no effort:

Theorem 4.2 (Radon–Nikodym[1])**.** *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 nonnegative measurable function f such that*

$$\nu(A) = \int_A f\, d\mu \tag{4.4}$$

*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. □

Example 4.3. Take $X := \mathbb{R}$. Let μ be the counting measure and ν Lebesgue measure. Then $\nu \ll \mu$ but there is no f with $d\nu = f\, d\mu$. If there were such an f, there must be a point $x_0 \in \mathbb{R}$ with $f(x_0) > 0$ and we have $0 = \nu(\{x_0\}) = \int_{\{x_0\}} f\, d\mu = f(x_0) > 0$, a contradiction. Hence the Radon–Nikodym theorem can fail if μ is not σ-finite. ⋄

Theorem 4.3 (Lebesgue decomposition)**.** *Let μ, ν be two σ-finite measures on a measurable space (X, Σ). Then ν can be uniquely decomposed as $\nu = \nu_{sing} + \nu_{ac}$, where μ 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, since A is arbitrary, $\tilde{f} = f$ 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 $\tilde{\nu}_{sing} = \nu - \tilde{\nu}_{ac} = \nu - \nu_{ac} = \nu_{sing}$. □

[1] Otto M. Nikodym (1887–1974), Polish mathematician

Problem* 4.1. *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 with respect to Lebesgue measure equals the ordinary derivative $\mu'(x)$.*

Problem 4.2. *Suppose ν is an inner regular measures. Show that $\nu \ll \mu$ if and only if $\mu(K) = 0$ implies $\nu(K) = 0$ for every compact set.*

Problem 4.3. *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 4.4. *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 4.5 (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 4.6. *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 $\mu \ll \nu$.

4.2. 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 4.1). Our aim in this section is to generalize this result to arbitrary Borel measures on $\mathfrak{B}^n$.

Let μ be a Borel measure on $\mathbb{R}^n$. We call

$$(D\mu)(x) := \lim_{\varepsilon \downarrow 0} \frac{\mu(B_\varepsilon(x))}{|B_\varepsilon(x)|} \tag{4.5}$$

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 $A \in \mathfrak{B}^n$.)

Example 4.4. Consider a Borel measure on $\mathfrak{B}$ and suppose its distribution $\mu(x)$ (as defined in (1.17)) is differentiable at x. Then

$$(D\mu)(x) = \lim_{\varepsilon \downarrow 0} \frac{\mu((x-\varepsilon, x+\varepsilon))}{2\varepsilon} = \lim_{\varepsilon \downarrow 0} \frac{\mu(x+\varepsilon) - \mu(x-\varepsilon)}{2\varepsilon} = \mu'(x).$$

$\diamond$

To compute the derivative of μ, we introduce the **upper** and **lower derivative**,

$$(\overline{D}\mu)(x) := \limsup_{\varepsilon\downarrow 0} \frac{\mu(B_\varepsilon(x))}{|B_\varepsilon(x)|} \quad \text{and} \quad (\underline{D}\mu)(x) := \liminf_{\varepsilon\downarrow 0} \frac{\mu(B_\varepsilon(x))}{|B_\varepsilon(x)|}. \tag{4.6}$$

Clearly μ is differentiable at x if $(\underline{D}\mu)(x) = (\overline{D}\mu)(x) < \infty$. Next note that they are measurable: In fact, this follows from

$$(\overline{D}\mu)(x) = \lim_{n\to\infty} \sup_{0<\varepsilon<1/n} \frac{\mu(B_\varepsilon(x))}{|B_\varepsilon(x)|} \tag{4.7}$$

since the supremum on the right-hand side is lower semicontinuous with respect to x (cf. Problem 1.23) as $x \mapsto \mu(B_\varepsilon(x))$ is lower semicontinuous (Problem 4.7). Similarly for $(\underline{D}\mu)(x)$.

Next, the following geometric fact of $\mathbb{R}^n$ will be needed.

Lemma 4.4 (Wiener[2] covering lemma)**.** *Given open balls $B_1 := B_{r_1}(x_1)$, ..., $B_m := B_{r_m}(x_m)$ in $\mathbb{R}^n$, there is a subset of disjoint balls B_{j_1}, ..., B_{j_k} such that*

$$\bigcup_{j=1}^m B_j \subseteq \bigcup_{\ell=1}^k B_{3r_{j_\ell}}(x_{j_\ell}) \tag{4.8}$$

Proof. Assume that the balls B_j are ordered by decreasing radius. Start with $B_{j_1} = B_1$ and remove all balls from our list which intersect B_{j_1}. Observe that the removed balls are all contained in $B_{3r_1}(x_1)$. Proceeding like this, we obtain the required subset. □

The upshot of this lemma is that we can select a disjoint subset of balls which still controls the Lebesgue volume of the original set up to a universal constant 3^n (recall $|B_{3r}(x)| = 3^n|B_r(x)|$).

Now we can show

Lemma 4.5. *Let μ be a Borel measure on $\mathbb{R}^n$ and $\alpha > 0$. For every Borel set A we have*

$$|\{x \in A \,|\, (\overline{D}\mu)(x) > \alpha\}| \le 3^n \frac{\mu(A)}{\alpha} \tag{4.9}$$

and

$$|\{x \in A \,|\, (\overline{D}\mu)(x) > 0\}| = 0, \ \textit{whenever } \mu(A) = 0. \tag{4.10}$$

Proof. Let $A_\alpha := \{x \in A | (\overline{D}\mu)(x) > \alpha\}$. We will show

$$|K| \le 3^n \frac{\mu(O)}{\alpha}$$

[2]Norbert Wiener (1894–1964), American mathematician

for every open set O with $A \subseteq O$ and every compact set $K \subseteq A_\alpha$. The first claim then follows from outer regularity of μ and inner regularity of 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 from these balls. Moreover, by Lemma 4.4 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 $A_0 = \cup_{j=1}^\infty A_{1/j}$ and by the first part $|A_{1/j}| = 0$ for every j if $\mu(A) = 0$. □

Theorem 4.6 (Lebesgue differentiation theorem)**.** *Let f be (locally) integrable, then for a.e. $x \in \mathbb{R}^n$ we have*

$$\lim_{r \downarrow 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)| d^n y = 0. \tag{4.11}$$

The points where (4.11) *holds are called* ***Lebesgue points*** *of f.*

Proof. Decompose f as $f = g + h$, where g is continuous and $\|h\|_1 < \varepsilon$ (Theorem 3.18) and abbreviate

$$D_r(f)(x) := \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)| d^n y.$$

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 $d\mu = |h| d^n x$. 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 4.5 plus $|\{x \,|\, |h(x)| \ge \alpha\}| \le \alpha^{-1}\|h\|_1$ (Problem 4.10), 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. □

Example 4.5. It is easy to see that every point of continuity of f is a Lebesgue point. However, the converse is not true. To see this consider a function $f : \mathbb{R} \to [0, 1]$ which is given by a sum of nonoverlapping spikes centered at $x_j = 2^{-j}$ with base length $2b_j$ and height 1. Explicitly

$$f(x) = \sum_{j=1}^{\infty} \max(0, 1 - b_j^{-1}|x - x_j|)$$

with $b_{j+1} + b_j \le 2^{-j-1}$ (such that the spikes don't overlap). By construction $f(x)$ will be continuous except at $x = 0$. However, if we let the b_j's decay sufficiently fast such that the area of the spikes inside $(-r, r)$ is $o(r)$, the point $x = 0$ will nevertheless be a Lebesgue point. For example, $b_j = 2^{-2j-1}$ will do. $\diamond$

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 4.7. *Let f be (locally) integrable. Then at every Lebesgue point we have*

$$f(x) = \lim_{j\to\infty} \frac{1}{|A_j(x)|} \int_{A_j(x)} f(y) d^n y \tag{4.12}$$

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

$$\frac{1}{|A_j(x)|} \int_{A_j(x)} |f(y) - f(x)| d^n y \le \frac{1}{\varepsilon |B_{r_j}(x)|} \int_{B_{r_j}(x)} |f(y) - f(x)| d^n y$$

and the claim follows. $\square$

If we take f to be the Radon–Nikodym derivative of an absolutely continuous measure, then these results can also be understood as differentiation results for measures. To complement these results we provide the corresponding statement for singular measures.

Lemma 4.8. *Let μ be singular with respect to Lebesgue measure. Then at almost every point x (with respect to Lebesgue measure) we have*

$$\lim_{j\to\infty} \frac{\mu(A_j(x))}{|A_j(x)|} = 0, \tag{4.13}$$

whenever $A_j(x)$ shrinks to x nicely. In particular, we have $(\overline{D}\mu)(x) = 0$ for almost every x.

Proof. By definition μ is supported on a set N of Lebesgue measure zero. Hence choosing $A = \mathbb{R}^n \setminus N$ in the second part of Lemma 4.5 shows $(\overline{D}\mu)(x) = 0$ for a.e. $x \in A$. The remaining claim now follows from $\limsup_{j\to\infty} \frac{\mu(A_j(x))}{|A_j(x)|} \le \varepsilon \limsup_{j\to\infty} \frac{\mu(B_{r_j}(x))}{|B_{r_j}(x)|} \le \varepsilon(\overline{D}\mu)(x)$. □

Combining these two lemmas with Theorem 4.6 shows

Theorem 4.9. *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,*

$$\mu_{ac}(A) = \int_A (D\mu)(x) d^n x. \tag{4.14}$$

In particular, μ is singular with respect to Lebesgue measure if and only if $D\mu = 0$ a.e. with respect to Lebesgue measure.

In the special case of Borel measures on $\mathbb{R}$ the last result reads

Corollary 4.10. *Let μ be a Borel measure on $\mathbb{R}$. Then its distribution function is differentiable a.e. with respect to Lebesgue measure and the derivative equals the Radon–Nikodym derivative.*

Proof. Let f be the Radon–Nikodym derivative. Since the sets $(x, x+r)$ shrink nicely to x as $r \to 0$, Lemma 4.7 implies

$$\lim_{r\to 0} \frac{\mu((x, x+r))}{r} = \lim_{r\to 0} \frac{\mu(x+r) - \mu(x)}{r} = f(x)$$

a.e. Since the same is true for the sets $(x-r, x)$, $\mu(x)$ is differentiable a.e. and $\mu'(x) = f(x)$. □

Using the upper and lower derivatives, we can also give supports for the absolutely and singularly continuous parts.

Theorem 4.11. *The set $\{x | 0 < (D\mu)(x) < \infty\}$ is a support for the absolutely continuous and $\{x | (\underline{D}\mu)(x) = \infty\}$ is a support for the singular part.*

Proof. The first part is immediate from the previous theorem. For the second part first note that by $(\underline{D}\mu)(x) \ge (\underline{D}\mu_{sing})(x)$ we can assume that μ is purely singular. It suffices to 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 $|V_j \setminus K| \le \frac{1}{j}$. For every $x \in K$ there is some $\varepsilon = \varepsilon(x)$ such that $B_\varepsilon(x) \subseteq V_j$

and $\mu(B_{3\varepsilon}(x)) \le k|B_{3\varepsilon}(x)|$. By compactness, finitely many of these balls cover K and hence

$$\mu(K) \le \sum_i \mu(B_{\varepsilon_i}(x_i)).$$

Selecting disjoint balls as in Lemma 4.4 further shows

$$\mu(K) \le \sum_\ell \mu(B_{3\varepsilon_{i_\ell}}(x_{i_\ell})) \le k3^n \sum_\ell |B_{\varepsilon_{i_\ell}}(x_{i_\ell})| \le k3^n |V_j|.$$

Letting $j \to \infty$, we see $\mu(K) \le k3^n|K|$ and by regularity we even have $\mu(A) \le k3^n|A|$ 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 $\mu(A_k) = 0$. □

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.

Example 4.6. Let $X := \mathbb{R}$, $\Sigma := \mathfrak{B}$. If $d\mu(x) := \sum_n \alpha_n d\theta(x - x_n)$ is a sum of Dirac measures, then the set $\{x_n\}$ is clearly a minimal support for μ. Moreover, it is clearly the smallest support as none of the x_n can be removed. If we choose $\{x_n\}$ to be the rational numbers, then $\mathrm{supp}(\mu) = \mathbb{R}$, but $\mathbb{R}$ is not a minimal support, as we can remove the irrational numbers.

On the other hand, if we consider the Lebesgue measure λ, then $\mathbb{R}$ is a minimal support. However, the same is true if we remove any set of measure zero, for example, the Cantor set. In particular, since we can remove any single point, we see that, just like supports, minimal supports are not unique. ◇

Lemma 4.12. *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

$$|M_\varepsilon| = \int_{M_\varepsilon} d^n x \le \frac{1}{\varepsilon} \int_{M_\varepsilon} (D\mu)(x) d^n x = \frac{1}{\varepsilon}\mu_{ac}(M_\varepsilon) \le \frac{1}{\varepsilon}\mu_{ac}(M_0) = 0$$

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 (Problem 4.8).

Example 4.7. The **Cantor function** (also known as **devil's staircase**) 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. Explicitly, we have $f_0(x) = x$ and $f_{n+1} = K(f_n)$, where

$$K(f)(x) := \begin{cases} \frac{1}{2}f(3x), & 0 \le x \le \frac{1}{3}, \\ \frac{1}{2}, & \frac{1}{3} \le x \le \frac{2}{3}, \\ \frac{1}{2}(1 + f(3x-2)), & \frac{2}{3} \le x \le 1. \end{cases}$$

Since K is a contraction on the set of bounded functions, $\|K(g_1)-K(g_0)\|_\infty \le \frac{1}{2}\|g_1 - g_0\|_\infty$, we can define the Cantor function as $\mathfrak{f} = \lim_{n\to\infty} f_n$. By construction $\mathfrak{f}$ is a continuous function which is constant on every subinterval of $[0,1] \setminus C$. Since C is of Lebesgue measure zero, this set is of full Lebesgue measure and hence $\mathfrak{f}' = 0$ 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* 4.7. *Show that*

$$\mu(B_\varepsilon(x)) \le \liminf_{y\to x} \mu(B_\varepsilon(y)) \le \limsup_{y\to x} \mu(B_\varepsilon(y)) \le \mu(\overline{B_\varepsilon(x)}).$$

In particular, conclude that $x \mapsto \mu(B_\varepsilon(x))$ is lower semicontinuous for $\varepsilon > 0$.

Problem* 4.8. *Show that $M := \{x | 0 < (D\mu)(x)\}$ is a minimal support of μ.*

Problem 4.9. *Suppose $\overline{D}\mu \le \alpha$. Show that $d\mu = f\, d^n x$ with $0 \le f \le \alpha$.*

Problem* 4.10 (Markov (also Chebyshev) inequality)**.** *For $f \in L^1(\mathbb{R}^n)$ and $\alpha > 0$ show*

$$|\{x \in A|\, |f(x)| \ge \alpha\}| \le \frac{1}{\alpha}\int_A |f(x)| d^n x.$$

Somewhat more general, assume $g(x) \ge 0$ is nondecreasing and $g(\alpha) > 0$. Then

$$\mu(\{x \in A|\, f(x) \ge \alpha\}) \le \frac{1}{g(\alpha)}\int_A g \circ f\, d\mu.$$

Problem* 4.11. *Let $f \in L^p_{loc}(\mathbb{R}^n)$, $1 \le p < \infty$, then for a.e. $x \in \mathbb{R}^n$ we have*

$$\lim_{r\downarrow 0} \frac{1}{|B_r(x)|}\int_{B_r(x)} |f(y) - f(x)|^p d^n y = 0.$$

The same conclusion hols if the balls are replaced by sets $A_j(x)$ which shrink nicely to x.

Problem 4.12. *Show that the Cantor function is Hölder continuous $|\mathfrak{f}(x) - \mathfrak{f}(y)| \le |x-y|^\alpha$ with exponent $\alpha = \log_3(2)$. (Hint: Show that if a bijection*

$g : [0,1] \to [0,1]$ satisfies a Hölder estimate $|g(x) - g(y)| \le M|x-y|^\alpha$, then so does $K(g)$: $|K(g)(x) - K(g)(y)| \le \frac{3^\alpha}{2} M|x-y|^\alpha$.)

4.3. Complex measures

Let (X, Σ) be some measurable space. A map $\nu : \Sigma \to \mathbb{C}$ is called a **complex measure** if it is σ-additive:

$$\nu(\dot{\bigcup_{n=1}^{\infty}} A_n) = \sum_{n=1}^{\infty} \nu(A_n), \quad A_n \in \Sigma. \tag{4.15}$$

Choosing $A_n = \emptyset$ for all n in (4.15) 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_n, that is, it converges unconditionally and thus absolutely by the Riemann series theorem.

Example 4.8. 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 2.3 and use dominated in place of monotone convergence). In fact, we will show that every complex measure is of this form. ◇

Example 4.9. 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. ◇

When dealing with complex functions f an important object is the positive function $|f|$. 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. However, using $|\nu(A \cap [-1,0))| + |\nu(A \cap [0,1])|$ we do get a positive measure. Motivated by this we introduce the **total variation** of a measure defined as

$$|\nu|(A) := \sup\Big\{ \sum_{k=1}^{n} |\nu(A_k)| \Big| A_k \in \Sigma \text{ disjoint}, \ A = \dot{\bigcup_{k=1}^{n}} A_k \Big\}. \tag{4.16}$$

Note that by construction we have

$$|\nu(A)| \le |\nu|(A). \tag{4.17}$$

Moreover, the total variation is monotone $|\nu|(A) \le |\nu|(B)$ if $A \subseteq B$ and for a positive measure μ we have of course $|\mu|(A) = \mu(A)$.

Theorem 4.13. *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. 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 $B_{n,k}$ of A_n such that

$$|\nu|(A_n) \le \sum_{k=1}^{m_n} |\nu(B_{n,k})| + \frac{\varepsilon}{2^n}.$$

Then

$$\sum_{n=1}^N |\nu|(A_n) \le \sum_{n=1}^N \sum_{k=1}^{m_n} |\nu(B_{n,k})| + \varepsilon \le |\nu|(\dot\bigcup_{n=1}^N A_n) + \varepsilon \le |\nu|(A) + \varepsilon$$

since $\dot\bigcup_{n=1}^N \dot\bigcup_{k=1}^{m_n} B_{n,k} = \dot\bigcup_{n=1}^N A_n$. As 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 $A \setminus B$ such that $|\nu(B)| \ge 1$ and $|\nu|(A \setminus B) = \infty$. 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 $\nu(\dot\bigcup_n B_n)$, a contradiction.

It remains to show existence of this splitting. Let A with $|\nu|(A) = \infty$ be given. Then there is a partition of 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 the above inequality reads $\nu(A_+) + |\nu(A_-)| \ge 2 + |\nu(A_+) - |\nu(A_-)||$ implying (show this) that 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. $\square$

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

$$\nu = \nu_r + \mathrm{i}\nu_i, \qquad \nu_r(A) := \mathrm{Re}(\nu(A)), \quad \nu_i(A) := \mathrm{Im}(\nu(A)). \tag{4.18}$$

Second we can split every real (also called **signed**) measure according to

$$\nu = \nu_+ - \nu_-, \qquad \nu_\pm(A) := \frac{|\nu|(A) \pm \nu(A)}{2}. \tag{4.19}$$

By (4.17) both ν_- and ν_+ are positive measures. This splitting is also known as **Jordan decomposition** of a signed measure. In summary, we can split every complex measure ν into four positive measures

$$\nu = \nu_{r,+} - \nu_{r,-} + \mathrm{i}(\nu_{i,+} - \nu_{i,-}) \tag{4.20}$$

which is also known as Jordan decomposition.

Of course such a decomposition of a signed measure is not unique (we can always add a positive measure to both parts), however, the Jordan decomposition is unique in the sense that it is the smallest possible decomposition.

Lemma 4.14. *Let ν be a complex measure and μ a positive measure satisfying $|\nu(A)| \le \mu(A)$ for all measurable sets A. Then $|\nu| \le \mu$. (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 Jordan 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 $\sum_k |\nu(A_k)| \le \sum_k \mu(A_k) = \mu(A)$ implying $|\nu|(A) \le \mu(A)$. $\square$

Moreover, we also have:

Theorem 4.15. *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 (4.15). Let A_m be given disjoint sets and set $\tilde{A}_n := \dot\bigcup_{m=1}^n A_m$, $A := \dot\bigcup_{m=1}^\infty A_m$. Since we can interchange limits with finite sums, (4.15) 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 4.17). □

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 $\nu(A) = 0$. We say that ν is singular with respect to μ if $|\nu|$ is, that is, there is a measurable set N such that $\mu(N) = 0$ and $|\nu|(X \setminus N) = 0$.

Lemma 4.16. *If μ is a positive and ν a complex measure then $\nu \ll \mu$ if and only if $|\nu| \ll \mu$. Similarly, $\nu \ll \mu$ if and only if this holds for all four measures in the Jordan decomposition.*

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 $|\nu(A)| \le |\nu|(A) = 0$. The second claim now is immediate from Problem 4.17. □

Now we can prove the complex version of the Radon–Nikodym theorem:

Theorem 4.17 (Complex Radon–Nikodym)**.** *Let (X, Σ) be a measurable space, μ a positive σ-finite measure and ν a complex measure. Then there exists a function $f \in L^1(X, d\mu)$ and a set N of μ measure zero, such that*

$$\nu(A) = \nu(A \cap N) + \int_A f\, d\mu. \tag{4.21}$$

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.*

In particular, ν can be uniquely decomposed as $\nu = \nu_{sing} + \nu_{ac}$, where $\nu_{sing}(A) := \nu(A \cap N)$ is singular and $d\nu_{ac} := f\, d\mu$ is absolutely continuous with respect to μ.

Proof. We start with the case where ν is a signed measure. Let $\nu = \nu_+ - \nu_-$ be its Jordan decomposition. Then by Theorem 4.1 there are sets $N_\pm$ and functions $f_\pm$ such that $\nu_\pm(A) = \nu_\pm(A \cap N_\pm) + \int_A f_\pm d\mu$. Since $\nu_\pm$ are finite measures we must have $\int_X f_\pm d\mu \le \nu_\pm(X)$ and hence $f_\pm \in L^1(X, d\mu)$. Moreover, since $N := N_- \cup N_+$ has μ measure zero the remark after Theorem 4.1

implies $\nu_\pm(A) = \nu_\pm(A\cap N) + \int_A f_\pm d\mu$ and hence $\nu(A) = \nu(A\cap N) + \int_A f\, d\mu$ where $f := f_+ - f_- \in L^1(X, d\mu)$.

If ν is complex we can split it into real and imaginary part and use the same reasoning to reduce it to the singed case. If ν is absolutely continuous we have $|\nu|(N) = 0$ and hence $\nu(A\cap N) = 0$. Uniqueness of the decomposition (and hence of f) follows literally as in the proof of Theorem 4.3. □

If ν is absolutely continuous with respect to μ the total variation of $d\nu = f\, d\mu$ is just $d|\nu| = |f| d\mu$:

Lemma 4.18. *Let $d\nu = d\nu_{sing} + f\, d\mu$ be the Lebesgue decomposition of a complex measure ν with respect to a positive σ-finite measure μ. Then*

$$|\nu|(A) = |\nu_{sing}|(A) + \int_A |f| d\mu. \tag{4.22}$$

Proof. We first show $|\nu| = |\nu_{sing}| + |\nu_{ac}|$. Let A be given and let A_k be a partition of A as in (4.16). By the definition of the total variation we can find a partition $A_{sing,k}$ of $A\cap N$ such that $\sum_k |\nu(A_{sing,k})| \ge |\nu_{sing}|(A) - \frac{\varepsilon}{2}$ for arbitrary $\varepsilon > 0$ (note that $\nu_{sing}(A_{sing,k}) = \nu(A_{sing,k})$ as well as $\nu_{sing}(A\cap N') = 0$). Similarly, there is such a partition $A_{ac,k}$ of $A\cap N'$ such that $\sum_k |\nu(A_{ac,k})| \ge |\nu_{ac}|(A) - \frac{\varepsilon}{2}$. Then combining both partitions into a partition A_k for A we obtain $|\nu|(A) \ge \sum_k |\nu(A_k)| \ge |\nu_{sing}|(A) + |\nu_{ac}|(A) - \varepsilon$. Since $\varepsilon > 0$ is arbitrary we conclude $|\nu|(A) \ge |\nu_{sing}|(A) + |\nu_{ac}|(A)$ and as the converse inequality is trivial the first claim follows.

It remains to show $d|\nu_{ac}| = |f|\, d\mu$. Given a partition $A = \dot\bigcup_n A_n$ 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

$$A_k^n := \{x \in A | \frac{k-1}{n} < \frac{\arg(f(x)) + \pi}{2\pi} \le \frac{k}{n}\}, \quad 1 \le k \le n.$$

Then the simple functions

$$s_n(x) := \sum_{k=1}^n \mathrm{e}^{-2\pi\mathrm{i}\frac{k-1}{n} + \mathrm{i}\pi} \chi_{A_k^n}(x)$$

converge to $\mathrm{sign}(f(x)^*)$ for every $x \in A$ 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 4.13):

Corollary 4.19 (polar decomposition)**.** *If ν is a complex measure, then $d\nu = h\, d|\nu|$, where $|h| = 1$.*

If ν is a signed measure, then h is real-valued and we obtain:

Corollary 4.20. *If ν is a signed measure, then $d\nu = h\, d|\nu|$, where $h^2 = 1$. In particular, $d\nu_\pm = \chi_{X_\pm} d|\nu|$, where $X_\pm := h^{-1}(\{\pm 1\})$.*

The decomposition $X = X_+ \uplus X_-$ from the previous corollary is known as **Hahn decomposition**[3] and it is characterized by the property that $\pm\nu(A) \ge 0$ if $A \subseteq X_\pm$. This decomposition is not unique since we can shift sets of $|\nu|$ measure zero from one to the other.

We also briefly mention that the concept of regularity generalizes in a straightforward manner to complex Borel measures. If X is a topological space with its Borel σ-algebra we call ν **(outer/inner) regular** if $|\nu|$ is. It is not hard to see (Problem 4.18):

Lemma 4.21. *A complex measure is regular if and only if all measures in its Jordan decomposition are.*

The subspace of regular Borel measures will be denoted by $\mathcal{M}_{reg}(X)$. Note that it is closed and hence again a Banach space (Problem 4.19).

Clearly we can use Corollary 4.19 to define the integral of a bounded function f with respect to a complex measure $d\nu = h\, d|\nu|$ as

$$\int f\, d\nu := \int f h\, d|\nu|. \tag{4.23}$$

In fact, it suffices to assume that f is integrable with respect to $d|\nu|$ and we obtain

$$\left|\int f\, d\nu\right| \le \int |f| d|\nu|. \tag{4.24}$$

For bounded functions this implies

$$\left|\int_A f\, d\nu\right| \le \|f\|_\infty |\nu|(A). \tag{4.25}$$

Finally, there is an interesting equivalent definition of absolute continuity:

Lemma 4.22. *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*

$$\mu(A) < \delta \quad \Rightarrow \quad |\nu(A)| < \varepsilon, \qquad \forall A \in \Sigma. \tag{4.26}$$

[3] Hans Hanhn (1879–1934), Austrian mathematician and philosopher

Proof. Suppose $\nu \ll \mu$ implying that it is of the from $d\nu = f\, d\mu$. Let $X_n = \{x \in X | \, |f(x)| \le n\}$ and note that $|\nu|(X \setminus X_n) \to 0$ since $X_n \nearrow X$ and $|\nu|(X) < \infty$. Given $\varepsilon > 0$ we can choose n such that $|\nu|(X \setminus X_n) \le \frac{\varepsilon}{2}$ and $\delta = \frac{\varepsilon}{2n}$. Then, if $\mu(A) < \delta$ we have

$$|\nu(A)| \le |\nu|(A \cap X_n) + |\nu|(X \setminus X_n) \le n\, \mu(A) + \frac{\varepsilon}{2} < \varepsilon.$$

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 fact, it can fail for positive measures as the simple counterexample $d\nu(\lambda) = \lambda^2 d\lambda$ on $\mathbb{R}$ shows.

Problem* 4.13. *Prove Corollary 4.19. (Hint: Use the complex Radon–Nikodym theorem to get existence of h. Then show that $1 - |h|$ vanishes a.e.)*

Problem 4.14 (Markov 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*

$$\mu(\{x \in A | \, |f(x)| \ge \alpha\}) \le \frac{|\nu|(A)}{\alpha}.$$

Problem 4.15. *Let ν be a complex and μ a positive measure and suppose $|\nu(A)| \le C\mu(A)$ for all $A \in \Sigma$. Then $d\nu = f\, d\mu$ with $\|f\|_\infty \le C$. (Hint: First show $|\nu|(A) \le C\mu(A)$ and then use Problem 4.3.)*

Problem 4.16. *Let ν be a signed measure and $\nu_\pm$ its Jordan 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* 4.17. *Let ν be a complex measure with Jordan decomposition* (4.20)*. Show the estimate*

$$\frac{1}{\sqrt{2}} \nu_s(A) \le |\nu|(A) \le \nu_s(A), \quad \nu_s = \nu_{r,+} + \nu_{r,-} + \nu_{i,+} + \nu_{i,-}.$$

Show that $|\nu(A)| \le C$ for all measurable sets A implies $\|\nu\| \le 4C$.

Problem* 4.18. *Show Lemma 4.21. (Hint: Problems 1.25 and 4.17.)*

Problem* 4.19. *Let X be a topological space. Show that $\mathcal{M}_{reg}(X) \subseteq \mathcal{M}(X)$ is a closed subspace.*

Problem 4.20. *Let $S \subseteq \Sigma$ be a collection of sets which contains X, generates Σ, and which is closed under finite intersections. Show that a complex measure is uniquely determined by its values on S.*

Moreover, for a positive measure μ and an integrable function f:

$$f = 0 \text{ a.e.} \quad \Leftrightarrow \quad \int_A f\, d\mu = 0, \qquad A \in S.$$

(Hint: Compare with Theorem 1.3.)

Problem 4.21. *Let μ be a complex Borel measures $\mathbb{R}^n$ and $f \in L^1_{loc}(\mathbb{R}^n)$. Define the convolution of μ and f as*

$$(\mu * f)(x) := \int_{\mathbb{R}^n} f(x-y) d\mu(y),$$

whenever $f(x-.)$ is integrable with respect to μ. Generalize Young's inequality to this case: If $f \in L^p(\mathbb{R}^n)$, $1 \le p \le \infty$, then $\mu * f \in L^p(\mathbb{R}^n)$ with

$$\|\mu * f\|_p \le |\mu|(\mathbb{R}^n) \|f\|_p.$$

Problem 4.22. *Define the convolution of two complex Borel measures μ and ν on $\mathbb{R}^n$ via*

$$(\mu * \nu)(A) := \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \chi_A(x+y) d\mu(x) d\nu(y).$$

Note $|\mu * \nu|(\mathbb{R}^n) \le |\mu|(\mathbb{R}^n)|\nu|(\mathbb{R}^n)$. Show that this implies

$$\int_{\mathbb{R}^n} h(x) d(\mu * \nu)(x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} h(x+y) d\mu(x) d\nu(y)$$

for any bounded measurable function h. Conclude that it coincides with our previous definition in case μ and ν are absolutely continuous with respect to Lebesgue measure.

4.4. 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

$$V(P,f) := \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|. \tag{4.27}$$

Note that the triangle inequality implies that adding points to a partition increases the variation: if $P_1 \subseteq P_2$ then $V(P_1,f) \le V(P_2,f)$. The supremum over all partitions

$$V_a^b(f) := \sup_{\text{partitions } P \text{ of } [a,b]} V(P,f) \tag{4.28}$$

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

$$V_a^b(\alpha f) = |\alpha| V_a^b(f), \qquad V_a^b(f+g) \le V_a^b(f) + V_a^b(g) \tag{4.29}$$

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 $P = \{a,x,b\}$)

$$V_a^b(f) = 0 \quad \Leftrightarrow \quad f(x) \equiv c. \tag{4.30}$$

Moreover, any function of bounded variation is in particular bounded (consider again the partition $P = \{a,x,b\}$)

$$\sup_{x\in[a,b]} |f(x)| \le |f(a)| + V_a^b(f). \tag{4.31}$$

Theorem 4.23. *The functions of bounded variation $BV[a,b]$ together with the norm*

$$\|f\|_{BV} := |f(a)| + V_a^b(f) \tag{4.32}$$

are a Banach space. Moreover, by (4.31) *we have $\|f\|_\infty \le \|f\|_{BV}$.*

Proof. By (4.30) we have $\|f\|_{BV} = 0$ if and only if f is constant and $|f(a)| = 0$, that is $f = 0$. Moreover, by (4.29) the norm is homogenous and satisfies the triangle inequality. So let f_n be a Cauchy sequence. Then f_n converges uniformly and pointwise to some bounded function f. Moreover, choose N such that $\|f_n - f_m\|_{BV} < \varepsilon$ whenever $m,n \ge N$. Then for $n \ge N$ and for any fixed partition

$$\begin{aligned} |f(a) - f_n(a)| + V(P, f - f_n) &= \lim_{m\to\infty} \big(|f_m(a) - f_n(a)| + V(P, f_m - f_n)\big) \\ &\le \sup_{m\ge N} \|f_n - f_m\|_{BV} < \varepsilon. \end{aligned}$$

Consequently $\|f - f_n\|_{BV} < \varepsilon$ which shows $f \in BV[a,b]$ as well as $f_n \to f$ in $BV[a,b]$. □

Observe $V_a^a(f) = 0$ as well as (Problem 4.24)

$$V_a^b(f) = V_a^c(f) + V_c^b(f), \qquad c \in [a,b], \tag{4.33}$$

and it will be convenient to set

$$V_b^a(f) = -V_a^b(f). \tag{4.34}$$

Example 4.10. Every Lipschitz continuous function is of bounded variation. In fact, if $|f(x) - f(y)| \le L|x-y|$ for $x,y \in [a,b]$, then $V_a^b(f) \le L(b-a)$. However, (Hölder) continuity is not sufficient (cf. Problem 4.26 and Theorem 4.29 below). ◇

Example 4.11. By the inverse triangle inequality we have

$$V_a^b(|f|) \le V_a^b(f)$$

whereas the converse is not true: The function $f : [0,1] \to \{-1,1\}$ which is 1 on the rational and -1 on the irrational numbers satisfies $V_0^1(f) = \infty$ (show this) and $V_0^1(|f|) = V_0^1(1) = 0$.

From $2^{-1/2}(|\mathrm{Re}(z)| + |\mathrm{Im}(z)|) \le |z| \le |\mathrm{Re}(z)| + |\mathrm{Im}(z)|$ we infer

$$2^{-1/2}\big(V_a^b(\mathrm{Re}(f)) + V_a^b(\mathrm{Im}(f))\big) \le V_a^b(f) \le V_a^b(\mathrm{Re}(f)) + V_a^b(\mathrm{Im}(f))$$

which shows that f is of bounded variation if and only if $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are. $\diamond$

Example 4.12. 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 g is of bounded variation with variation given by $V_a^b(g) = g(a) - g(b)$. Moreover, the sum $f + g$ is of bounded variation with variation given by $V_a^b(f+g) \le V_a^b(f) + V_a^b(g)$. The following theorem shows that the converse is also true. $\diamond$

Theorem 4.24 (Jordan)**.** *Let $f : [a,b] \to \mathbb{R}$ be of bounded variation, then f can be decomposed as*

$$f(x) = f_+(x) - f_-(x), \qquad f_\pm(x) := \frac{1}{2}\left(V_a^x(f) \pm f(x)\right), \tag{4.35}$$

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. $\square$

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

$$V_a^b(f) := \lim_{c\downarrow a, d\uparrow b} V_c^d(f). \tag{4.36}$$

In this respect the following lemma is of interest (whose proof is left as an exercise):

Lemma 4.25. *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 if 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 1.13 (since all four functions are bounded, so are the associated measures).

Theorem 4.26. *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* (1.17). *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 B.42 from [**22**]), and thus for all Borel sets by outer regularity. Hence Lemma 4.14 implies $\mu = |\nu|$ and hence $|\nu|(x) = V_0^x(f)$. □

As a consequence of Corollary 4.10 we note:

Corollary 4.27. *Functions of bounded variation are differentiable a.e. with respect to Lebesgue measure.*

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

$$\sum_k |y_k - x_k| < \delta \quad \Rightarrow \quad \sum_k |f(y_k) - f(x_k)| < \varepsilon \tag{4.37}$$

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, $AC[a,b] \subset C[a,b]$.

Example 4.13. Every Lipschitz continuous function is absolutely continuous. In fact, if $|f(x) - f(y)| \le L|x - y|$ for $x, y \in [a,b]$, then we can choose

$\delta = \frac{\varepsilon}{L}$. In particular, $C^1[a,b] \subset AC[a,b]$. Note that Hölder continuity is neither sufficient (cf. Problem 4.26 and Theorem 4.29 below) nor necessary (cf. Problem 4.31). ⋄

Theorem 4.28. *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 subinterval). Moreover, in this case the distribution function $\nu(x)$ is differentiable almost everywhere and*

$$\nu(x) = \nu(0) + \int_0^x \nu'(y)dy \tag{4.38}$$

with ν' integrable, $\int_{\mathbb{R}} |\nu'(y)|dy = |\nu|(\mathbb{R})$.

Proof. Suppose the measure ν is absolutely continuous. Now (4.37) follows from (4.26) 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 (4.26). To this end fix ε and choose δ such that $\nu(x)$ satisfies (4.37). By outer regularity it suffices to consider the case where A is open. Moreover, by Problem B.42 from **[22]**, 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

$$|\nu(O)| = \Big| \sum_k \big(\nu(y_k) - \nu(x_k)\big)\Big| \le \sum_k |\nu(y_k) - \nu(x_k)| \le \varepsilon$$

as required.

The rest follows from Corollary 4.10. □

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 4.29. *A function $f : [a,b] \to \mathbb{C}$ is absolutely continuous if and only if it is of the form*

$$f(x) = f(a) + \int_a^x g(y)dy \tag{4.39}$$

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

$$V_a^x(f) = \int_a^x |g(y)|dy. \tag{4.40}$$

Proof. This is just a reformulation of the previous result. To see the last claim combine the last part of Theorem 4.26 with Lemma 4.18. □

In particular, since the fundamental theorem of calculus fails for the Cantor function, this function is an example of a continuous function which is not absolutely continuous. Note that even if f is differentiable everywhere the fundamental theorem of calculus might fail (Problem 4.32).

Finally, we note that in this case the **integration by parts formula** continues to hold.

Lemma 4.30. *Let $f, g \in BV[a,b]$, then*

$$\int_{[a,b)} f(x-)dg(x) = f(b-)g(b-) - f(a-)g(a-) - \int_{[a,b)} g(x+)df(x) \quad (4.41)$$

as well as

$$\int_{[a,b)} f(x+)dg(x) = f(b-)g(b-) - f(a-)g(a-) - \int_{[a,b)} g(x-)df(x). \quad (4.42)$$

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. □

Note that if we require $f(x) = \frac{f(x+)+f(x-)}{2}$ and $g(x) = \frac{g(x+)+g(x-)}{2}$ then we can add both formulas to obtain the more symmetric form $\int_{[a,b)} f\,dg = (fg)(b-) - (fg)(a-) - \int_{[a,b)} g\,df$. If both $f, g \in AC[a,b]$ this takes the usual form

$$\int_a^b f(x)g'(x)dx = f(b)g(b) - f(a)g(a) - \int_a^b g(x)f'(x)dx. \quad (4.43)$$

For further generalizations of classical calculus rules to absolutely continuous functions see the problems.

Problem 4.23. *Compute $V_a^b(f)$ for $f(x) = \operatorname{sign}(x)$ on $[a,b] = [-1,1]$.*

Problem* 4.24. *Show* (4.33).

Problem 4.25. *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 $j \ge 2$ but not for $j = 1$.*

Problem* 4.26. *Let $\alpha \in (0,1)$ and $\beta > 1$ with $\alpha\beta \le 1$. Set $M := \sum_{k=1}^\infty k^{-\beta}$, $x_0 := 0$, and $x_n := M^{-1}\sum_{k=1}^n k^{-\beta}$. Then we can define a function on $[0,1]$ as follows: Set $g(0) := 0$, $g(x_n) := n^{-\beta}$, and*

$$g(x) := c_n \left|x - t_n x_n - (1-t_n)x_{n+1}\right|, \qquad x \in [x_n, x_{n+1}],$$

where c_n and $t_n \in [0,1]$ are chosen such that g is (Lipschitz) continuous. Show that $f = g^\alpha$ is Hölder continuous of exponent α but not of bounded variation. (Hint: What is the variation on each subinterval?)

Problem* 4.27 (Takagi function)**.** *The Takagi function (also blancmange function) is the fixed point of the contraction $K(f)(x) := s(x) + \frac{1}{2}f(2x)$ for one-periodic functions, where $s(x) := \mathrm{dist}(x,\mathbb{Z}) = \min_{n\in\mathbb{Z}}|x-n|$:*

$$b(x) := \lim_{n\to\infty} b_n(x), \qquad b_n(x) = \sum_{k=0}^n \frac{s(2^k x)}{2^k}. \tag{4.44}$$

(i) Show that the Takagi function is Hölder continuous $|b(x)-b(y)| \le M_\alpha|x-y|^\alpha$ with $M_\alpha = (2^{1-\alpha}-1)^{-1}$ for any $\alpha \in (0,1)$. (Hint: Show that the contraction leaves the space of periodic functions satisfying such an estimate invariant. Note $|s(x)-s(y)| \le 2^{\alpha-1}|x-y|^\alpha$.)

(ii) Show that the Takagi function is not of bounded variation. (Hint: Note that b_n is piecewise linear on the intervals of the partition $P_n = \{k2^{-n-1} | 0 \le k \le 2^{n+1}\}$. Now investigate the derivative on this intervals and in particular how they change in each step. You should get $V_0^1(b_n) = V(P_n, b_n) = \sum_{k=0}^{\lfloor (n+1)/2\rfloor}\binom{2k}{k}2^{-2k}$.)

(iii) Show that the Takagi function is nowhere differentiable. By Corollary 4.27 this also shows that b is not of bounded variation. (Hint: Consider a difference quotient $\frac{b(z)-b(y)}{z-y}$, where y, z are dyadic rationals.)

Problem 4.28. *Show that if $f \in BV[a,b]$ then so is f^*, $|f|$ and*

$$V_a^b(f^*) = V_a^b(f), \qquad V_a^b(|f|) \le V_a^b(f).$$

Moreover, show

$$V_a^b(\mathrm{Re}(f)) \le V_a^b(f), \qquad V_a^b(\mathrm{Im}(f)) \le V_a^b(f).$$

Problem 4.29. *Show that if $f, g \in BV[a,b]$ then so is $f\,g$ and*

$$V_a^b(f\,g) \le V_a^b(f)\sup|g| + V_a^b(g)\sup|f|.$$

Hence, together with the norm $\|f\|_{BV} := \|f\|_\infty + V_a^b(f)$ the space $BV[a,b]$ is a Banach algebra.

Problem 4.30. *Show that for $f \in BV(\mathbb{R})$ we have*

$$\int_{\mathbb{R}} |f(x+y) - f(x)| dx \leq V_{-\infty}^{\infty}(f)|y|.$$

Problem* 4.31. *Show that if $f \in AC[a,b]$ and $f' \in L^p(a,b)$,$p > 1$, then f is Hölder continuous:*

$$|f(x) - f(y)| \leq \|f'\|_p |x-y|^{1-\frac{1}{p}}.$$

Show that the function $f(x) = -\log(x)^{-1}$ is absolutely continuous but not Hölder continuous on $[0, \frac{1}{2}]$.

Problem* 4.32. *Consider $f(x) := x^2 \sin(\frac{\pi}{x^2})$ on $[0,1]$ (here $f(0) = 0$). Show that f is differentiable everywhere and compute its derivative. Show that its derivative is* not *integrable. In particular, this function is not absolutely continuous and the fundamental theorem of calculus does not hold for this function.*

Problem 4.33. *Show that the function from the previous problem is Hölder continuous of exponent $\frac{1}{2}$. (Hint: Consider $0 < x < y$. There is an $x' < y$ with $f(x') = f(x)$ and $(x')^{-2} - y^{-2} \leq 2\pi$. Hence $(x')^{-1} - y^{-1} \leq \sqrt{2\pi}$). Now use the Cauchy–Schwarz inequality to estimate $|f(y) - f(x)| = |f(y) - f(x')| = |\int_{x'}^{y} 1 \cdot f'(t) dt|$.)*

Problem 4.34. *Suppose $f \in AC[a,b]$. Show that f' vanishes a.e. on $f^{-1}(c)$ for every c. (Hint: Split the set $f^{-1}(c)$ into isolated points and the rest.)*

Problem 4.35. *A function $f : [a,b] \to \mathbb{R}$ is said to have the **Luzin** N **property** if it maps Lebesgue null sets to Lebesgue null sets. Show that absolutely continuous functions have the Luzin N property. Show that the Cantor function does not have the Luzin N property. (Hint: Use (4.37) and recall that: A set $A \subseteq \mathbb{R}$ is a null set if and only if for every ε there exists a countable set of intervals I_j which cover A and satisfy $\sum_j |I_j| < \varepsilon$.)*

Problem* 4.36. *The variation of a function $f : [a,b] \to \mathbb{R}^n$ is defined by replacing the absolute value by the Euclidean norm in the definiton. Show that $f : [a,b] \to \mathbb{R}^n$ is of bounded variation if and only if every component is of bounded variation.*

*Recall that a curve $\gamma : [a,b] \to \mathbb{R}^n$ is **rectifiable** if $V_a^b(\gamma) < \infty$. In this case $V_a^b(\gamma)$ is called the **arc length** of γ. Conclude that γ is rectifiable if and only if each of its coordinate functions is of bounded variation. Show that if each coordinate function is absolutely continuous, then*

$$V_a^b(\gamma) = \int_a^b |\gamma'(t)| dt.$$

(Hint: For the last part note that one inequality is easy. Then reduce it to the case when γ' is a step function.)

Problem 4.37. *Show that if $f, g \in AC[a,b]$ then so is $f\,g$ and the* ***product rule*** $(f\,g)' = f'g + f\,g'$ *holds. Conclude that $AC[a,b]$ is a closed subalgebra of the Banach algebra $BV[a,b]$. (Hint: Integration by parts. For the second part use Problem 4.29 and* (4.40)*.)*

Problem 4.38. *Show that $f \in AC[a,b]$ is nondecreasing iff $f' \geq 0$ a.e. and prove the* ***substitution rule***

$$\int_a^b g(f(x))f'(x)dx = \int_{f(a)}^{f(b)} g(y)dy$$

in this case. Conclude that if h is absolutely continuous, then so is $h \circ f$ and the ***chain rule*** $(h \circ f)' = (h' \circ f)f'$ *holds.*

Moreover, show that $f \in AC[a,b]$ is strictly increasing iff $f' > 0$ a.e. In this case f^{-1} is also absolutely continuous and the ***inverse function rule***

$$(f^{-1})' = \frac{1}{f' \circ f^{-1}}$$

holds. (Hint: (2.75)*.)*

Problem 4.39. *Consider $f(x) := x^2 \sin(\frac{\pi}{x})$ on $[0,1]$ (here $f(0) = 0$) and $g(x) = \sqrt{|x|}$. Show that both functions are absolutely continuous, but $g \circ f$ is not. Hence the monotonicity assumption in Problem 4.38 is important. Show that if f is absolutely continuous and g Lipschitz, then $g \circ f$ is absolutely continuous.*

Problem 4.40. *Consider $f(x) := \frac{1}{2}(c(x)+x)$ on $[0,1]$, where c is the Cantor function. Show that the inverse f^{-1} is Lipschitz continuous with Lipschitz constant $\frac{1}{2}$. In particular f^{-1} is absolutely continuous while f is not. Hence the $f' > 0$ assumption in Problem 4.38 is important.*

Problem 4.41 (Characterization of the exponential function)**.** *Show that every nontrivial locally integrable function $f : \mathbb{R} \to \mathbb{C}$ satisfying*

$$f(x+y) = f(x)f(y), \qquad x, y \in \mathbb{R},$$

is of the from $f(x) = \mathrm{e}^{\alpha x}$ for some $\alpha \in \mathbb{C}$. (Hint: Start with $F(x) = \int_0^x f(t)dt$ and show $F(x+y) - F(x) = F(y)f(x)$. Conclude that f is absolutely continuous.)

Problem 4.42. *Let $X \subseteq \mathbb{R}$ be an interval, Y some measure space, and $f : X \times Y \to \mathbb{C}$ some measurable function. Suppose $x \mapsto f(x,y)$ is absolutely continuous for a.e. y such that*

$$\int_a^b \int_Y \left| \frac{\partial}{\partial x} f(x,y) \right| d\mu(y)dx < \infty \tag{4.45}$$

for every compact interval $[a,b] \subseteq X$ and $\int_Y |f(c,y)|d\mu(y) < \infty$ for one $c \in X$.

Show that

$$F(x) := \int_Y f(x,y)\, d\mu(y) \tag{4.46}$$

is absolutely continuous and

$$F'(x) = \int_Y \frac{\partial}{\partial x} f(x,y)\, d\mu(y) \tag{4.47}$$

in this case. (Hint: Fubini.)

Chapter 5

Even more measure theory

5.1. Hausdorff measure

The purpose of the Hausdorff measure is to measure lower dimensional objects in $\mathbb{R}^n$, which have Lebesgue measure 0 and hence are not seen by Lebesgue measure. The idea is, that if you cover a curve by sets of diameter δ, the number of sets required will be roughly proportional to δ^{-1} and to the length of the curve. Moreover, one expects to get (a multiple of) the length by looking at the quotient as $\delta \to 0$. Slightly more general, if we want to cover a submanifold of dimension $m \le n$, we expect that number of sets is related to δ^m and the area of the submanifold. So this suggests the following procedure to compute the area (up to a normalization constant): Choose sets of diameter at most δ which cover the submanifold. Count the number of sets weighted by their diameter to the power of m. Choose the cover optimal (i.e. with minimal overlap) and let $\delta \to 0$. Now let us turn to the details.

Throughout this section we will assume that (X, d) is a metric space. Recall that the **diameter** of a subset $A \subseteq X$ is defined by $\operatorname{diam}(A) := \sup_{x,y\in A} d(x, y)$ with the convention that $\operatorname{diam}(\emptyset) = 0$. A cover $\{A_j\}$ of A is called a **δ-cover** if it is countable and if $\operatorname{diam}(A_j) \le \delta$ for all j.

For $A \subseteq X$ and $\alpha \ge 0$, $\delta > 0$ we define

$$h_\delta^{\alpha,*}(A) := \inf \left\{ \sum\nolimits_j \operatorname{diam}(A_j)^\alpha \,\middle|\, \{A_j\} \text{ is a } \delta\text{-cover of } A \right\} \in [0, \infty], \tag{5.1}$$

which is an outer measure by Lemma 1.8. In the case $\alpha = 0$ and $A = \emptyset$ we also regard the empty cover as a valid cover such that $h_\delta^{0,*}(\emptyset) = 0$. As δ decreases the number of admissible covers decreases and hence $h^\alpha_\delta(A)$ increases as a function of δ. Thus the limit

$$h^{\alpha,*}(A) := \lim_{\delta\downarrow 0} h^{\alpha,*}_\delta(A) = \sup_{\delta>0} h^{\alpha,*}_\delta(A) \tag{5.2}$$

exists. Moreover, it is not hard to see that it is again an outer measure (Problem 5.1) and by Theorem 1.9 we get a measure. To show that the σ-algebra from Theorem 1.9 contains all Borel sets it suffices to show that μ^* is a metric outer measure (cf. Lemma 1.11).

Now if A_1, A_2 with $\operatorname{dist}(A_1, A_2) > 0$ are given and $\delta < \operatorname{dist}(A_1, A_2)$ then every set from a cover for $A_1 \cup A_1$ can have nonempty intersection with at most one of both sets. Consequently $h^{*,\alpha}_\delta(A_1 \cup A_2) = h^{*,\alpha}_\delta(A_1) + h^{*,\alpha}_\delta(A_2)$ for $\delta < \operatorname{dist}(A_1, A_2)$ implying $h^{*,\alpha}(A_1 \cup A_2) = h^{*,\alpha}(A_1) + h^{*,\alpha}(A_2)$. Hence $h^{\alpha,*}$ is a metric outer measure and the resulting measure h^α on the Borel σ-algebra is called the α-dimensional **Hausdorff measure**. Note that if X is a vector space with a translation invariant metric, then the diameter of a set is translation invariant and so will be h^α.

Example 5.1. For example, consider the case $\alpha = 0$. Suppose $A = \{x, y\}$ consists of two points. Then $h^0_\delta(A) = 1$ for $\delta \ge d(x, y)$ and $h^0_\delta(A) = 2$ for $\delta < |x - y|$. In particular, $h^0(A) = 2$. Similarly, it is not hard to see (show this) that $h^0(A)$ is just the number of points in A, that is, h^0 is the counting measure on X. ◇

Example 5.2. At the other extreme, if $X := \mathbb{R}^n$, we have $h^n(A) = c_n|A|$, where $|A|$ denotes the Lebesgue measure of A. In fact, since the square $(0, 1]$ has $\operatorname{diam}((0, 1]) = \sqrt{n}$ and we can cover it by k^n squares of side length $\frac{1}{k}$, we see $h^n((0, 1]) \le n^{n/2}$. Thus h^n is a translation invariant Borel measure which hence must be a multiple of Lebesgue measure. To identify c_n we will need the isodiametric inequality. ◇

For the rest of this section we restrict ourselves to the case $X = \mathbb{R}^n$. Note that while in dimension more than one it is not true that a set of diameter d is contained in a ball of diameter d (a counter example is an equilateral triangle), we at least have the following:

Lemma 5.1 (Isodiametric inequality)**.** *For every Borel set $A \in \mathfrak{B}^n$ we have*

$$|A| \le \frac{V_n}{2^n} \operatorname{diam}(A)^n.$$

In other words, a ball is the set with the largest volume when the diameter is kept fixed.

Proof. The trick is to transform A to a set with smaller diameter but same volume via Steiner symmetrization. To this end we build up A from slices obtained by keeping the second coordinate fixed: $A(y) = \{x \in \mathbb{R} | (x, y) \in A\}$.

Now we build a new set $\tilde{A}$ by replacing $A(y)$ with a symmetric interval with the same measure, that is, $\tilde{A} = \{(x,y) | \, |x| \le |A(y)|/2\}$. Note that by Theorem 2.11 $\tilde{A}$ is measurable with $|A| = |\tilde{A}|$. Hence the same is true for $\tilde{A} = \{(x,y) | \, |x| \le |A(y)|/2\} \setminus \{(0,y) | \, A(y) = \emptyset\}$ if we look at the complete Lebesgue measure, since the set we subtract is contained in a set of measure zero.

Moreover, if $\bar{I}$, $\bar{J}$ are closed intervals, then $\sup_{x_1 \in \bar{I}, x_2 \in \bar{J}} |x_2 - x_1| \ge \frac{|\bar{I}|}{2} + \frac{|\bar{J}|}{2}$ (without loss both intervals are compact and $i \le j$, where i, j are the midpoints of $\bar{I}$, $\bar{J}$; then the sup is at least $(j + \frac{|\bar{J}|}{2}) - (i - \frac{|\bar{I}|}{2})$). If I, J are Borel sets in $\mathfrak{B}^1$ and $\bar{I}$, $\bar{J}$ are their respective closed convex hulls, then $\sup_{x_1 \in I, x_2 \in J} |x_2 - x_1| = \sup_{x_1 \in \bar{I}, x_2 \in \bar{J}} |x_2 - x_1| \ge \frac{|\bar{I}|}{2} + \frac{|\bar{J}|}{2} \ge \frac{|I|}{2} + \frac{|J|}{2}$. Thus for $(\tilde{x}_1, y_1), (\tilde{x}_2, y_2) \in \tilde{A}$ we can find $(x_1, y_1), (x_2, y_2) \in A$ with $|x_1 - x_2| \ge |\tilde{x}_1 - \tilde{x}_2|$ implying $\operatorname{diam}(\tilde{A}) \le \operatorname{diam}(A)$.

In addition, if A is symmetric with respect to $x_j \mapsto -x_j$ for some $2 \le j \le n$, then so is $\tilde{A}$.

Now repeat this procedure with the remaining coordinate directions, to obtain a set $\tilde{A}$ which is symmetric with respect to reflection $x \mapsto -x$ and satisfies $|A| = |\tilde{A}|$, $\operatorname{diam}(\tilde{A}) \le \operatorname{diam}(A)$. By symmetry $\tilde{A}$ is contained in a ball of diameter $\operatorname{diam}(\tilde{A})$ and the claim follows. □

Lemma 5.2. *For every Borel set $A \in \mathfrak{B}^n$ we have*

$$h^n(A) = \frac{2^n}{V_n} |A|.$$

Proof. Using (1.37) (for $\mathcal{C}$ choose the collection of all open balls of radius at most δ) one infers $h^n_\delta(A) \le \frac{2^n}{V_n}|A|$ implying $c_n \le 2^n/V_n$. The converse inequality $h^n_\delta(A) \ge \frac{2^n}{V_n}|A|$ follows from the isodiametric inequality. □

We have already noted that the Hausdorff measure is translation invariant. Similarly, it is also invariant under orthogonal transformations (since the diameter is). Moreover, using the fact that for $\lambda > 0$ the map $\lambda : x \mapsto \lambda x$ gives rise to a bijection between δ-covers and (δ/λ)-covers, we easily obtain the following scaling property of Hausdorff measures.

Lemma 5.3. *Let $\lambda > 0$, $d \in \mathbb{R}^n$, $O \in O(n)$ an orthogonal matrix, and A be a Borel set of $\mathbb{R}^n$. Then*

$$h^\alpha(\lambda O A + d) = \lambda^\alpha h^\alpha(A). \tag{5.3}$$

Moreover, Hausdorff measures also behave nicely under uniformly Hölder continuous maps.

Lemma 5.4. *Suppose $f : A \to \mathbb{R}^n$ is uniformly Hölder continuous with exponent $\gamma > 0$, that is,*

$$|f(x) - f(y)| \le c|x - y|^\gamma \quad \text{for all } x, y \in A. \tag{5.4}$$

Then

$$h^\alpha(f(A)) \le c^\alpha h^{\alpha\gamma}(A). \tag{5.5}$$

Proof. A simple consequence of the fact that for every δ-cover $\{A_j\}$ of a Borel set A, the set $\{f(A \cap A_j)\}$ is a $(c\delta^\gamma)$-cover for the Borel set $f(A)$. □

Now we are ready to define the Hausdorff dimension. First note that h_δ^α is non increasing with respect to α for $\delta < 1$ and hence the same is true for h^α. Moreover, for $\alpha \le \beta$ we have $\sum_j \operatorname{diam}(A_j)^\beta \le \delta^{\beta-\alpha} \sum_j \operatorname{diam}(A_j)^\alpha$ and hence

$$h_\delta^\beta(A) \le \delta^{\beta-\alpha}\, h_\delta^\alpha(A) \le \delta^{\beta-\alpha}\, h^\alpha(A). \tag{5.6}$$

Thus if $h^\alpha(A)$ is finite, then $h^\beta(A) = 0$ for every $\beta > \alpha$. Hence there must be one value of α where the Hausdorff measure of a set jumps from ∞ to 0. This value is called the **Hausdorff dimension**

$$\dim_H(A) = \inf\{\alpha | h^\alpha(A) = 0\} = \sup\{\alpha | h^\alpha(A) = \infty\}. \tag{5.7}$$

It is also not hard to see that we have $\dim_H(A) \le n$ (Problem 5.3).

The following observations are useful when computing Hausdorff dimensions. First the Hausdorff dimension is monotone, that is, for $A \subseteq B$ we have $\dim_H(A) \le \dim_H(B)$. Furthermore, if A_j is a (countable) sequence of Borel sets we have $\dim_H(\bigcup_j A_j) = \sup_j \dim_H(A_j)$ (show this).

Using Lemma 5.4 it is also straightforward to show

Lemma 5.5. *Suppose $f : A \to \mathbb{R}^n$ is uniformly Hölder continuous with exponent $\gamma > 0$, that is,*

$$|f(x) - f(y)| \le c|x - y|^\gamma \quad \text{for all } x, y \in A, \tag{5.8}$$

then

$$\dim_H(f(A)) \le \frac{1}{\gamma} \dim_H(A). \tag{5.9}$$

Similarly, if f is bi-Lipschitz, that is,

$$a|x - y| \le |f(x) - f(y)| \le b|x - y| \quad \text{for all } x, y \in A, \tag{5.10}$$

then

$$\dim_H(f(A)) = \dim_H(A). \tag{5.11}$$

Example 5.3. The Hausdorff dimension of the Cantor set (see Example 1.28) is

$$\dim_H(C) = \frac{\log(2)}{\log(3)}. \tag{5.12}$$

To see this let $\delta = 3^{-n}$. Using the δ-cover given by the intervals forming C_n used in the construction of C we see $h^\alpha_\delta(C) \le (\frac{2}{3^\alpha})^n$. Hence for $\alpha = d = \log(2)/\log(3)$ we have $h^d_\delta(C) \le 1$ implying $\dim_H(C) \le d$.

The reverse inequality is a little harder. Let $\{A_j\}$ be a cover and $\delta < \frac{1}{3}$. It is clearly no restriction to assume that all V_j are open intervals. Moreover, finitely many of these sets cover C by compactness. Drop all others and fix j. Furthermore, increase each interval A_j by at most ε.

For V_j there is a k such that

$$\frac{1}{3^{k+1}} \le |A_j| < \frac{1}{3^k}.$$

Since the distance of two intervals in C_k is at least 3^{-k} we can intersect at most one such interval. For $n \ge k$ we see that V_j intersects at most $2^{n-k} = 2^n(3^{-k})^d \le 2^n 3^d |A_j|^d$ intervals of C_n.

Now choose n larger than all k (for all A_j). Since $\{A_j\}$ covers C, we must intersect all 2^n intervals in C_n. So we end up with

$$2^n \le \sum_j 2^n 3^d |A_j|^d,$$

which together with our first estimate yields

$$\frac{1}{2} \le h^d(C) \le 1.$$

Observe that this result can also formally be derived from the scaling property of the Hausdorff measure by solving the identity

$$\begin{aligned} h^\alpha(C) &= h^\alpha(C \cap [0, \tfrac{1}{3}]) + h^\alpha(C \cap [\tfrac{2}{3}, 1]) = 2\, h^\alpha(C \cap [0, \tfrac{1}{3}]) \\ &= \frac{2}{3^\alpha} h^\alpha(3(C \cap [0, \tfrac{1}{3}])) = \frac{2}{3^\alpha} h^\alpha(C) \end{aligned} \tag{5.13}$$

for α. However, this is possible only if we already know that $0 < h^\alpha(C) < \infty$ for some α. ⋄

Problem* 5.1. *Suppose $\{\mu^*_\alpha\}_\alpha$ is a family of outer measures on X. Then $\mu^* = \sup_\alpha \mu^*_\alpha$ is again an outer measure.*

Problem 5.2. *Let $L = [0, 1] \times \{0\} \subseteq \mathbb{R}^2$. Show that $h^1(L) = 1$.*

Problem* 5.3. *Show that $\dim_H(U) \le n$ for every $U \subseteq \mathbb{R}^n$.*

5.2. Infinite product measures

In Section 2.2 we have dealt with finite products of measures. However, in some situations even infnite products are of interest. For example, in probability theory one describes a single random experiment by a probability measure and performing n independent trials is modeled by taking the n-fold product. If one is interested in the behavior of certain quantities in the limit as $n \to \infty$ one is naturally lead to an infinite product.

Hence our goal is to to define a probability measure on the product space $\mathbb{R}^{\mathbb{N}} = \bigtimes_{\mathbb{N}} \mathbb{R}$. We can regard $\mathbb{R}^{\mathbb{N}}$ as the set of all sequences $x = (x_j)_{j\in\mathbb{N}}$. A **cylinder set** in $\mathbb{R}^{\mathbb{N}}$ is a set of the form $A \times \mathbb{R}^{\mathbb{N}} \subseteq \mathbb{R}^N$ with $A \subseteq \mathbb{R}^n$ for some n. We equip $\mathbb{R}^{\mathbb{N}}$ with the product topology, that is, the topology generated by open cylinder sets with A open (which are a base for the product topology since they are closed under intersections — note that the cylinder sets are precisely the finite intersections of preimages of projections). Then the Borel σ-algebra $\mathfrak{B}^{\mathbb{N}}$ on $\mathbb{R}^{\mathbb{N}}$ is the σ-algebra generated by cylinder sets with $A \in \mathfrak{B}^n$.

Now suppose we have probability measures μ_n on $(\mathbb{R}^n, \mathfrak{B}^n)$ which are **consistent** in the sense that

$$\mu_{n+1}(A \times \mathbb{R}) = \mu_n(A), \qquad A \in \mathfrak{B}^n. \tag{5.14}$$

Example 5.4. The prototypical example would be the case where μ is a probability measure on $(\mathbb{R}, \mathfrak{B})$ and $\mu_n = \mu \otimes \cdots \otimes \mu$ is the n-fold product. Slightly more general, one could even take probability measures ν_j on $(\mathbb{R}, \mathfrak{B})$ and consider $\mu_n = \nu_1 \otimes \cdots \otimes \nu_n$. $\diamond$

Theorem 5.6 (Kolmogorov extension theorem)**.** *Suppose that we have a consistent family of probability measures $(\mathbb{R}^n, \mathfrak{B}^n, \mu_n)$, $n \in \mathbb{N}$. Then there exists a unique probability measure μ on $(\mathbb{R}^{\mathbb{N}}, \mathfrak{B}^{\mathbb{N}})$ such that $\mu(A \times \mathbb{R}^{\mathbb{N}}) = \mu_n(A)$ for all $A \in \mathfrak{B}^n$.*

Proof. Consider the algebra $\mathcal{A}$ of all Borel cylinder sets which generates $\mathfrak{B}^{\mathbb{N}}$ as noted above. Then $\mu(A \times \mathbb{R}^{\mathbb{N}}) = \mu_n(A)$ for $A \times \mathbb{R}^{\mathbb{N}} \in \mathcal{A}$ defines an additive set function on $\mathcal{A}$. Indeed, by our consistency assumption different representations of a cylinder set will give the same value and (finite) additivity follows from additivity of μ_n. Hence it remains to verify that μ is a premeasure such that we can apply the extension results from Section 1.3.

Now in order to show σ-additivity it suffices to show continuity from above, that is, for given sets $A_n \in \mathcal{A}$ with $A_n \searrow \emptyset$ we have $\mu(A_n) \searrow 0$. Suppose to the contrary that $\mu(A_n) \searrow \varepsilon > 0$. Moreover, by repeating sets in the sequence A_n if necessary, we can assume without loss of generality that $A_n = \tilde{A}_n \times \mathbb{R}^{\mathbb{N}}$ with $\tilde{A}_n \subseteq \mathbb{R}^n$. Next, since μ_n is inner regular, we can find a compact set $\tilde{K}_n \subseteq \tilde{A}_n$ such that $\mu_n(\tilde{K}_n) \ge \frac{\varepsilon}{2}$. Furthermore, since A_n is decreasing we can arrange $K_n = \tilde{K}_n \times \mathbb{R}^{\mathbb{N}}$ to be decreasing as well: $K_n \searrow \emptyset$. However, by compactness of $\tilde{K}_n$ we can find a sequence with $x \in K_n$ for all n (Problem 5.4), a contradiction. $\square$

Example 5.5. The simplest example for the use of this theorem is a discrete **random walk** in one dimension. So we suppose we have a fictitious particle confined to the lattice $\mathbb{Z}$ which starts at 0 and moves one step to the left or right depending on whether a fictitious coin gives head or tail (the imaginative reader might also see the price of a share here). Somewhat more formal, we take $\mu_1 = (1-p)\delta_{-1} + p\delta_1$ with $p \in [0,1]$ being the probability of moving right and $1-p$ the probability of moving left. Then the infinite product will give a probability measure for the sets of all paths $x \in \{-1,1\}^{\mathbb{N}}$ and one might try to answer questions like if the location of the particle at step n, $s_n = \sum_{j=1}^n x_j$ remains bounded for all $n \in \mathbb{N}$, etc. $\diamond$

Example 5.6. Another classical example is the **Anderson model**. The discrete one-dimensional Schrödinger equation for a single electron in an external potential is given by the difference operator

$$(Hu)_n = u_{n+1} + u_{n-1} + q_n u_n, \qquad u \in \ell^2(\mathbb{Z}),$$

where the *potential* q_n is a bounded real-valued sequence. A simple model for an electron in a crystal (where the atoms are arranged in a periodic structure) is hence the case when q_n is periodic. But what happens if you introduce some random impurities (known as doping in the context of semiconductors)? This can be modeled by $q_n = q_n^0 + x_n(q_n^1 - q_n^0)$ where $x \in \{0,1\}^{\mathbb{N}}$ and we can take $\mu_1 = (1-p)\delta_0 + p\delta_1$ with $p \in [0,1]$ the probability of an impurity being present. $\diamond$

Problem* 5.4. *Suppose $K_n \subseteq \mathbb{R}^n$ is a sequence of nonempty compact sets which are nesting in the sense that $K_{n+1} \subseteq K_n \times \mathbb{R}$. Show that there is a sequence $x = (x_j)_{j\in\mathbb{N}}$ with $(x_1, \dots, x_n) \in K_n$ for all n. (Hint: Choose x_m by considering the projection of K_n onto the m'th coordinate and using the finite intersection property of compact sets.)*

5.3. Convergence in measure and a.e. convergence

Let (X, Σ, μ) be a measure space. A sequence of measurable functions f_n **converges in measure** to a measurable function f if

$$\lim_{n\to\infty} \mu(\{x \,|\, |f_n(x) - f(x)| \ge \varepsilon\}) = 0 \quad \forall \varepsilon > 0. \tag{5.15}$$

In case of a probability measure this is also known as **convergence in probability**. One writes $f_n \overset{\mu}{\to} f$ in this case. There is also the concept of **local converges in measure**

$$\lim_{n\to\infty} \mu(\{x \in A \,|\, |f_n(x) - f(x)| \ge \varepsilon\}) = 0 \qquad \forall \varepsilon > 0,\ \mu(A) < \infty. \tag{5.16}$$

In case of a finite measure this must in particular hold for $A = X$ and both concepts agree.

Example 5.7. Consider $X := \mathbb{R}$ with Lebesgue measure and $f_n := \chi_{(n,\infty)}$. Then $f_n \to 0$ locally in measure since for every Borel set A with finite measure we have $(n,\infty) \cap A \searrow \emptyset$ and hence $\mu((n,\infty)\cap A) \to 0$ by continuity of measures from above. Clearly f_n does not converge in measure since $\mu(\{x| \, |f_n(x)| \geq \varepsilon\}) = \infty$ for $\varepsilon < 1$. ⋄

We first note a convergence result:

Theorem 5.7 (Bounded convergence theorem)**.** *Suppose μ is a finite measure and f_n is a bounded sequence of measurable functions which converges in measure to a measurable function f. Then*

$$\lim_{n\to\infty} \int f_n d\mu = \int f \, d\mu. \tag{5.17}$$

Proof. First of all note that $|f_n| \leq M$ together with convergence in measure implies $|f| \leq M$ a.e. Now fix $\varepsilon > 0$ and set $A_n = \{x| \, |f_n(x) - f(x)| \geq \varepsilon\}$, then

$$\begin{aligned}\left|\int (f_n - f) d\mu\right| = \int |f_n - f| d\mu = \int_{A_{n,\varepsilon}} |f_n - f| d\mu + \int_{X\setminus A_{n,\varepsilon}} |f_n - f| d\mu \\ \leq 2M\mu(A_n) + \varepsilon\mu(X).\end{aligned}$$

Since $\mu(A_n) \to 0$ by assumption and $\varepsilon > 0$ is arbitrary the claim follows. □

Our next aim is to investigate the connection with almost everywhere convergence. To this end define

$$\liminf_{n\to\infty} A_n := \bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} A_n, \qquad \limsup_{n\to\infty} A_n := \bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} A_n. \tag{5.18}$$

That is, $x \in \liminf A_n$ if $x \in A_n$ eventually and $x \in \limsup A_n$ if $x \in A_n$ infinitely often. In terms of indicator functions this could be equivalently defined using

$$\chi_{\liminf_{n\to\infty} A_n} = \liminf_{n\to\infty} \chi_{A_n}, \qquad \chi_{\limsup_{n\to\infty} A_n} = \limsup_{n\to\infty} \chi_{A_n}. \tag{5.19}$$

In particular, $\liminf A_n \subseteq \limsup A_n$.

Example 5.8. If we have $A_n \nearrow A$ or $A_n \searrow A$, then one easily checks $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = A$. However, note that if we consider $A_n = \{\frac{1}{n}\} \subset \mathbb{R}$, then we have $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \emptyset$. ⋄

Lemma 5.8. *Let f_n be a sequence of measurable functions and f measurable and consider the following statements:*

(i) *$f_n(x) \to f(x)$ for a.e. $x \in X$.*

(ii) *$\mu(\limsup_{n\to\infty}\{x| \, |f_n(x) - f(x)| \geq \varepsilon\}) = 0$ for all $\varepsilon > 0$.*

(iii) *$\lim_{k\to\infty} \mu(\bigcup_{n\geq k}\{x| \, |f_n(x) - f(x)| \geq \varepsilon\}) = 0$ for all $\varepsilon > 0$.*

Then (i) $\Leftrightarrow$ (ii) $\Leftarrow$ (iii). If μ is finite, then all above three items are equivalent.

Proof. Abbreviate $A_n(\varepsilon) := \{x| \, |f_n(x) - f(x)| \geq \varepsilon\}$ as well as $A(\varepsilon) := \limsup_{n\to\infty} A_n(\varepsilon)$. Moreover, note that

$$\begin{aligned} A(\varepsilon) &= \bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} A_n(\varepsilon) = \{x| \, \forall k \exists n \geq k : |f_n(x) - f(x)| \geq \varepsilon\} \\ &= \{x| \limsup_{n\to\infty} |f_n(x) - f(x)| \geq \varepsilon\} \end{aligned}$$

and hence $\{x|f_n(x) \not\to f(x)\} = \bigcup_{\varepsilon>0} A(\varepsilon) = \bigcup_n A(\frac{1}{n})$. Thus (i) $\Rightarrow$ (ii) and the converse follows since measures are continuous from below. Finally, (iii) $\Rightarrow$ (ii) by continuity from above and the converse also follows by continuity from above if we know that $\mu(\bigcup_{n\geq k}\{x| \, |f_n(x) - f(x)| \geq \varepsilon\}) < \infty$ for some k. □

Example 5.9. Consider $X := \mathbb{R}$ with Lebesgue measure and $f_n := \chi_{(n,\infty)}$. Then $f_n(x) \to f(x)$ for all $x \in \mathbb{R}$ but $\mu(\bigcup_{n\geq k}\{x| \, |f_n(x)| \geq \varepsilon\}) = \mu([k,\infty)) = \infty$ for all k and all $\varepsilon \leq 1$. Hence the finiteness assumption is crucial for (ii) $\Rightarrow$ (iii) to hold. ◇

In particular, (iii) implies (5.15) and hence

Corollary 5.9. *Convergence a.e. implies local convergence in measure.*

The above example clearly shows that we do not get convergence in measure unless μ is finite. The converse is wrong even in the finite case:

Example 5.10. Let $X := [0,1]$ with Lebesgue measure and choose $f_n := \chi_{I_{m,k}}$ the characteristic functions of the intervals $I_{m,k} := [k2^{-m}, (k+1)2^{-m}]$ with m,k defined as the unique decomposition of $n = 2^m + k$ with $0 \leq m$ and $0 \leq k < 2^m$. Then $\lambda^1(\{x| \, |f_n(x)| \geq \varepsilon\}) = 0$ for $\varepsilon > 1$ and $\lambda^1(\{x| \, |f_n(x)| \geq \varepsilon\}) = 2^{-m}$ for $1 \geq \varepsilon > 0$. Hence in both cases $\lambda^1(\{x| \, |f_n(x)| \geq \varepsilon\}) \to 0$ as $n \to \infty$ (which implies $m = \lfloor \log_2(n) \rfloor \to \infty$). Hence $f_n \overset{\mu}{\to} f$ in measure. However, for every $x \in [0,1]$ and for every $m \in \mathbb{N}_0$ we can find a corresponding k_0 such that $x \in I_{m,k_0}$ and $x \notin I_{m,k}$ for $|k - k_0| > 1$. Hence $\liminf_{n\to\infty} f_n(x) = 0$ and $\limsup_{n\to\infty} f_n(x) = 1$. ◇

A useful criterion is given by the

Theorem 5.10 (Borel–Cantelli lemma)**.** *Suppose* $\sum_n \mu(A_n) < \infty$, *then* $\mu(\limsup_n A_n) = 0$.

Proof. This follows from

$$\mu(\limsup_n A_n) = \mu\Big(\bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} A_n\Big) \leq \inf_{k\in\mathbb{N}} \mu\Big(\bigcup_{n\geq k} A_n\Big) \leq \inf_{k\in\mathbb{N}} \sum_{n\geq k} \mu(A_n) = 0. \quad \square$$

Example 5.11. The example $A_n = [0, \frac{1}{n}]$ with Lebesgue measure on $[0,1]$ shows that the converse might fail. ◇

Corollary 5.11. *If $\sum_n \mu(\{x| \, |f_n(x) - f(x)| \geq \varepsilon\}) < \infty$ for all $\varepsilon > 0$, then $f_n(x) \to f(x)$ for a.e. x.*

Corollary 5.12. . *If $f_n \overset{\mu}{\to} f$ in measure, then there is a subsequence which converges a.e. Conversely, if from every subsequence of f_n we can choose a subsequence which converges to f a.e., then $f_n \overset{\mu}{\to} f$ in measure.*

Proof. If $f_n \overset{\mu}{\to} f$ we can find a subsequence n_k such that $\mu(\{x| \, |f_{n_k}(x) - f(x)| \geq \frac{1}{k}\}) \leq \frac{1}{k^2}$ and the first claim follows from the previous corollary.

Conversely, if f_n does not converge to f in measure, we can find $\varepsilon, \delta > 0$ and a subsequence n_k such that $\mu(\{x| \, |f_{n_k}(x) - f(x)| \geq \varepsilon\}) \geq \delta$ for all k. But from this sequence we cannot choose an a.e. convergent subsequence. □

Problem 5.5. *Let $\alpha, \beta \in \mathbb{C}$ and suppose $f_n(x) \to f(x)$ and $g_n(x) \to g(x)$ for a.e. Show that $\alpha f_n(x) + \beta g_n(x) \to \alpha f(x) + \beta g(x)$ as well as $f_n(x) g_n(x) \to f(x) g(x)$ a.e.*

Problem 5.6. *Let $\alpha, \beta \in \mathbb{C}$ and suppose $f_n \overset{\mu}{\to} f$ and $g_n \overset{\mu}{\to} g$. Show that $\alpha f_n + \beta g_n \overset{\mu}{\to} \alpha f + \beta g$. Show that $f_n(x) g_n(x) \overset{\mu}{\to} f(x) g(x)$ if either both sequences are bounded or μ is finite. (Hint: In the finite case one can use the previous problem and Corollary 5.12.)*

Problem 5.7. *Show that L^p convergence implies convergence in measure. Show that the converse fails. (See Problem 9.9 for an improvement.)*

Problem 5.8 (Fatou for sets)**.** *Show*

$$\mu(\liminf_n A_n) \leq \liminf_n \mu(A_n)$$

and

$$\limsup_n \mu(A_n) \leq \mu(\limsup_n A_n), \qquad \text{if} \quad \mu(\bigcup_n A_n) < \infty.$$

(Hint: Show $\liminf_n \chi_{A_n} = \chi_{\liminf_n A_n}$ and $\limsup_n \chi_{A_n} = \chi_{\limsup_n A_n}$.)

Problem 5.9. *Define the* ***Ky Fan metric***

$$d(f,g) := \min\{\varepsilon \geq 0 | \mu(\{x| \, |f(x) - g(x)| > \varepsilon\}) \leq \varepsilon\}.$$

Show:

(i) *The minimum in the definition is attained.*
(ii) *$d(f,g) = 0$ if and only if $f = g$ a.e.*
(iii) *d satisfies the triangle inequality: $d(f,h) \leq d(f,g) + d(g,h)$.*
(iv) *$f_n \overset{\mu}{\to} f$ in measure $\Leftrightarrow d(f_n, f) \to 0$.*

(Hint for (iii): Start with $\mu(\{|f-h| > d(f,g) + d(g,h)\}) \leq \mu(\{|f-g| > d(f,g)\}) + \mu(\dots) \leq \dots$)

Problem 5.10. *Suppose μ is σ-finite and choose $X_n \nearrow X$ with finite $\mu(X_n) > 0$. Show that*

$$d(f,g) := \sum_{n\in\mathbb{N}} \frac{1}{2^n \mu(X_n)} \int_{X_n} \frac{|f-g|}{1+|f-g|} d\mu$$

is a metric(if functions which are equal a.e. are identified) which induces local convergence in measure.

Problem 5.11. *Consider $X := [0,1]$ with Lebesgue measure. Show that a.e. convergence does not stem from a topology. (Hint: Recall that in a topological space a sequence x_n converges to x if and only if every subsequence has a subsequence which converges to x — Lemma B.5 from* [**22**]*.)*

5.4. Weak and vague convergence of measures

In this section X will be a metric space equipped with the Borel σ-algebra. We say that a sequence of finite Borel measures μ_n **converges weakly** to a finite Borel measure μ if

$$\int_X f d\mu_n \to \int_X f d\mu, \qquad \forall f \in C_b(X). \tag{5.20}$$

Since by the Riesz representation theorem the set of (complex) measures is the dual of $C(X)$ (under appropriate assumptions on X), this is what would be denoted by weak-$*$ convergence in functional analysis. However, we will not need this connection here. Nevertheless we remark that the weak limit is unique. To see this let C be a nonempty closed set and consider

$$f_n(x) := \max(0, 1 - n\,\mathrm{dist}(x, C)). \tag{5.21}$$

Then $f \in C_b(X)$, $f_n \downarrow \chi_C$, and (dominated convergence)

$$\mu(C) = \lim_{n\to\infty} \int_X g_n d\mu \tag{5.22}$$

shows that μ is uniquely determined by the integral for continuous functions (recall Lemma 1.21 and its corollary). For later use observe that f_n is even Lipschitz continuous, $|f_n(x) - f_n(y)| \le n\,d(x,y)$ (cf. Lemma B.25 from [**22**]).

Moreover, choosing $f \equiv 1$ shows

$$\mu(X) = \lim_{n\to\infty} \mu_n(X). \tag{5.23}$$

However, this might not be true for arbitrary sets in general. To see this look at the case $X = \mathbb{R}$ with $\mu_n = \delta_{1/n}$. Then μ_n converges weakly to $\mu = \delta_0$. But $\mu_n((0,1)) = 1 \not\to 0 = \mu((0,1))$ as well as $\mu_n([-1,0]) = 0 \not\to 1 = \mu([-1,0])$. So mass can appear/disappear at boundary points but this is the worst that can happen:

Theorem 5.13 (Portmanteau)**.** *Let X be a metric space and μ_n, μ finite Borel measures. The following are equivalent:*

(i) μ_n *converges weakly to* μ.

(ii) $\int_X f\, d\mu_n \to \int_X f\, d\mu$ *for every bounded Lipschitz continuous* f.

(iii) $\limsup_n \mu_n(C) \le \mu(C)$ *for all closed sets* C *and* $\mu_n(X) \to \mu(X)$.

(iv) $\liminf_n \mu_n(O) \ge \mu(O)$ *for all open sets* O *and* $\mu_n(X) \to \mu(X)$.

(v) $\mu_n(A) \to \mu(A)$ *for all Borel sets* A *with* $\mu(\partial A) = 0$.

(vi) $\int_X f\, d\mu_n \to \int_X f\, d\mu$ *for all bounded functions* f *which are continuous at* μ*-a.e.* $x \in X$.

Proof. (i) $\Rightarrow$ (ii) is trivial. (ii) $\Rightarrow$ (iii): Define f_n as in (5.21) and start by observing

$$\limsup_n \mu_n(C) = \limsup_n \int \chi_F \mu_n \le \limsup_n \int f_m \mu_n = \int f_m \mu.$$

Now taking $m \to \infty$ establishes the claim. Moreover, $\mu_n(X) \to \mu(X)$ follows choosing $f \equiv 1$. (iii) $\Leftrightarrow$ (iv): Just use $O = X \setminus C$ and $\mu_n(O) = \mu_n(X) - \mu_n(C)$. (iii) and (iv) $\Rightarrow$ (v): By $A^\circ \subseteq A \subseteq \overline{A}$ we have

$$\begin{aligned}\limsup_n \mu_n(A) &\le \limsup_n \mu_n(\overline{A}) \le \mu(\overline{A}) \\ &= \mu(A^\circ) \le \liminf_n \mu_n(A^\circ) \le \liminf_n \mu_n(A)\end{aligned}$$

provided $\mu(\partial A) = 0$. (v) $\Rightarrow$ (vi): By considering real and imaginary parts separately we can assume f to be real-valued. Moreover, adding an appropriate constant we can even assume $0 \le f_n \le M$. Set $A_r = \{x \in X | f(x) > r\}$ and denote by D_f the set of discontinuities of f. Then $\partial A_r \subseteq D_f \cup \{x \in X | f(x) = r\}$. Now the first set has measure zero by assumption and the second set is countable by Problem 2.20. Thus the set of all r with $\mu(\partial A_r) > 0$ is countable and thus of Lebesgue measure zero. Then by Problem 2.20 (with $\phi(r) = 1$) and dominated convergence

$$\int_X f\, d\mu_n = \int_0^M \mu_n(A_r) dr \to \int_0^M \mu(A_r) dr = \int_X f\, d\mu$$

Finally, (vi) $\Rightarrow$ (i) is trivial. □

Next we want to extend our considerations to unbounded measures. In this case boundedness of f will not be sufficient to guarantee integrability and hence we will require f to have compact support. If $x \in X$ is such that $f(x) \ne 0$, then $f^{-1}(B_r(f(x))$ will be a relatively compact neighborhood of x whenever $0 < r < |f(x)|$. Hence $C_b(X)$ will not have sufficiently many functions with compact support unless we assume X to be locally compact, which we will do for the remainder of this section.

Let μ_n be a sequence of Borel measures on a locally compact metric space X. We will say that μ_n **converges vaguely** to a Borel measure μ if

$$\int_X f\, d\mu_n \to \int_X f\, d\mu, \qquad \forall f \in C_c(X). \tag{5.24}$$

As with weak convergence (cf. Problem 6.6) we can conclude that the integral over functions with compact supports determines $\mu(K)$ for every compact set. Hence the vague limit will be unique if μ is inner regular (which we already know to always hold if X is locally compact and separable by Corollary 1.23).

We first investigate the connection with weak convergence.

Lemma 5.14. *Let X be a locally compact separable metric space and suppose $\mu_m \to \mu$ vaguely. Then $\mu(X) \le \liminf_n \mu_n(X)$ and (5.24) holds for every $f \in C_0(X)$ in case $\mu_n(X) \le M$. If in addition $\mu_n(X) \to \mu(X)$, then (5.24) holds for every $f \in C_b(X)$, that is, $\mu_n \to \mu$ weakly.*

Proof. For every compact set K we can find a nonnegative function $g \in C_c(X)$ which is one on K by Urysohn's lemma. Hence $\mu(K) \le \int g d\mu = \lim_n \int g d\mu_n \le \liminf_n \mu_n(X)$. Letting $K \nearrow X$ shows $\mu(X) \le \liminf_n \mu_n(X)$. Next, let $f \in C_0(X)$ and fix $\varepsilon > 0$. Then there is a compact set K such that $|f(x)| \le \varepsilon$ for $x \in X \setminus K$. Choose g for K as before and set $f = f_1 + f_2$ with $f_1 = gf$. Then $|\int f d\mu - \int f d\mu_n| \le |\int f_1 d\mu - \int f_1 d\mu_n| + 2\varepsilon M$ and the first claim follows.

Similarly, for the second claim, let $|f| \le C$ and choose a compact set K_0 with a corresponding function g_0 such that $\mu(X \backslash K_0) < \varepsilon$. Then $\limsup_n \mu_n(X \setminus K) \le \lim_n \int (1 - g_0) d\mu_n = \int (1 - g_0) d\mu \le \mu(X \setminus K_0)$, where $K = \operatorname{supp}(g_0)$. Hence we have $\mu_n(X \backslash K) < \varepsilon$ for $n \ge N$. Choose g for K as before and set $f = f_1 + f_2$ with $f_1 = gf$. Then $|\int f d\mu - \int f d\mu_n| \le |\int f_1 d\mu - \int f_1 d\mu_n| + 2\varepsilon C$ and the second claim follows. □

Example 5.12. The example $X = \mathbb{R}$ with $\mu_n = \delta_n$ shows that in the first claim f cannot be replaced by a bounded continuous function. Moreover, the example $\mu_n = n\,\delta_n$ also shows that the uniform bound cannot be dropped. ⋄

The analog of Theorem 5.13 reads as follows.

Theorem 5.15. *Let X be a locally compact metric space and μ_n, μ Borel measures. The following are equivalent:*

(i) *μ_n converges vagly to μ.*

(ii) *$\int_X f\, d\mu_n \to \int_X f\, d\mu$ for every Lipschitz continuous f with compact support.*

(iii) *$\limsup_n \mu_n(C) \le \mu(C)$ for all compact sets K and $\liminf_n \mu_n(O) \ge \mu(O)$ for all relatively compact open sets O.*

(iv) $\mu_n(A) \to \mu(A)$ *for all relative compact Borel sets* A *with* $\mu(\partial A) = 0$.

(v) $\int_X f\, d\mu_n \to \int_X f\, d\mu$ *for all bounded functions* f *with compact support which are continuous at* μ*-a.e.* $x \in X$.

Proof. (i) $\Rightarrow$ (ii) is trivial. (ii) $\Rightarrow$ (iii): The claim for compact sets follows as in the proof of Theorem 5.13. To see the case of open sets let $K_n = \{x \in X | \operatorname{dist}(x, X \setminus O) \geq n^{-1}\}$. Then $K_n \subseteq O$ is compact and we can look at

$$g_n(x) := \frac{\operatorname{dist}(x, X \setminus O)}{\operatorname{dist}(x, X \setminus O) + \operatorname{dist}(x, K_n)}.$$

Then g_n is supported on O and $g_n \nearrow \chi_O$. Then

$$\liminf_n \mu_n(O) = \liminf_n \int \chi_O \mu_n \geq \liminf_n \int g_m \mu_n = \int g_m \mu.$$

Now taking $m \to \infty$ establishes the claim.

The remaining directions follow literally as in the proof of Theorem 5.13 (concerning (iv) $\Rightarrow$ (v) note that A_r is relatively compact). □

Finally we look at Borel measures on $\mathbb{R}$. In this case, we have the following equivalent characterization of vague convergence.

Lemma 5.16. *Let* μ_n *be a sequence of Borel measures on* $\mathbb{R}$. *Then* $\mu_n \to \mu$ *vaguely if and only if the distribution functions (normalized at a point of continuity of* μ*) converge at every point of continuity of* μ.

Proof. Suppose $\mu_n \to \mu$ vaguely. Then the distribution functions converge at every point of continuity of μ by item (iv) of Theorem 5.15.

Conversely, suppose that the distribution functions converge at every point of continuity of μ. To see that in fact $\mu_n \to \mu$ vaguely, let $f \in C_c(\mathbb{R})$. Fix some $\varepsilon > 0$ and note that, since f is uniformly continuous, there is a $\delta > 0$ such that $|f(x) - f(y)| \leq \varepsilon$ whenever $|x - y| \leq \delta$. Next, choose some points $x_0 < x_1 < \cdots < x_k$ such that $\operatorname{supp}(f) \subset (x_0, x_k)$, μ is continuous at x_j, and $x_j - x_{j-1} \leq \delta$ (recall that a monotone function has at most countable discontinuities). Furthermore, there is some N such that $|\mu_n(x_j) - \mu(x_j)| \leq$

$\frac{\varepsilon}{2k}$ for all j and $n \geq N$. Then

$$\begin{aligned}\left|\int f d\mu_n - \int f d\mu\right| \leq & \sum_{j=1}^{k} \int_{(x_{j-1},x_j]} |f(x) - f(x_j)| d\mu_n(x) \\ & + \sum_{j=1}^{k} |f(x_j)||\mu((x_{j-1},x_j]) - \mu_n((x_{j-1},x_j])| \\ & + \sum_{j=1}^{k} \int_{(x_{j-1},x_j]} |f(x) - f(x_j)| d\mu(x).\end{aligned}$$

Now, for $n \geq N$, the first and the last terms on the right-hand side are both bounded by $(\mu((x_0,x_k]) + \frac{\varepsilon}{k})\varepsilon$ and the middle term is bounded by $\max|f|\varepsilon$. Thus the claim follows. □

Moreover, every bounded sequence of measures has a vaguely convergent subsequence (this is a special case of **Helly's selection theorem** — a generalization will be provided in Theorem 6.11).

Lemma 5.17. *Suppose μ_n is a sequence of finite Borel measures on $\mathbb{R}$ such that $\mu_n(\mathbb{R}) \leq M$. Then there exists a subsequence n_j which converges vaguely to some measure μ with $\mu(\mathbb{R}) \leq \liminf_j \mu_{n_j}(\mathbb{R})$.*

Proof. Let $\mu_n(x) = \mu_n((-\infty,x])$ be the corresponding distribution functions. By $0 \leq \mu_n(x) \leq M$ there is a convergent subsequence for fixed x. Moreover, by the standard diagonal series trick (enumerate the rationals, extract a subsequence which converges for the first rational, form this subsequence extract another one which converges also on the second rational, etc.; finally choose the diagonal sequence, i.e., the fisrt element form the first subsequence, the second element from the second subsequence, etc.), we can assume that $\mu_n(y)$ converges to some number $\mu(y)$ for each rational y. For irrational x we set $\mu(x) = \inf\{\mu(y) | x < y \in \mathbb{Q}\}$. Then $\mu(x)$ is monotone, $0 \leq \mu(x_1) \leq \mu(x_2) \leq M$ for $x_1 < x_2$. Indeed for $x_1 \leq y_1 < x_2 \leq y_2$ with $y_j \in \mathbb{Q}$ we have $\mu(x_1) \leq \mu(y_1) = \lim_n \mu_n(y_1) \leq \lim_n \mu_n(y_2) = \mu(y_2)$. Taking the infimum over y_2 gives the result.

Furthermore, for every $\varepsilon > 0$ and $x - \varepsilon < y_1 \leq x \leq y_2 < x + \varepsilon$ with $y_j \in \mathbb{Q}$ we have

$$\mu(x-\varepsilon) \leq \lim_n \mu_n(y_1) \leq \liminf \mu_n(x) \leq \limsup \mu_n(x) \leq \lim_n \mu_n(y_2) \leq \mu(x+\varepsilon)$$

and thus

$$\mu(x-) \leq \liminf \mu_n(x) \leq \limsup \mu_n(x) \leq \mu(x+)$$

which shows that $\mu_n(x) \to \mu(x)$ at every point of continuity of μ. So we can redefine μ to be right continuous without changing this last fact. The bound for $\mu(\mathbb{R})$ follows since for every point of continuity x of μ we have $\mu(x) = \lim_n \mu_n(x) \le \liminf_n \mu_n(\mathbb{R})$. □

Example 5.13. The example $d\mu_n(x) = d\Theta(x-n)$ for which $\mu_n(x) = \Theta(x-n) \to 0$ shows that we can have $\mu(\mathbb{R}) = 0 < 1 = \mu_n(\mathbb{R})$. ◇

A common assumption to exclude the phenomena from the previous example is tightness, see Problem 5.15.

Problem 5.12. *Suppose $\mu_n \to \mu$ vaguely and let I be a bounded interval with boundary points x_0 and x_1. Then*

$$\limsup_n \left| \int_I f d\mu_n - \int_I f d\mu \right| \le |f(x_1)|\mu(\{x_1\}) + |f(x_0)|\mu(\{x_0\})$$

for any $f \in C([x_0, x_1])$.

Problem 5.13. *Let $\mu_n(X) \le M$ and suppose* (5.24) *holds for all $f \in U \subseteq C(X)$. Then* (5.24) *holds for all f in the closed linear span of U.*

Problem 5.14. *Let $\mu_n(\mathbb{R}), \mu(\mathbb{R}) \le M$ and suppose the Cauchy transforms converge*

$$\int_{\mathbb{R}} \frac{1}{x-z} d\mu_n(x) \to \int_{\mathbb{R}} \frac{1}{x-z} d\mu(x)$$

for $z \in U$, where $U \subseteq \mathbb{C}\backslash\mathbb{R}$ is a set which has a limit point. Then $\mu_n \to \mu$ vaguely. (Hint: Problem B.66 from **[22]***.)*

Problem 5.15. *Let X be a metric space. A sequence of finite measures is called **tight** if for every $\varepsilon > 0$ there is a compact set $K \subseteq X$ such that* $\sup_n \mu_n(X \setminus K) \le \varepsilon$. *Show that if a vaguely convergent sequence is tight, then $\mu_n(X) \to \mu(X)$. Show that the converse holds if X is σ-compact.*

Problem 5.16. *Let ϕ be a mollifier and μ a finite measure. Show that $\phi_\varepsilon * \mu$ converges weakly to μ. Here $\phi * \mu$ is the absolutely continuous measure with density $\int_{\mathbb{R}} \phi(x-y) d\mu(y)$ (cf. Problem 3.34).*

5.5. The Bochner integral

In this section we want to extend the Lebesgue integral to the case of functions with values in a normed space. This extension is known as Bochner integral. Since a normed space has no order we cannot use monotonicity and hence are restricted to finite values for the integral. Other than that, we only need some small adaptions.

Let (X, Σ, μ) be a measure space and Y a Banach space equipped with the Borel σ-algebra $\mathfrak{B}(Y)$. Note that if f, g are measurable then so is any linear combination (Problem 5.17) and also any composition with a continuous function. Moreover, the limit of measurable functions is again measurable (Problem 5.18).

As in (2.1), a measurable function $s : X \to Y$ is called **simple** if its image is finite; that is, if

$$s = \sum_{j=1}^{p} \alpha_j \, \chi_{A_j}, \qquad \mathrm{Ran}(s) =: \{\alpha_j\}_{j=1}^{p}, \quad A_j := s^{-1}(\alpha_j) \in \Sigma. \tag{5.25}$$

Also the integral of a simple function can be defined as in (2.2) provided we ensure that it is finite. To this end we call s **integrable** if $\mu(A_j) < \infty$ for all j with $\alpha_j \neq 0$. Now, for an integrable simple function s as in (5.25) we define its **integral** as

$$\int_A s \, d\mu := \sum_{j=1}^{p} \alpha_j \, \mu(A_j \cap A). \tag{5.26}$$

As before we use the convention $0 \cdot \infty = 0$ (where 0 is the zero vector from Y).

Lemma 5.18. *The integral has the following properties:*

(i) $\int_A s \, d\mu = \int_X \chi_A \, s \, d\mu$.

(ii) $\int_{\bigcup_{n=1}^\infty A_n} s \, d\mu = \sum_{n=1}^\infty \int_{A_n} s \, d\mu$.

(iii) $\int_A \alpha \, s \, d\mu = \alpha \int_A s \, d\mu$, $\alpha \in \mathbb{C}$.

(iv) $\int_A (s + t) d\mu = \int_A s \, d\mu + \int_A t \, d\mu$.

(v) $\| \int_A s \, d\mu \| \le \int_A \|s\| d\mu$.

Proof. The first four items follow literally as in Lemma 2.1. (v) follows from the triangle inequality. □

Now we extend the integral via approximation by simple functions. However, while a complex-valued measurable function f can always be approximated by simple functions s_n, this might no longer be true in our present setting. In fact, note that a sequence of simple functions s_n involves only a countable number of values from Y and since the limit must be in the closure of the span of these values, $\mathrm{Ran}(f) \subseteq \overline{\bigcup_n \mathrm{Ran}(s_n)}$, the range of f must be separable. Moreover, we also need to ensure finiteness of the integral.

If μ is finite, the latter requirement is easily satisfied by considering only bounded functions. Consequently we could equip the space of integrable simple functions $S(X, \mu, Y)$ with the supremum norm $\|s\|_\infty = \sup_{x \in X} \|s(x)\|$ and use the fact that the integral is a bounded linear functional,

$$\| \int_A s \, d\mu \| \le \mu(A) \|s\|_\infty, \tag{5.27}$$

to extend it to the completion of the simple functions, known as the regulated functions $R(X, \mu, Y)$. Hence the integrable functions will be the bounded functions which are uniform limits of integrable simple functions. Note that the range of a regulated function will even be totally bounded and hence relatively compact. This, together with measurability, characterizes regulated functions (see Lemma 5.20 below). While this gives a theory suitable for many cases, we want to do better and look at functions which are pointwise limits of simple functions.

Consequently, we call a function f **integrable** if there is a sequence of integrable simple functions s_n which converges pointwise to f such that

$$\int_X \|f - s_n\| d\mu \to 0. \tag{5.28}$$

In this case item (v) from Lemma 5.18 ensures that $\int_X s_n d\mu$ is a Cauchy sequence and we can define the **Bochner integral** of f to be

$$\int_X f\, d\mu := \lim_{n\to\infty} \int_X s_n d\mu. \tag{5.29}$$

If there are two sequences of integrable simple functions as in the definition, we could combine them into one sequence (taking one as even and the other one as odd elements) to conclude that the limit of the first two sequences equals the limit of the third sequence. In other words, the definition of the integral is independent of the sequence chosen. The integrable functions will be denoted by $\mathcal{L}^1(X, d\mu, Y)$.

Lemma 5.19. *The integral is linear and Lemma 5.18 holds for integrable functions s, t.*

Proof. All items except for (ii) are immediate. (ii) is also immediate for finite unions. The general case will follow from the dominated convergence theorem to be shown below. □

Before we proceed, we try to shed some light on when a function is integrable.

Lemma 5.20. *A function $f : X \to Y$ is the pointwise limit of simple functions if and only if it is measurable and its range is separable. It is the uniform limit of simple functions if and only if it is measurable and its range is relatively compact. Moreover, the sequence s_n can be chosen such that $\|s_n(x)\| \le 2\|f(x)\|$ for every $x \in X$ and $\mathrm{Ran}(s_n) \subseteq \mathrm{Ran}(f) \cup \{0\}$.*

Proof. It remains to establish the converse. Let $\{y_j\}_{j\in\mathbb{N}}$ be a dense set for the range. Note that the balls $B_{1/m}(y_j)$ will cover the range for every fixed $m \in \mathbb{N}$. Furthermore, we will augment this set by $y_0 = 0$ to make sure that any value less than $1/m$ is replaced by 0 (since otherwise one might destroy decay properties of f). By iteratively removing what has already been covered we get a disjoint cover $A_j^m \subseteq B_{1/m}(y_j)$ (some sets might be empty) such that $A_{n,m} := \dot{\bigcup}_{j\le n} A_j^m = \bigcup_{j\le n} B_{1/m}(y_j)$. Now consider the simple function s_n defined as follows

$$s_n(x) = \begin{cases} y_j, & \text{if } f(x) \in A_j^m \setminus \bigcup_{m<k\le n} A_{n,k}, \\ 0, & \text{else.} \end{cases}$$

That is, we first search for the largest $m \le n$ with $f(x) \in A_{n,m}$ and look for the unique j with $f(x) \in A_j^m$ (i.e., the first j with $f(x) \in B_{1/m}(y_j)$). If such an m exists we have $s_n(x) = y_j$ and otherwise $s_n(x) = 0$ (i.e., if $x \notin A_{n,1}$). By construction s_n is measurable and $\|f(x) - s_n(s)\| < \frac{1}{m}$ for $x \in A_{n,m}$ with $m \le n$. Hence s_n converges pointwise to f. Moreover, to see $\|s_n(x)\| \le 2\|f(x)\|$ we consider two cases. If $s_n(x) = 0$ then the claim is trivial. If $s_n(x) \ne 0$ then $f(x) \in A_j^m$ with $j > 0$ and hence $\|f(x)\| \ge 1/m$ implying

$$\|s_n(x)\| = \|y_j\| \le \|y_j - f(x)\| + \|f(x)\| \le \frac{1}{m} + \|f(x)\| \le 2\|f(x)\|.$$

If the range of f is relatively compact, then for every m there is an n such that $A_{n,m}$ covers the whole range. Hence $\|f(x) - s_n(x)\| < \frac{1}{m}$ for all x. Conversely, note that the range of a uniform limit of simple functions is totally bounded (i.e. it has a finite ε-cover for every $\varepsilon > 0$) and hence relatively compact. □

Functions $f : X \to Y$ which are the pointwise limit of simple functions are also called **strongly measurable**. Notice that we do not require the simple functions to be integrable! Some authors do.

Since the limit of measurable functions is again measurable (Problem 5.18) and still has separable range, the limit of strongly measurable functions is strongly measurable. Moreover, our lemma also shows that continuous functions on separable spaces are strongly measurable (recall Lemma 1.18 and the fact that the continuous image of a separable set is separable; cf. Problem B.29 from [**22**]). Another fact worth mentioning is, that if Y_0 is the closure of the linear span of $\mathrm{Ran}(f)$, then restricting Y to Y_0 does neither affect integrability nor the integral. In particular, the integral of f is in Y_0.

Now we are in the position to give a useful characterization of integrability.

Lemma 5.21 (Bochner)**.** *A function $f : X \to Y$ is integrable if and only if it is strongly measurable and $\|f\|$ is integrable.*

Proof. We have already seen that an integrable function has the stated properties. Conversely, the sequence s_n from the previous lemma is integrable (note $\|s_n(x)\| = \sum_j \|\alpha_j^n\| \chi_{A_j^n}(x) \le 2\|f(x)\|$) and satisfies (5.28) by the dominated convergence theorem. □

Another useful observation is the fact that the integral behaves nicely under linear transforms. Recall that a linear operator $A : \mathfrak{D}(A) \subseteq Y \to Z$ defined on a domain $\mathfrak{D}(A)$ is is called closed, if its graph $\{(x, Ax)|x \in \mathfrak{D}(A)\}$ is a closed subset of $Y \times Z$.

Theorem 5.22 (Hille). *Let Y, Z be Banach spaces with their respective Borel σ-algebras. Let $A : \mathfrak{D}(A) \subseteq Y \to Z$ be a closed linear operator. Suppose $f : X \to Y$ is integrable with* $\mathrm{Ran}(f) \subseteq \mathfrak{D}(A)$ *and $Af : X \to Z$ is also integrable. Then*

$$A \int_X f\,d\mu = \int_X (Af)d\mu, \qquad f \in \mathfrak{D}(A). \tag{5.30}$$

If $A \in \mathscr{L}(Y, Z)$, then f integrable implies Af is integrable.

Proof. By assumption $(f, Af) : X \to Y \times Z$ is integrable and by the proof of Lemma 5.21 the sequence of simple functions can be chosen to have its range in the graph of A. In other words, there exists a sequence of simple functions (s_n, As_n) such that

$$\int_X s_n d\mu \to \int_X f\,d\mu, \qquad A\int_X s_n d\mu = \int_X As_n d\mu \to \int_X Af\,d\mu.$$

Hence the claim follows from closedness of A.

If A is bounded and s_n is a sequence of simple functions satisfying (5.28), then $t_n = As_n$ is a corresponding sequence of simple functions for Af since $\|t_n - Af\| \le \|A\|\|s_n - f\|$. This shows the last claim. □

Clearly the assumptions of this theorem are satisfied if $A \in \mathscr{L}(Y, Z)$, for example, if we choose a bounded linear functional from Y^*.

Next, we note that the dominated convergence theorem holds for the Bochner integral.

Theorem 5.23. *Suppose f_n are integrable and $f_n \to f$ pointwise. Moreover, suppose there is nonnegative integrable function g with $\|f_n(x)\| \le g(x)$. Then f is integrable and*

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f\,d\mu \tag{5.31}$$

Proof. The pointwise limit f is strongly measurable and by $\|f(x)\| \le g(x)$ it is even integrable. Moreover, the usual dominated convergence theorem shows $\int_X \|f_n - f\|d\mu \to 0$ from which the claim follows. □

There is also yet another useful characterization of strong measurability. To this end we call a function $f : X \to Y$ **weakly measurable** if $\ell \circ f$ is measurable for every $\ell \in Y^*$.

Theorem 5.24 (Pettis). *A function $f : X \to Y$ is strongly measurable if and only if it is weakly measurable and its range is separable.*

Proof. Since every measurable function is weakly measurable, Lemma 5.20 shows one direction. To see the converse direction let $\{y_k\}_{k\in\mathbb{N}} \subseteq f(X)$ be dense. Define $s_n : \overline{f(X)} \to \{y_1, \dots, y_n\}$ via $s_n(y) = y_k$, where $k = k_{y,n}$ is the smallest k such that $\|y - y_k\| = \min_{1\le j\le n} \|y - y_j\|$. By density we have $\lim_{n\to\infty} s_n(y) = y$ for every $y \in \overline{f(X)}$. Now set $f_n = s_n \circ f$ and note that $f_n \to f$ pointwise. Moreover, for $1 \le k \le n$ we have

$$\begin{aligned} f_n^{-1}(y_k) =& \{x \in X | \|f(x) - y_k\| = \min_{1\le j\le n} \|f(x) - y_j\|\} \cap \\ & \{x \in X | \|f(x) - y_l\| < \min_{1\le j\le n} \|f(x) - y_j\|, 1 \le l < k\} \end{aligned}$$

and f_n will be measurable once we show that $\|f - y\|$ is measurable for every $y \in \overline{f(Y)}$. To show this last claim choose (Problem 5.21) a countable set $\{y_k'\} \in Y^*$ of unit vectors such that $\|y\| = \sup_k y_k'(y)$. Then $\|f - y\| = \sup_k y'(f - y)$ from which the claim follows. □

Finally, there is also a version of Fubini:

Theorem 5.25 (Fubini). *Let f be a strongly measurable function on $X_1 \times X_2$ and let μ_1, μ_2 be σ-finite measures on X_1, X_2, respectively.*

Then

$$\int_{X_1} \|f(x_1, x_2)\| d\mu_1(x_1) \in \mathcal{L}^1(X_2, d\mu_2), \tag{5.32}$$

respectively,

$$\int_{X_2} \|f(x_1, x_2)\| d\mu_2(x_2) \in \mathcal{L}^1(X_1, d\mu_1), \tag{5.33}$$

if and only if $f \in \mathcal{L}^1(X_1 \times X_2, d\mu_1 \otimes d\mu_2, Y)$. In this case

$$\begin{aligned} \iint_{X_1\times X_2} f(x_1, x_2) d\mu_1 \otimes \mu_2(x_1, x_2) &= \int_{X_2} \left(\int_{X_1} f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2) \\ &= \int_{X_1} \left(\int_{X_2} f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1). \end{aligned} \tag{5.34}$$

Proof. The first claim is Tonelli's theorem applied to $\|f\|$. Moreover, by Theorem 2.11 and linearity the last claim holds for simple functions. Now choose a sequences of simple functions as in Lemma 5.20 and apply dominated convergence (twice for the iterated integrals). □

Finally note that since the integral does not see null sets, one could also work with functions which satisfy the above requirements only away from null sets.

Also note that many results can be reduced to the classical ones by applying linear functionals. One example is the **fundamental lemma of the calculus of variations** (cf. Lemma 3.23)

Lemma 5.26. *Suppose $X \subseteq \mathbb{R}^n$ is open and $f : X \to Y$ is locally integrable. Then*

$$\int_X \varphi(x) f(x) d^n x = 0, \qquad \forall \varphi \in C_c^\infty(X),\ \varphi \geq 0, \tag{5.35}$$

if and only if $f(x) = 0$ (a.e.).

Proof. Note that by Problem 5.17 the integrand φf is measurable. Moreover, by $\|\varphi(x)f(x)\| = |\varphi(x)|\|f(x)\|$ and $\operatorname{Ran}(\varphi f) \subseteq Y_0$, where $Y_0 := \overline{\operatorname{span}\operatorname{Ran}(f)}$, we conclude that φf is integrable. In particular, we can replace Y by Y_0 and assume that Y is separable without loss of generality. Now choose a countable set of linear functionals $\ell_k \in Y^*$ as in Problem 5.21. Then

$$\int_X \varphi(x) \ell_k(f(x)) d^n x = \ell_k\left(\int_X \varphi(x) f(x) d^n x\right) = 0$$

for all k, which implies $\ell_k(f(x)) = 0$ away from some null set N_k. Hence $\|f(x)\| = \sup_k |\ell_k(f(x))| = 0$ away from the null set $\bigcup_k N_k$. □

Another result that continues to hold (with literally the same proof) is the Lebesgue differentiation theorem (Theorem 4.6 and Lemma 4.7):

Theorem 5.27. *Lebesgue Let $f : \mathbb{R}^n \to Y$ be (locally) integrable, then for a.e. $x \in \mathbb{R}^n$ we have*

$$\lim_{r \downarrow 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y) - f(x)| d^n y = 0. \tag{5.36}$$

The points where (5.36) holds are called ***Lebesgue points*** *of f. At every Lebesgue point we have*

$$f(x) = \lim_{j \to \infty} \frac{1}{|A_j(x)|} \int_{A_j(x)} f(y) d^n y \tag{5.37}$$

whenever $A_j(x)$ shrinks to x nicely, that is, 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 $|A_j| \geq \varepsilon |B_{r_j}(x)|$.

Problem* 5.17. *Let (X, Σ_X) be a measurable space and Y a Banach space. Then the set of all measurable functions forms a vector space. Moreover, if $f : X \to Y$ and $\varphi : X \to \mathbb{C}$ are measurable, then so is φf. If in addition, Y is a Banach algebra and $f, g : X \to Y$ are measurable, then so is fg. (Hint: Compare Corollary 1.19.)*

Problem* 5.18. *Let (X, Σ_X) be a measurable space and Y be a metric space equipped with the Borel σ-algebra. Show that the pointwise limit $f : X \to Y$ of measurable functions $f_n : X \to Y$ is measurable. (Hint: Show that* *for $U \subseteq Y$ open we have that $f^{-1}(U) = \bigcup_{m=1}^\infty \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty f_k^{-1}(U_m)$, where $U_m := \{y \in U | d(y, Y \setminus U) > \frac{1}{m}\}$.)*

Problem 5.19. *Suppose $\mu(A) < \infty$ and f integrable. Show*

$$\int_A f\, d\mu \in \mu(A)\overline{\operatorname{conv}(f(A))}.$$

Problem 5.20. *Let μ be σ-finite. Let f be integrable and $C \subset X$ closed. Show that if*

$$\int_A f\, d\mu \in \mu(A)C, \qquad \forall A \in \Sigma : \mu(A) < \infty,$$

then $f(x) \in C$ for a.e. $x \in X$. (Hint: Show that the preimage of any ball in $Y \setminus C$ is a null set.)

Problem* 5.21. *Let X be a separable Banach space. Show that there is a countable set $\ell_k \in X^*$ such that $\|x\| = \sup_k |\ell_k(x)|$ for all x.*

5.6. Lebesgue–Bochner spaces

In this section we briefly discuss Lebesgue spaces of functions with values in a Banach space Y. As in the complex-valued case we define the L^p norm by

$$\|f\|_p := \left(\int_X \|f\|^p\, d\mu\right)^{1/p}, \qquad 1 \le p, \tag{5.38}$$

and denote by $\mathcal{L}^p(X, d\mu, Y)$ the set of all strongly measurable functions for which $\|f\|_p$ is finite. Note that $\mathcal{L}^p(X, d\mu, Y)$ is a vector space, since $\|f + g\|^p \le 2^p \max(\|f\|, \|g\|)^p = 2^p \max(\|f\|^p, \|g\|^p) \le 2^p(\|f\|^p + \|g\|^p)$. Again Lemma 2.6 (applied to $\|f\|$) implies that we need to identify functions which are equal almost everywhere: Let

$$\mathcal{N}(X, d\mu, Y) := \{f \text{ strongly measurable} | f(x) = 0\ \mu\text{-almost everywhere}\} \tag{5.39}$$

and consider the quotient space

$$L^p(X, d\mu, Y) := \mathcal{L}^p(X, d\mu, Y)/\mathcal{N}(X, d\mu, Y). \tag{5.40}$$

If $d\mu$ is the Lebesgue measure on $X \subseteq \mathbb{R}^n$, we simply write $L^p(X, Y)$. Observe that $\|f\|_p$ is well defined on $L^p(X, d\mu, Y)$ and hence we have a normed space (the triangle inequality will be established below).

Lemma 5.28. *The integrable simple functions are dense in $L^p(X, d\mu, Y)$, $1 \le p < \infty$.*

Suppose X is a second countable topological space (i.e., it has a countable basis) and μ is an outer regular Borel measure. Then for every countable base $\mathcal{B}$ and total set $V \subseteq Y$ the set of simple functions $\alpha\chi_O$ with $\alpha \in V$, $O \in \mathcal{B}$, *and $\mu(O)$ finite is total. If in addition Y is separable, then $L^p(X, d\mu, Y)$, $1 \le p < \infty$, is separable.*

Proof. Let $f \in L^p(X, d\mu, Y)$. By Lemma 5.20 there is a sequence of simple functions such that $s_n(x) \to f(x)$ pointwise and $\|s_n(x)\| \le 2\|f(x)\|$. In particular, $s_n \in L^p(X, d\mu, Y)$ and thus integrable since $\|s_n\|_p^p = \|s_n\|_1$. Moreover, $\|f(x) - s_n(x)\|^p \le 3^p \|f(x)\|^p$ and thus $s_n \to f$ in $L^p(X, d\mu, Y)$ by dominated convergence. The rest follows as in Lemma 3.14. □

Similarly we define $L^\infty(X, d\mu, Y)$ together with the **essential supremum**

$$\|f\|_\infty := \inf\{C \,|\, \mu(\{x | \, \|f(x)\| > C\}) = 0\}. \tag{5.41}$$

Next note that the usual integral inequalities follow by applying the scalar case after using Lemma 5.18 (v):

Theorem 5.29 (Hölder's inequality)**.** *Let p and q be dual indices, $\frac{1}{p} + \frac{1}{q} = 1$, with $1 \le p \le \infty$. If*

(i) *$f \in L^p(X, d\mu, Y^*)$ and $g \in L^q(X, d\mu, Y)$, or*
(ii) *$f \in L^p(X, d\mu)$ and $g \in L^q(X, d\mu, Y)$, or*
(iii) *$f \in L^p(X, d\mu, Y)$ and $g \in L^q(X, d\mu, Y)$ and Y is a Banach algebra,*

then $f g$ is integrable and

$$\|f\, g\|_1 \le \|f\|_p \|g\|_q. \tag{5.42}$$

Theorem 5.30 (Minkowski's integral inequality)**.** *Suppose, (X, Σ_X, μ) and (Y, Σ_Y, ν) are two σ-finite measures, Z a Banach space, and $f : X \times Y \to Z$ is strongly $\mu \otimes \nu$ measurable. Let $1 \le p \le \infty$. Then*

$$\left\| \int_Y f(.,y) d\nu(y) \right\|_p \le \int_Y \|f(.,y)\|_p d\nu(y), \tag{5.43}$$

where the p-norm is computed with respect to μ. In particular, this says that $f(x,.)$ is integrable for a.e x and $\int_Y f(.,y)d\nu(y) \in L^p(X, d\mu, Z)$ if the integral on the right is finite.

Corollary 5.31 (Minkowski's inequality)**.** *Let $f, g \in L^p(X, d\mu, Y)$, $1 \le p \le \infty$. Then*

$$\|f + g\|_p \le \|f\|_p + \|g\|_p. \tag{5.44}$$

Moreover, literally the same proof as for the complex-valued case shows:

Theorem 5.32 (Riesz–Fischer)**.** *The space $L^p(X, d\mu, Y)$, $1 \le p \le \infty$, is a Banach space.*

Corollary 5.33. *If $\|f_n - f\|_p \to 0$, $1 \le p \le \infty$, then there is a subsequence f_{n_j} (of representatives) which converges pointwise almost everywhere and a nonnegative function $G \in L^p(X, d\mu)$ such that $\|f_{n_j}(x)\| \le G(x)$ almost everywhere.*

Of course, if Y is a Hilbert space, then $L^2(X, d\mu, Y)$ is a Hilbert space with scalar product

$$\langle f, g\rangle = \int_X \langle f(x), g(x)\rangle_Y d\mu(x). \tag{5.45}$$

Also Theorem 3.18 extends with literally the same proof:

Theorem 5.34. *Let X be a locally compact metric space and let μ be a regular Borel measure. Then the set $C_c(X, Y)$ of continuous functions with compact support and values in a dense linear subspace $\tilde{Y} \subseteq Y$ is dense in $L^p(X, d\mu, Y)$, $1 \le p < \infty$.*

And mollification works:

Lemma 5.35. *If $\phi \in L^1(\mathbb{R}^n)$ and $f \in L^p(\mathbb{R}^n, Y)$ with $1 \le p < \infty$. Then*

$$\phi * f = \int_{\mathbb{R}^n} \phi(x-y) f(y) d^n y \tag{5.46}$$

satisfies ***Young's inequality***

$$\|\phi * f\|_p \le \|\phi\|_1 \|f\|_p. \tag{5.47}$$

If ϕ_ε is an approximate identity we have

$$\lim_{\varepsilon \downarrow 0} \phi_\varepsilon * f = f \tag{5.48}$$

with the limit taken in L^p. In the case $p = \infty$ the claim holds for $f \in C_0(\mathbb{R}^n, Y)$.

Moreover, if $\phi \in C_c^k(\mathbb{R}^n)$ and $f \in L^1(\mathbb{R}^n, Y)$, then $\phi * f \in C^k(\mathbb{R}^n, Y)$ and

$$\partial_\alpha(\phi * f) = (\partial_\alpha \phi) * f \tag{5.49}$$

for any partial derivative of order at most k.

Next we turn to absolutely continuous functions. Let $I = [a, b]$ be a compact interval together with Lebesgue measure. We set

$$AC(I, Y) := \{f : I \to Y | \exists g \in L^1(I, Y) : f(x) = f(a) + \int_{[a,x)} g(y) dy\}. \tag{5.50}$$

By dominated convergence, one sees that $AC(I, Y) \subseteq C(I, Y)$. Moreover, by the Lebesgue differentiation theorem (Theorem 5.27) the same proof as for Corollary 4.10 shows that an absolutely continuous function is differentiable with $f' = g$ a.e. In particular, g is uniquely determined. If I is an open interval, we call $f : I \to Y$ absolutely continuous if it is absolutely continuous on every compact subinterval and we set

$$W^{1,p}(I,Y) := \{f \in AC(I,Y) \cap L^p(I,Y) | f' \in L^p(I,Y)\}, \tag{5.51}$$

which is a Banach space when equipped with the norm

$$\|f\|_{W^{1,p}} := \begin{cases} (\|f\|_p^p + \|f'\|_p^p)^{1/p}, & 1 \le p < \infty, \\ \max(\|f\|_\infty, \|f'\|_\infty), & p = \infty. \end{cases} \tag{5.52}$$

It is a Hilbert space if $p = 2$ and Y is a Hilbert space.

Problem 5.22. *Let $Y = C[a,b]$ and $f : [0,1] \to Y$ integrable. Compute the Bochner Integral $\int_0^1 f(x)dx$.*

Problem 5.23. *Show that an absolutely continuous function $f \in AC([a,b],Y)$ satisfies that for every $\varepsilon > 0$ there is a corresponding $\delta > 0$ such that*

$$\sum_k |y_k - x_k| < \delta \quad \Rightarrow \quad \sum_k \|f(y_k) - f(x_k)\| < \varepsilon$$

for every countable collection of pairwise disjoint intervals $(x_k, y_k) \subset [a,b]$. (Hint: Have a look at the proof of Lemma 4.22.)

Problem 5.24. *The variation of a function $f : [a,b] \to Y$ with Y a Banach space is defined as*

$$V_a^b(f) := \sup_{\text{partitions } P \text{ of } [a,b]} V(P,f), \quad V(P,f) := \sum_{k=1}^m \|f(x_k) - f(x_{k-1})\|.$$

Verify that everything said in Section 4.4 up to the first part of Example 4.11 continues to hold if absolute values are replaced by norms.

Problem 5.25. *Let $U \subseteq \mathbb{R}^n$ be a domain and $I \subseteq \mathbb{R}$ an interval. Show that $C^j(I, L^p(U)) \cap L^r(I, L^q(U))$ together with the norm*

$$\|f\| := \|f\|_{C^j(I,L^p(U))} + \|f\|_{L^r(I,L^q(U))}$$

is a Banach space for $j,k,l \in \mathbb{N}_0$, $1 \le p,q,r \le \infty$. (Hint: Work with test functions from C_c^∞ and recall Problem 3.35. Remark: The same arguments works if $L^p(U)$ and $L^q(U)$ are replaced by the Sobolev spaces $W^{k,p}(U)$ and $W^{l,q}(U)$), respectively.)

Problem 5.26. *Let $U \subseteq \mathbb{R}^n$ be a domain and $I \subseteq \mathbb{R}$ an interval. Show that $L^r(I, L^p(U))$, $1 \le r,p < \infty$, can be identified with the set of all measurable functions $f : I \times U \to \mathbb{C}$ for which*

$$\left(\int_I \left(\int_U |f(t,x)|^p d^n x \right)^{r/p} \right)^{1/r} < \infty.$$

In particular, $L^p(I, L^p(U)) \cong L^p(I \times U)$ for $1 \le p < \infty$.

The first claim also holds for $p, r = \infty$ if the corresponding integrals are replaced by the (essential) sup and the additional assumption that the range $\bigcup_{t\in I}\{f(t,.)\} \subseteq L^\infty(U)$ is separable in the case $p = \infty$. Show that this additional condition is necessary.

Problem 5.27. *Let $U \subseteq \mathbb{R}^n$ be a domain and $I \subseteq \mathbb{R}$ an interval. Show that for a strongly measurable function f and $1 \le p, r < \infty$ we have*

$$\|f\|_{L^r(I,L^p(U))} = \sup_{\|g\|_{L^{r'}(I,L^{p'}(U))}=1} \int_I \int_U |f(t,x)||g(t,x)| d^n x\, dt,$$

where p', r' are the corresponding dual indices. Moreover, it suffices to take the sup over functions which have support in a compact rectangle.

Problem 5.28. *Show that if $f \in L^{p_0}(X, d\mu, Y) \cap L^{p_1}(X, d\mu, Y)$ for some $p_0 < p_1$ then $f \in L^p(X, d\mu, Y)$ for every $p \in [p_0, p_1]$ and we have the* ***Lyapunov inequality***

$$\|f\|_p \le \|f\|_{p_0}^{1-\theta} \|f\|_{p_1}^{\theta},$$

where $\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$, $\theta \in (0,1)$. (Hint: Classical Lyapunov inequality from Problem 3.13.)

Problem 5.29. *Let $U \subseteq \mathbb{R}^n$ be a domain and $I \subseteq \mathbb{R}$ an interval. Suppose $f \in L^{p_0}(I, L^{q_0}(U)) \cap L^{p_1}(I, L^{q_1}(U))$. Show that $f \in L^{p_\theta}(I, L^{q_\theta}(U))$ for $\theta \in [0,1]$, where*

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \qquad \frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

(Hint: Lyapunov and generalized Hölder inequality — Problems 3.13 and 3.10.)

Problem 5.30. *Suppose X is a Hilbert space. For $f \in L^1(\mathbb{R}^n, X)$ define its Fourier transform as*

$$\mathcal{F}(f)(p) \equiv \hat{f}(p) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}px} f(x) d^n x.$$

Show that $\hat{f} \in C_b(\mathbb{R}^n, X)$. Moreover, show that the Fourier transform extends to a unitary operator on $L^2(\mathbb{R}^n, X)$. (Hint: For the first part just follow the argument for the case $X = \mathbb{C}$. For the second part it suffices to consider simple functions in which case the claim again reduces to the case $X = \mathbb{C}$).

Chapter 6

The dual of L^p

6.1. 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. For $1 \le p < \infty$ this is indeed the case:

Theorem 6.1. *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*

$$\ell_g(f) := \int_X g f \, d\mu \tag{6.1}$$

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 Corollary 3.5.

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 $A := \dot{\bigcup}_{j=1}^\infty A_j$, 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 $A_n = \{x| |g(x)| < n\}$, then $g_n = g\chi_{A_n} \in L^q$ and by Corollary 3.5 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 $\|g_n\|_q \le \|\ell\|$ independent of n, we have $g \in L^q$. By construction $\ell(f\chi_{X_n}) = \ell_g(f\chi_{X_n})$ for every $f \in L^p$ and letting $n \to \infty$ shows $\ell(f) = \ell_g(f)$. □

Corollary 6.2. *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. □

Note that in the case $0 < p < 1$, where L^p fails to be a Banach space, the dual might even be empty (see Problem 6.2)!

Problem 6.1. *Let $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$ with $\frac{1}{p} + \frac{1}{q} = 1$. Show that*

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x+n) g(x) dx = 0.$$

In other words, the sequence $f(.+n)$ converges weakly to 0 for $1 \le p < \infty$.

Problem* 6.2. *Formally extend the definition of $L^p(0,1)$ to $p \in (0,1)$. Show that $\|.\|_p$ does not satisfy the triangle inequality. However, show that it is a **quasinormed space**, that is, it satisfies all requirements for a normed space except for the triangle inequality which is replaced by*

$$\|a + b\| \le K(\|a\| + \|b\|)$$

with some constant $K \ge 1$. Show, in fact,

$$\|a + b\|_p \le 2^{1/p-1}(\|a\|_p + \|b\|_p), \qquad p \in (0,1).$$

Moreover, show that $L^p(0,1)^* = \{0\}$ in this case. (Hint: For the first part show $\alpha+\beta \le (\alpha^p+\beta^p)^{1/p} \le 2^{1/p-1}(\alpha+\beta)$ for $0 < p < 1$ and $\alpha, \beta \ge 0$. For the *second suppose there were a nontrivial $\ell \in L^p(0,1)^*$. Start with $f_0 \in L^p$ such that $|\ell(f_0)| \geq 1$. Set $g_0 = \chi_{(0,s]} f$ and $h_0 = \chi_{(s,1]} f$, where $s \in (0,1)$ is chosen such that $\|g_0\|_p = \|h_0\|_p = 2^{-1/p}\|f_0\|_p$. Then $|\ell(g_0)| \geq \frac{1}{2}$ or $|\ell(h_0)| \geq \frac{1}{2}$ and we set $f_1 = 2g_0$ in the first case and $f_1 = 2h_0$ else. Iterating this procedure gives a sequence f_n with $|\ell(f_n)| \geq 1$ and $\|f_n\|_p = 2^{-n(1/p-1)}\|f_0\|_p$.)*

Problem 6.3. *Suppose $K : L^p(Y, d\nu) \to L^p(X, d\mu)$ is an integral operator with kernel $K(x,y)$ satisfying the Schur criterion (Lemma 3.25). Show that for $1 \leq p < \infty$ the adjoint operator $K' : L^q(X, d\mu) \to L^q(Y, d\nu)$ is given by*

$$(K'f)(x) = \int_X K(y,x) f(y) d\mu(y), \qquad f \in L^q(X, d\mu).$$

6.2. 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 6.1. Let ν be a complex measure. Then

$$\ell_\nu(f) := \int_X f d\nu \tag{6.2}$$

is a bounded linear functional on $B(X)$ (the Banach space of bounded measurable functions) with norm

$$\|\ell_\nu\| = |\nu|(X) \tag{6.3}$$

by (4.25) and Corollary 4.19. 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)

$$\nu\Big(\dot{\bigcup_{k=1}^n} A_k\Big) = \sum_{k=1}^n \nu(A_k), \qquad A_j \cap A_k = \emptyset,\ j \neq k. \tag{6.4}$$

A content is called positive if $\nu(A) \geq 0$ for all $A \in \Sigma$ and given ν we can define its **total variation** $|\nu|(A)$ as in (4.16). The same proof as in Theorem 4.13 shows that $|\nu|$ is a positive 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 $|\nu|(X) < \infty$. As in (4.18), (4.19) we can split every content into a complex linear combination of four positive contents.

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

$$\int_A s\, d\nu := \sum_{k=1}^n \alpha_k \nu(A_k \cap A). \tag{6.5}$$

As in the proof of Lemma 2.1 one shows that the integral is linear. Moreover,

$$|\int_A s\, d\nu| \le |\nu|(A)\, \|s\|_\infty \tag{6.6}$$

and our integral is a densely defined bounded linear functional and hence there is a unique extension to all of $B(X)$ such that

$$|\int_X f\, d\nu| \le |\nu|(X)\, \|f\|_\infty \tag{6.7}$$

by Theorem 1.16 from **[22]** (compare Problem 2.4). 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 6.3. *Every bounded linear functional $\ell \in B(X)^*$ is of the form*

$$\ell(f) = \int_X f\, d\nu \tag{6.8}$$

for some unique finite 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, (6.8) holds for characteristic functions and by

$$\ell(\sum_{k=1}^n \alpha_k \chi_{A_k}) = \sum_{k=1}^n \alpha_k \nu(A_k) = \sum_{k=1}^n |\nu(A_k)|, \qquad \alpha_k := \operatorname{sign}(\nu(A_k)),$$

we see $|\nu|(X) \le \|\ell\|$.

Since the characteristic functions are total, (6.8) holds everywhere by continuity and (6.7) shows $\|\ell\| \le |\nu|(X)$. □

It is also easy to tell when ν is positive. To this end call ℓ a **positive functional** if $\ell(f) \ge 0$ whenever $f \ge 0$.

Corollary 6.4. *Let $\ell \in B^*(X)$ be associated with the finite content ν. Then ν will be a positive content if and only if ℓ is a positive functional. Moreover, every $\ell \in B^*(X)$ can be written as a complex linear combination of four positive functionals.*

Proof. Clearly $\ell \geq 0$ implies $\nu(A) = \ell(\chi_A) \geq 0$. Conversely $\nu(A) \geq 0$ implies $\ell(s) \geq 0$ for every simple $s \geq 0$. Now for $f \geq 0$ we can find a sequence of simple functions s_n such that $\|s_n - f\|_\infty \to 0$. Moreover, by $||s_n| - f| \leq |s_n - f|$ we can assume s_n to be nonnegative. But then $\ell(f) = \lim_{n\to\infty} \ell(s_n) \geq 0$ as required.

The last part follows by splitting the content ν into a linear combination of positive contents. □

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 6.4), 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 6.2. Consider $B(\mathbb{R})$ and define

$$\ell(f) = \lim_{\varepsilon\downarrow 0}(\lambda f(-\varepsilon) + (1-\lambda) f(\varepsilon)), \qquad \lambda \in (0,1], \tag{6.9}$$

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:

$$\lambda = \nu([-1,0)) = \nu(\bigcup_{n=1}^\infty [-\frac{1}{n}, -\frac{1}{n+1})) \neq \sum_{n=1}^\infty \nu([-\frac{1}{n}, -\frac{1}{n+1})) = 0. \tag{6.10}$$

Observe that the corresponding distribution function (defined as in (1.17)) 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! ⋄

Based on this observation we can give a simple proof of the Riesz representation for compact intervals. The general version will be shown in the next section.

Theorem 6.5 (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*

$$\ell(f) = \int_I f\, d\nu \tag{6.11}$$

for some unique complex Borel measure ν and $\|\ell\| = |\nu|(I)$.

Proof. By the Hahn–Banach theorem we can extend ℓ to a bounded linear functional $\overline{\ell} \in B(I)^*$ and we have a corresponding content $\tilde{\nu}$. Splitting this content into positive parts 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 (1.17). By finite additivity of $\tilde{\nu}$ it will be nondecreasing and we can use

Theorem 1.13 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

$$f_n(x) = f(x_0^n)\chi_{[x_0^n, x_1^n)} + f(x_1^n)\chi_{[x_1^n, x_2^n)} + \dots + f(x_{n-1}^n)\chi_{[x_{n-1}^n, x_n^n]}.$$

converges uniformly to f by continuity of f. 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 4.19). Now choose continuous functions $h_n(x) \to h(x)$ pointwise a.e. (Theorem 3.18). Using $\tilde{h}_n = \frac{h_n}{\max(1,|h_n|)}$ we 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 (6.7). □

Problem* 6.4. *Let M be a closed subspace of a Banach space X. Show that $(X/M)^* \cong \{\ell \in X^* | M \subseteq \operatorname{Ker}(\ell)\}$ (cf. Theorem 4.21 from* **[22]***).*

Problem 6.5. *Let μ and ν be σ-finite measures on X and Y, respectively. Let $p, q < \infty$ and let p', q' be the corresponding dual indices. Suppose that $K : X \times Y \to \mathbb{C}$ is measurable with $K(x,.) \in L^{q'}(Y, d\nu)$ for a.e. x. Moreover, suppose that*

$$(Kf)(x) := \int_X K(x,y) f(y) d\nu(y)$$

is a bounded operator from $L^q(, d\nu) \to L^p(X, d\mu)$ and the same holds for $|K|$. Then, if we identify the duals $L^p(X, d\mu)^* \cong L^q(X, d\mu)$ and $L^p(X, d\mu)^* \cong L^q(X, d\mu)$, the adjoint of K is given by

$$(K' f)(y) := \int_Y K(x,y) g(x) d\nu(x).$$

6.3. The Riesz–Markov representation theorem

In this section section we want to generalize Theorem 6.5. To this end X will be a metric space with the Borel σ-algebra. Given a Borel measure μ the integral

$$\ell(f) := \int_X f\, d\mu \tag{6.12}$$

will define a linear functional ℓ on the set of continuous functions with compact support $C_c(X)$. If μ were bounded we could drop the requirement for f to have compact support, but we do not want to impose this restriction here. However, in an arbitrary metric space there might not be many continuous functions with compact support. In fact, if $f \in C_c(X)$ and $x \in X$ is such that $f(x) \neq 0$, then $f^{-1}(B_r(f(x))$ will be a relatively compact neighborhood of x whenever $0 < r < |f(x)|$. So in order to be able to *see* all of X, we will assume that every point has a relatively compact neighborhood, that is, X is locally compact.

Moreover, note that positivity of μ implies that ℓ is **positive** in the sense that $\ell(f) \geq 0$ if $f \geq 0$. This raises the question if there are any other requirements for a linear functional to be of the form (6.12). The purpose of this section is to prove that there are none, that is, there is a one-to-one connection between positive linear functionals on $C_c(X)$ and positive Borel measures on X.

As a preparation let us reflect how μ could be recovered from ℓ as in (6.12). Given a Borel set A it seems natural to try to approximate the characteristic function χ_A by continuous functions form the inside or the outside. However, if you try this for the rational numbers in the case $X = \mathbb{R}$, then this works neither from the inside nor the outside. So we have to be more modest. If K is a compact set, we can choose a sequence $f_n \in C_c(X)$ with $f_n \downarrow \chi_K$ (Problem 6.6). In particular,

$$\mu(K) = \lim_{n\to\infty} \int_X f_n d\mu \tag{6.13}$$

by dominated convergence. So we can recover the measure of compact sets from ℓ and hence μ if it is inner regular. In particular, for every positive linear functional there can be at most one inner regular measure. This shows how μ should be defined given a linear functional ℓ. Nevertheless it will be more convenient for us to approximate characteristic functions of open sets from the inside since we want to define an outer measure and use the Carathéodory construction. Hence, given a positive linear functional ℓ we define

$$\rho(O) := \sup\{\ell(f) | f \in C_c(X),\ f \prec O\} \tag{6.14}$$

for any open set O. Here $f \prec O$ is short hand for $f \leq \chi_O$ and $\mathrm{supp}(f) \subseteq O$. Since ℓ is positive, so is ρ. Note that it is not clear that this definition will indeed coincide with $\mu(O)$ if ℓ is given by (6.12) unless O has a compact exhaustion. However, this is of no concern for us at this point.

Lemma 6.6. *Given a positive linear functional ℓ on $C_c(X)$ the set function ρ defined in* (6.14) *has the following properties:*

(i) $\rho(\emptyset) = 0$,

(ii) *monotonicity* $\rho(O_1) \le \rho(O_2)$ *if* $O_1 \subseteq O_2$,

(iii) ρ *is finite for relatively compact sets,*

(iv) $\rho(O) \le \sum_n \rho(O_n)$ *for every countable open cover* $\{O_n\}$ *of* O, *and*

(v) *additivity* $\rho(O_1 \cupdot O_2) = \rho(O_1) + \rho(O_2)$ *if* $O_1 \cap O_2 = \emptyset$.

Proof. (i) and (ii) are clear. To see (iii) note that if $\overline{O}$ is compact, then by Urysohn's lemma there is a function $f \in C_c(X)$ which is one on $\overline{O}$ implying $\rho(O) \le \ell(f)$. To see (iv) let $f \in C_c(X)$ with $f \prec O$. Then finitely many of the sets $O_1, \dots, O_N$ will cover $K := \operatorname{supp}(f)$. Set

$$h_j(x) := \frac{\operatorname{dist}(x, X \setminus O_j)}{\sum_{k=1}^N \operatorname{dist}(x, X \setminus O_k) + \operatorname{dist}(x, K)}, \qquad 1 \le j \le N.$$

Then $\chi_K \le h_1 + \cdots + h_N$ and hence

$$\ell(f) = \sum_{j=1}^N \ell(h_j f) \le \sum_{j=1}^N \rho(O_j) \le \sum_n \rho(O_n).$$

To see (v) note that $f_1 \prec O_1$ and $f_2 \prec O_2$ implies $f_1 + f_2 \prec O_1 \cupdot O_2$ and hence $\ell(f_1) + \ell(f_2) = \ell(f_1 + f_2) \le \rho(O_1 \cupdot O_2)$. Taking the supremum over f_1 and f_2 shows $\rho(O_1) + \rho(O_2) \le \rho(O_1 \cupdot O_2)$. The reverse inequality follows from the previous item. □

Lemma 6.7. *Let ℓ be a positive linear functional on $C_c(X)$ and let ρ be defined as in* (6.14). *Then*

$$\mu^*(A) := \inf\Big\{\rho(O) \Big| A \subseteq O,\ O \text{ open}\Big\}. \tag{6.15}$$

defines a metric outer measure on X.

Proof. Consider the outer measure (Lemma 1.8)

$$\nu^*(A) := \inf\Big\{\sum_{n=1}^\infty \rho(O_n) \Big| A \subseteq \bigcup_{n=1}^\infty O_n,\ O_n \text{ open}\Big\}.$$

Then we clearly have $\nu^*(A) \le \mu^*(A)$. Moreover, if $\nu^*(A) < \mu^*(A)$ we can find an open cover $\{O_n\}$ such that $\sum_n \rho(O_n) < \mu^*(A)$. But for $O = \bigcup_n O_n$ we have $\mu^*(A) \le \rho(O) \le \sum_n \rho(O_n)$, a contradiction. Hence $\mu^* = \nu^*$ and we have an outer measure.

To see that μ^* is a metric outer measure let A_1, A_2 with $\mathrm{dist}(A_1, A_2) > 0$ be given. Then, there are disjoint open sets $O_1 \supseteq A_1$ and $O_2 \supseteq A_2$. Hence for $A_1 \uplus A_2 \subseteq O$ we have $\mu^*(A_1) + \mu^*(A_2) \le \rho(O_1 \cap O) + \rho(O_2 \cap O) \le \rho(O)$ and taking the infimum over all O we have $\mu^*(A_1) + \mu^*(A_2) \le \mu^*(A_1 \uplus A_2)$. The converse follows from subadditivity and hence μ^* is a metric outer measure. □

So Theorem 1.9 gives us a corresponding measure μ defined on the Borel σ-algebra by Lemma 1.11. By construction this Borel measure will be outer regular and it will also be inner regular as the next lemma shows. Note that if one is willing to make the extra assumption of separability for X, this will come for free from Corollary 1.23.

Lemma 6.8. *The Borel measure μ associated with μ^* from* (6.15) *is regular.*

Proof. Since μ is outer regular by construction it suffices to show

$$\mu(O) = \sup_{K \subseteq O,\, K \text{ compact}} \mu(K)$$

for every open set $O \subseteq X$. Now denote the last supremum by α and observe $\alpha \le \mu(O)$ by monotonicity. For the converse we can assume $\alpha < \infty$ without loss of generality. Then, by the definition of $\mu(O) = \rho(O)$ we can find some $f \in C_c(X)$ with $f \prec O$ such that $\mu(O) \le \ell(f) + \varepsilon$. Since $K := \mathrm{supp}(f) \subseteq O$ is compact we have $\mu(O) \le \ell(f) + \varepsilon \le \mu(K) + \varepsilon \le \alpha + \varepsilon$ and as $\varepsilon > 0$ is arbitrary this establishes the claim. □

Now we are ready to show

Theorem 6.9 (Riesz–Markov representation)**.** *Let X be a locally compact metric space. Then every positive linear functional $\ell : C_c(X) \to \mathbb{C}$ gives rise to a unique regular Borel measure μ such that* (6.12) *holds.*

Proof. We have already constructed a corresponding Borel measure μ and it remains to show that ℓ is given by (6.12). To this end observe that if $f \in C_c(X)$ satisfies $\chi_O \le f \le \chi_C$, where O is open and C closed, then $\mu(O) \le \ell(f) \le \mu(C)$. In fact, every $g \prec O$ satisfies $\ell(g) \le \ell(f)$ and hence $\mu(O) = \rho(O) \le \ell(f)$. Similarly, for every $\tilde{O} \supseteq C$ we have $f \prec \tilde{O}$ and hence $\ell(f) \le \rho(\tilde{O})$ implying $\ell(f) \le \mu(C)$.

Now the next step is to split f into smaller pieces for which this estimate can be applied. To this end suppose $0 \le f \le 1$ and define $g_k^n := \min(f, \frac{k}{n})$ for $0 \le k \le n$. Clearly $g_0^n = 0$ and $g_n^n = f$. Setting $f_k^n := g_k^n - g_{k-1}^n$ for $1 \le k \le n$ we have $f = \sum_{k=1}^n f_k^n$ and $\frac{1}{n}\chi_{C_k^n} \le f_k^n \le \frac{1}{n}\chi_{O_{k-1}^n}$ where $O_k^n = \{x \in X | f(x) > \frac{k}{n}\}$ and $C_k^n = \overline{O_k^n} = \{x \in X | f(x) \ge \frac{k}{n}\}$. Summing over

k we have $\frac{1}{n}\sum_{k=1}^{n}\chi_{C_k^n} \le f \le \frac{1}{n}\sum_{k=0}^{n-1}\chi_{O_k^n}$ as well as

$$\frac{1}{n}\sum_{k=1}^{n}\mu(O_k^n) \le \ell(f) \le \frac{1}{n}\sum_{k=0}^{n-1}\mu(C_k^n)$$

Hence we obtain

$$\int f\,d\mu - \frac{\mu(O_0^n)}{n} \le \frac{1}{n}\sum_{k=1}^{n}\mu(O_k^n) \le \ell(f) \le \frac{1}{n}\sum_{k=0}^{n-1}\mu(C_k^n) \le \int f\,d\mu + \frac{\mu(C_0^n)}{n}$$

and letting $n \to \infty$ establishes the claim since $C_0^n = \mathrm{supp}(f)$ is compact and hence has finite measure. □

Note that this might at first sight look like a contradiction since (6.12) gives a linear functional even if μ is not regular. However, in this case the Riesz–Markov theorem merely says that there will be a corresponding regular measure which gives rise to the same integral for continuous functions. Moreover, using (6.13) one even sees that both measures agree on compact sets.

As a consequence we can also identify the dual space of $C_0(X)$ (i.e. the closure of $C_c(X)$ as a subspace of $C_b(X)$). Note that $C_0(X)$ is separable if X is locally compact and separable (Lemma B.37 from **[22]**). Also recall that a complex measure is regular if all four positive measures in the Jordan decomposition (4.20) are. By Lemma 4.21 this is equivalent to the total variation being regular.

Theorem 6.10 (Riesz–Markov Jr.[1] representation)**.** *Let X be a locally compact metric space. Every bounded linear functional $\ell \in C_0(X)^*$ is of the form*

$$\ell(f) = \int_X f\,d\nu \qquad (6.16)$$

for some unique regular complex Borel measure ν and $\|\ell\| = |\nu|(X)$. Moreover, ℓ will be positive if and only if ν is.

If X is compact this holds for $C(X) = C_0(X)$.

Proof. First of all observe that (4.25) shows that for every regular complex measure ν equation (6.16) gives a linear functional ℓ with $\|\ell\| \le |\nu|(X)$. This functional will be positive if ν is. Moreover, we have $d\nu = h\,d|\nu|$ (Corollary 4.19) and by Theorem 3.18 we can find a sequence $h_n \in C_c(X)$ with $h_n(x) \to h(x)$ pointwise a.e. Using $\tilde{h}_n := \frac{h_n}{\max(1,|h_n|)}$ we even get such a sequence with $|\tilde{h}_n| \le 1$. Hence $\ell(\tilde{h}_n^*) = \int \tilde{h}_n^* h\,d|\nu| \to \int |h|^2 d|\nu| = |\nu|(X)$ implying $\|\ell\| \ge |\nu|(X)$.

[1] Andrey Markov Jr. (1903–1979), Soviet mathematician

Conversely, let ℓ be given. By the Hahn–Banach theorem we can extend ℓ to a bounded linear functional $\overline{\ell} \in B(X)^*$ which can be written as a linear combinations of positive functionals by Corollary 6.4. Hence it is no restriction to assume ℓ is positive. But for positive ℓ the previous theorem implies existence of a corresponding regular measure ν such that (6.16) holds for all $f \in C_c(X)$. Since $\overline{C_c(X)} = C_0(X)$ (6.16) holds for all $f \in C_0(X)$ by continuity. □

Example 6.3. Note that the dual space of $C_b(X)$ will in general be larger. For example, consider $C_b(\mathbb{R})$ and define $\ell(f) = \lim_{x\to\infty} f(x)$ on the subspace of functions from $C_b(\mathbb{R})$ for which this limit exists. Extend ℓ to a bounded linear functional on all of $C_b(\mathbb{R})$ using Hahn–Banach. Then ℓ restricted to $C_0(\mathbb{R})$ is zero and hence there is no associated measure such that (6.16) holds. ◇

The above result implies that for bounded sequences, vague convergence is the same as weak-$*$ convergence in $C_0(X)^*$. One direction following from Lemma 5.14 and the other from the fact that every weak-$*$ convergent sequence is bounded.

As a consequence we can extend Helly's selection theorem. We call a sequence of complex measures ν_n vaguely convergent to a measure ν if

$$\int_X f d\nu_n \to \int_X f d\nu, \qquad f \in C_c(X). \tag{6.17}$$

This generalizes our definition for positive measures from Section 5.4. Moreover, note that in the case that the sequence is bounded, $|\nu_n|(X) \le M$, we get (6.17) for all $f \in C_0(X)$. Indeed, choose $g \in C_c(X)$ such that $\|f-g\|_\infty < \varepsilon$ and note that $\limsup_n |\int_X f d\nu_n - \int_X f d\nu| \le \limsup_n |\int_X (f-g) d\nu_n - \int_X (f-g) d\nu| \le \varepsilon(M + |\nu|(X))$.

Theorem 6.11. *Let X be a locally compact metric space. Then every bounded sequence ν_n of regular complex measures, that is $|\nu_n|(X) \le M$, has a vaguely convergent subsequence whose limit is regular. If all ν_n are positive, every limit of a convergent subsequence is again positive.*

Proof. Let $Y = C_0(X)$. Then we can identify the space of regular complex measure $\mathcal{M}_{reg}(X)$ as the dual space Y^* by the Riesz–Markov theorem. Moreover, every bounded sequence has a weak-$*$ convergent subsequence by the Banach–Alaoglu theorem (Theorem 6.10 from **[22]** — if X and hence $C_0(X)$ is separable, then Lemma 4.36 from **[22]** will suffice) and this subsequence converges in particular vaguely.

If the measures are positive, then $\ell_n(f) = \int f\, d\nu_n \ge 0$ for every $f \ge 0$ and hence $\ell(f) = \int f\, d\nu \ge 0$ for every $f \ge 0$, where $\ell \in Y^*$ is the limit of some convergent subsequence. Hence ν is positive by the Riesz–Markov representation theorem. □

Recall once more that in the case where X is locally compact and separable, regularity will automatically hold for every Borel measure.

Example 6.4. This theorem applies in particular to the case of a sequence of probability measures μ_n. If one additionally assumes that the sequence is tight (cf. Problem 5.15) then there is a weakly convergent subsequence whose limit is again a probability measure. ⋄

Problem* 6.6. *Let X be a locally compact metric space. Show that for every compact set K there is a sequence $f_n \in C_c(X)$ with $0 \le f_n \le 1$ and $f_n \downarrow \chi_K$. (Hint: Urysohn's lemma.)*

Problem 6.7. *Show the Lemma 6.6 holds also in case X is a regular topological space. (Hint: Urysohn's lemma.)*

Problem 6.8. *A function*

$$F(z) := bz + a + \int_{\mathbb{R}} \frac{1+zx}{x-z} d\mu(x), \qquad z \in \mathbb{C} \setminus \mathbb{R},$$

*with $b \ge 0$, $a \in \mathbb{R}$ and μ a finite (nonnegative) measure is called a **Herglotz–Nevanlinna function**. It is easy to check that $F(z)$ is analytic (cf. Problem 2.18), satisfies $F(z^*) = F(z)^*$, and maps the upper half plane to itself (unless $b = 0$ and $\mu = 0$), that is $\mathrm{Im}(F(z)) \ge 0$ for $\mathrm{Im}(z) > 0$. In fact, it can be shown (this is the Herglotz representation theorem) that any function with these properties is of the above form.*

*Show that if F_n is a sequence of Herglotz–Nevanlinna functions which converges pointwise to a function F for z in some set $U \subseteq \mathbb{C}_+$ which has a limit point in $\mathbb{C}_+$, then F is also a Herglotz–Nevanlinna function and $a_n \to a$, $b_n \to b$ and $\mu_n \to \mu$ weakly. (Hint: Note that you can get a bound on $\mu_n(\mathbb{R})$ by considering $F_n(\mathrm{i})$. Now use Helly's selection theorem (Lemma 5.17 will do) and note that by the identity theorem the limit is uniquely determined by the values of F on U. Finally recall that vague convergence of measures is just weak-$*$ convergence on $C_0(\mathbb{R})^*$ and use Lemma B.5 from **[22]**.)*

Chapter 7

Sobolev spaces

7.1. Warmup: Differentiable and Hölder continuous functions

Given $U \subseteq \mathbb{R}^n$ the set of all bounded continuous functions $C_b(U)$ together with the sup norm

$$\|f\|_\infty := \sup_{x\in U} |f(x)| \tag{7.1}$$

is a Banach space (cf. Corollary B.36 from **[22]**). The space of continuous functions with compact support $C_c(U) \subseteq C_b(U)$ is in general not dense and its closure will be denoted by $C_0(U)$. If U is open $C_0(U)$ can be interpreted as the functions in $C_b(U)$ which vanish at the boundary

$$C_0(U) := \{f \in C(U) | \forall \varepsilon > 0, \exists K \subseteq U \text{ compact} : |f(x)| < \varepsilon,\ x \in U \setminus K\}. \tag{7.2}$$

Of course $\mathbb{R}^n$ could be replaced by any topological space up to this point.

Moreover, for U open 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

$$\|f\|_{1,\infty} := \|f\|_\infty + \sum_{j=1}^n \|\partial_j f\|_\infty \tag{7.3}$$

is 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 7.3) the above expression defines indeed a norm. It is also not hard to see that $C_b^1(U)$ is complete. In fact, let f^m be a Cauchy sequence, then $f^m(x)$ converges uniformly to some continuous function $f(x)$ and the same is true for the partial derivatives $\partial_j f^m(x) \to g_j(x)$. Moreover, since $f^m(x) = f^m(c, x_2, \dots, x_n) + \int_c^{x_1} \partial_1 f^m(t, x_2, \dots, x_n)dt \to f(x) = f(c, x_2, \dots, x_n) + \int_c^{x_1} g_1(t, x_2, \dots, x_n)dt$ we obtain $\partial_1 f(x) = g_1(x)$. The remaining derivatives follow analogously and thus $f^m \to f$ in $C_b^1(U)$.

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

$$\partial_\alpha f := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}, \qquad |\alpha| := \alpha_1 + \cdots + \alpha_n, \tag{7.4}$$

for $|\alpha| \le k$. In this context $\alpha \in \mathbb{N}_0^n$ is called a **multi-index** and $|\alpha|$ is called its **order**. Also recall that by the classical theorem of Schwarz the order in which these derivatives are performed is irrelevant. With this notation the above considerations can be easily generalized to higher order derivatives:

Theorem 7.1. *Let $U \subseteq \mathbb{R}^n$ be open. 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*

$$\|f\|_{k,\infty} := \sum_{|\alpha| \le k} \sup_{x \in U} |\partial_\alpha f(x)|. \tag{7.5}$$

An important subspace is $C_0^k(U)$ which we define as the closure of $C_c^k(U)$:

$$C_0^k(U) := \overline{C_c^k(U)}. \tag{7.6}$$

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

$$[f]_\gamma := \sup_{x \ne y \in U} \frac{|f(x) - f(y)|}{|x - y|^\gamma} \tag{7.7}$$

is finite. Clearly, any Hölder continuous function is uniformly continuous and, in the special case $\gamma = 1$, we obtain the **Lipschitz continuous** functions. Note that for $\gamma = 0$ the Hölder condition boils down to boundedness and also the case $\gamma > 1$ is not very interesting (Problem 7.2).

Example 7.1. By the mean value theorem every function $f \in C_b^1(U)$ is Lipschitz continuous with $[f]_\gamma \le \|\nabla f\|_\infty$, where $\nabla f = (\partial_1 f, \dots, \partial_n f)$ denotes the gradient. ⋄

Example 7.2. The prototypical example of a Hölder continuous function is of course $f(x) = x^\gamma$ on $[0, \infty)$ with $\gamma \in (0, 1]$. In fact, without loss of generality we can assume $0 \le x < y$ and set $t = \frac{x}{y} \in [0, 1)$. Then we have

$$\frac{y^\gamma - x^\gamma}{(y - x)^\gamma} \le \frac{1 - t^\gamma}{(1 - t)^\gamma} \le \frac{1 - t}{1 - t} = 1.$$

From this one easily gets further examples since the composition of two Hölder continuous functions is again Hölder continuous (the exponent being the product). ◇

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

Theorem 7.2. *Let $U \subseteq \mathbb{R}^n$ be open. 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*

$$\|f\|_{k,\gamma,\infty} := \|f\|_{k,\infty} + \sum_{|\alpha|=k} [\partial_\alpha f]_\gamma. \tag{7.8}$$

As before, observe that the closure of $C_c^k(U)$ is $C_0^{k,\gamma}(U) := C_b^{k,\gamma}(U) \cap C_0^k(U)$. Moreover, as also already noted before, in the case $\gamma = 0$ we get a norm which is equivalent to $\|f\|_{k,\infty}$ and we will set $C_b^{k,0}(U) := C_b^k(U)$ for notational convenience later on.

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

Example 7.3. So while locally, differentiability is stronger than Lipschitz continuity, globally the situation depends on the domain: The sign function is in $C_b^1(\mathbb{R}\setminus\{0\})$ but it is not in $C_b^{0,1}(\mathbb{R}\setminus\{0\})$. In fact it is not even uniformly continuous. Also observe that the fact that its derivative is Lipschitz continuous on $\mathbb{R}\setminus\{0\}$ does not help. ◇

Moreover, every Hölder continuous function is uniformly continuous and hence has a unique extension to the closure $\overline{U}$ (cf. Theorem B.39 from **[22]**). In this sense, the spaces $C_b^{0,\gamma}(U)$ and $C_b^{0,\gamma}(\overline{U})$ are naturally isomorphic. Consequently, we can also understand $C_b^{k,\gamma}(\overline{U})$ in this fashion since for a function from $C_b^{k,\gamma}(U)$ all derivatives have a continuous extension to $\overline{U}$. For a function in $C_b^k(U)$ this will not work in general and hence we define $C_b^k(\overline{U})$ as the functions from $C_b^k(U)$ for which all derivatives have a continuous extensions to $\overline{U}$. Note that with this definition $C_b^k(\overline{U})$ is still a Banach space (since $C_b(\overline{U})$ is a closed subspace of $C_b(U)$). Finally, since Hölder continuous functions on a bounded domain are automatically bounded, we can drop the subscript b in this situation.

Theorem 7.3. *Suppose $U \subset \mathbb{R}^n$ is bounded. Then $C^{0,\gamma_2}(\overline{U}) \subseteq C^{0,\gamma_1}(\overline{U}) \subseteq C(\overline{U})$ for $0 < \gamma_1 < \gamma_2 \le 1$ with the embeddings being compact.*

Proof. That we have continuous embeddings follows since $|x-y|^{-\gamma_1} = |x-y|^{-\gamma_2+(\gamma_2-\gamma_1)} \le (2r)^{\gamma_2-\gamma_1}|x-y|^{-\gamma_2}$ if $U \subseteq B_r(0)$. Moreover, that the embedding $C^{0,\gamma_1}(\overline{U}) \subseteq C(\overline{U})$ is compact follows from the Arzelà–Ascoli theorem (Theorem B.40 from **[22]**). To see the remaining claim let f_m be a bounded sequence in $C^{0,\gamma_2}(\overline{U})$, explicitly $\|f_m\|_\infty \le C$ and $[f_m]_{\gamma_2} \le C$. Hence by the Arzelà–Ascoli theorem we can assume that f_m converges uniformly to some $f \in C(\overline{U})$. Moreover, taking the limit in $|f_m(x) - f_m(y)| \le C|x-y|^{\gamma_2}$ we see that we even have $f \in C^{0,\gamma_2}(\overline{U})$. To see that f is the limit of f_m in $C^{0,\gamma_1}(\overline{U})$ we need to show $[g_m]_{\gamma_1} \to 0$, where $g_m := f_m - f$. Now observe that

$$\begin{aligned}[g_m]_{\gamma_1} &\le \sup_{x\ne y\in U:|x-y|\ge\varepsilon} \frac{|g_m(x)-g_m(y)|}{|x-y|^{\gamma_1}} + \sup_{x\ne y\in U:|x-y|<\varepsilon} \frac{|g_m(x)-g_m(y)|}{|x-y|^{\gamma_1}}\\ &\le 2\|g_m\|_\infty \varepsilon^{-\gamma_1} + [g_m]_{\gamma_2}\varepsilon^{\gamma_2-\gamma_1} \le 2\|g_m\|_\infty \varepsilon^{-\gamma_1} + 2C\varepsilon^{\gamma_2-\gamma_1},\end{aligned}$$

implying $\limsup_{m\to\infty}[g_m]_{\gamma_1} \le 2C\varepsilon^{\gamma_2-\gamma_1}$ and since $\varepsilon > 0$ is arbitrary this establishes the claim. □

As pointed out in the example before, the embedding $C_b^1(U) \subseteq C_b^{0,1}(U)$ is continuous and combining this with the previous result immediately gives

Corollary 7.4. *Suppose $U \subset \mathbb{R}^n$ is bounded, $k_1, k_2 \in \mathbb{N}_0$, and $0 \le \gamma_1, \gamma_2 \le 1$. Then $C^{k_2,\gamma_2}(\overline{U}) \subseteq C^{k_1,\gamma_1}(\overline{U})$ for $k_1+\gamma_1 \le k_2+\gamma_2$ with the embeddings being compact if the inequality is strict.*

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

Problem 7.1. *Show*

$$C_0^k(U) = \overline{C_c^\infty(U)} = \{f \in C_b^k(U) | \partial^\alpha f \in C_0(U),\ 0 \le |\alpha| \le k\}.$$

(Hint: Use mollification and observe that derivatives come for free from Lemma 3.20.)

Problem 7.2. *Let $U \subseteq \mathbb{R}^n$ be open. Suppose $f : U \to \mathbb{C}$ is Hölder continuous with exponent $\gamma > 1$. Show that f is constant on every connected component of U.*

Problem* 7.3. *Suppose X is a vector space and $\|.\|_j$, $1 \le j \le m$, is a finite family of seminorms. Show that $\|x\| := \sum_{j=1}^m \|x\|_j$ is a seminorm. It is a norm if and only if $\|x\|_j = 0$ for all j implies $x = 0$.*

Problem* 7.4. *Let $U \subseteq \mathbb{R}^n$. Show that $C_b(U)$ is a Banach space when equipped with the sup norm. Show that $\overline{C_c(U)} = C_0(U)$. (Hint: The function $m_\varepsilon(z) = \operatorname{sign}(z)\max(0, |z|-\varepsilon) \in C(\mathbb{C})$ might be useful.)*

Problem 7.5. *Let $U \subseteq \mathbb{R}^n$. Show that the product of two bounded Hölder continuous functions is again Hölder continuous with*

$$[fg]_\gamma \le \|f\|_\infty [g]_\gamma + [f]_\gamma \|g\|_\infty.$$

Problem 7.6. *Let $\phi \in L^1(\mathbb{R}^n)$ and $f \in C_b^{0,\gamma}(\mathbb{R}^n)$. Show*

$$[\phi * f]_\gamma \le \|\phi\|_1 [f]_\gamma.$$

7.2. Basic properties

Throughout this chapter $U \subseteq \mathbb{R}^n$ will be nonempty and open and we will use the notation $V \subset\subset U$ if V is a relatively compact set with $\overline{V} \subset U$.

Our aim is to extend the Lebesgue spaces to include derivatives. To this end, for a locally integrable function $f \in L^1_{loc}(U)$, a locally integrable function $h \in L^1_{loc}(U)$ satisfying

$$\int_U \varphi(x) h(x) d^n x = (-1)^{|\alpha|} \int_U (\partial_\alpha \varphi)(x) f(x) d^n x, \quad \forall \varphi \in C_c^\infty(U), \tag{7.9}$$

is called the **weak derivative** or the derivative in the sense of distributions of f. Note that by Lemma 3.23 such a function is unique if it exists. Moreover, if $f \in C^k(U)$ then integration by parts (2.67) shows that the weak derivative coincides with the usual derivative. Also note that the order in which the partial derivatives are taken is irrelevant for φ (by the classical theorem of Schwarz) and hence the same is true for weak derivatives. This is no contradiction to the classical counterexamples for the theorem of Schwarz since weak derivatives are only defined up to equivalence a.e.

Example 7.4. Consider $f(x) := x^2 \sin(\frac{\pi}{x^2})$ on $U := (-1, 1)$ (here $f(0) := 0$). Then it is straightforward to verify that

$$f'(x) = \begin{cases} 2x \sin(\frac{\pi}{x^2}) - \frac{2\pi}{x} \cos(\frac{\pi}{x^2}), & x \ne 0, \\ 0, & x = 0, \end{cases}$$

that is, f is everywhere differentiable on U. Of course f is weakly differentiable on $(-1, 0)$ as well as on $(0, 1)$ implying that the weak derivative of f on U must equal f' (consider test functions φ supported away from 0). However, one can check that f' is not integrable (Problem 7.8) and hence f is not weakly differentiable. ◇

Example 7.5. Consider $U := \mathbb{R}$. If $f(x) := |x|$, then $\partial f(x) = \operatorname{sign}(x)$ as a weak derivative. If we try to take a second derivative we are lead to

$$\int_{\mathbb{R}} \varphi(x) h(x) dx = -\int_{\mathbb{R}} \varphi'(x) \operatorname{sign}(x) dx = 2\varphi(0)$$

and it is easy to see that no locally integrable function can satisfy this requirement. ◇

Example 7.6. In fact, in one dimension the class of weakly differentiable functions can be identified with the class of antiderivatives of integrable functions, that is, the class of **absolutely continuous** functions

$$AC[a,b] := \{f(x) = f(a) + \int_a^x h(y)dy | h \in L^1(a,b)\},$$

where $a < b$ are some real numbers. It is easy to see that every absolutely continuous function is in particular continuous, $AC[a,b] \subset C[a,b]$. Moreover, using Lebesgue's differentiation theorem one can show that an absolutely continuous function is differentiable a.e. with $f'(x) = h(x)$ (and hence h is uniquely defined a.e.). We refer to Section 4.4 for further details.

If $f, g \in AC[a,b]$, we have the integration by parts formula (4.43), which shows that every absolutely continuous function has a weak derivative which equals the a.e. derivative. In fact, with a little more effort one can show (Problem 7.10) that the converse is also true, that is, $W^{1,1}(a,b) = AC[a,b]$ and $W^{1,p}(a,b) = \{f \in AC[a,b] | f' \in L^p(a,b)\}$. Consequently $W^{k,1}(a,b) = \{f \in C^{k-1}[a,b] | f^{(k-1)} \in AC[a,b]\}$ and $W^{k,p}(a,b) = \{f \in C^{k-1}[a,b] | f^{(k-1)} \in AC[a,b],\ f^{(k)} \in L^p(a,b)\}$. ⋄

Example 7.7. One can verify (Problem 7.11) that $f(x) := |x|^{-\gamma}$ is weakly differentiable for $\gamma < \frac{n-p}{p}$. Hence in higher dimensions weakly differentiable functions might not be continuous. ⋄

Now we can define the **Sobolev space** $W^{k,p}(U)$ as the set of all functions in $L^p(U)$ which have weak derivatives up to order k in $L^p(U)$. Clearly $W^{k,p}(U)$ is a linear space since $f, g \in W^{k,p}(U)$ implies $af + bg \in W^{k,p}(U)$ for $a, b \in \mathbb{C}$ and $\partial_\alpha(af + bg) = a\partial_\alpha f + b\partial_\alpha g$ for all $|\alpha| \le k$. Moreover, for $f \in W^{k,p}(U)$ we define its norm

$$\|f\|_{k,p} := \begin{cases} \left(\sum_{|\alpha|\le k} \|\partial_\alpha f\|_p^p\right)^{1/p}, & 1 \le p < \infty, \\ \max_{|\alpha|\le k} \|\partial_\alpha f\|_\infty, & p = \infty. \end{cases} \tag{7.10}$$

We will also use the gradient $\nabla u = (\partial_1 u, \dots, \partial_n u)$ and by $\|\nabla u\|_p$ we will always mean $\||\nabla u|\|_p$, where $|\nabla u|$ denotes the Euclidean norm.

It is easy to check that with this definition $W^{k,p}$ becomes a normed linear space. Of course for $p = 2$ we have a corresponding scalar product

$$\langle f, g\rangle_{W^{k,2}} := \sum_{|\alpha|\le k} \langle \partial_\alpha f, \partial_\alpha g\rangle_{L^2} \tag{7.11}$$

and one reserves the special notation $H^k(U) := W^{k,2}(U)$ for this case. Similarly we define local versions of these spaces $W^{k,p}_{loc}(U)$ as the set of all functions in $L^p_{loc}(U)$ which have weak derivatives up to order k in $L^p_{loc}(U)$.

Theorem 7.5. *For each $k \in \mathbb{N}_0$, $1 \le p \le \infty$ the Sobolev space $W^{k,p}(U)$ is complete, that is, it is a Banach space. It is separable for $1 \le p < \infty$ as well as reflexive and uniformly convex for $1 < p < \infty$.*

Proof. Let f_m be a Cauchy sequence in $W^{k,p}$. Then $\partial_\alpha f_m$ is a Cauchy sequence in L^p for every $|\alpha| \le k$. Consequently $\partial_\alpha f_m \to f_\alpha$ in L^p. Moreover, letting $m \to \infty$ in

$$\int_U \varphi f_\alpha d^n x = \lim_{m\to\infty} \int_U \varphi(\partial_\alpha f_m) d^n x = \lim_{m\to\infty} (-1)^{|\alpha|} \int_U (\partial_\alpha \varphi) f_m d^n x$$
$$= (-1)^{|\alpha|} \int_U (\partial_\alpha \varphi) f_0 d^n x, \qquad \varphi \in C_c^\infty(U),$$

shows $f_0 \in W^{k,p}$ with $\partial_\alpha f_0 = f_\alpha$ for $|\alpha| \le k$. By construction $f_m \to f_0$ in $W^{k,p}$ which implies that $W^{k,p}$ is complete.

Concerning the last claim note that $W^{k,p}(U)$ can be regarded as a subspace of $\bigoplus_{p,|\alpha|\le k} L^p(U)$ which has the claimed properties by Lemma 3.14 and Theorem 3.11 (see also the remark after the proof), Corollary 6.2 (cf. also Problem 4.30 from [**22**]). □

As a consequence of the proof we record:

Corollary 7.6. *Fix some multi-index α. Let $f_n \in L^p(U)$ be a sequence such that $\partial_\alpha f_n \in L^p(U)$ exists. Then $f_n \to f$, $\partial_\alpha f_n \to g$ in $L^p(U)$ implies that $\partial_\alpha f = g$ exists.*

Of course we have the natural embedding $W^{k,p}(U) \hookrightarrow L^p(U)$ and if $V \subseteq U$ is nonempty and open, then $f \in W^{k,p}(U)$ implies $f|_V \in W^{k,p}(V)$ (since $C_c^\infty(V) \subseteq C_c^\infty(U)$). Sometimes it is also useful to look at functions with values in $\mathbb{C}^n$ in which case we define $W^{k,p}(U,\mathbb{C}^n)$ as the corresponding direct sum.

Regarding $W^{k,p}(U)$ as a subspace of $\bigoplus_{p,|\alpha|\le k} L^p(U)$ also provides information on its dual space. Indeed, if $M \subseteq X$ is a closed subset of a Banach space X, then every linear functional on X gives rise to a linear functional on M by restricting its domain. Clearly two functionals give rise to the same restriction if their difference vanishes on M. Conversely, any functional on M can be extended to a functional on X by Hahn–Banach and thus any element from M^* arises in this way. Hence we have $M^* = X^*/M^\perp$, where $M^\perp$ denotes the annihilator of a subspace, that is, the set of all linear functionals which vanish on the subspace (cf. also Theorem 4.21 from [**22**]). Concerning the norm of a functional, note that when extending a functional from M to X, the norm can only increase. Moreover, the extension obtained from Hahn–Banach preserves the norm and hence the norm is given by taking the minimum over all extensions. In a strictly convex space the functional where the norm is attained is even unique (Problem 6.39 from [**22**]). Applied to $W^{k,p}(U)$ this gives

$$W^{k,p}(U)^* = \left(\bigoplus_{q,|\alpha|\le k} L^p(U)^*\right) / W^{k,p}(U)^\perp, \qquad \frac{1}{p} + \frac{1}{q} = 1, \tag{7.12}$$

and for $1 \le p < \infty$, such that $L^p(U)^* = L^q(U)$, we even get:

Lemma 7.7. *For $1 \le p < \infty$, every linear functional $\ell \in W^{k,p}(U)^*$ can be represented as*

$$\ell(f) = \sum_{|\alpha| \le k} \int_U g_\alpha(x)(\partial_\alpha f)(x) d^n x, \tag{7.13}$$

with some functions $g_\alpha \in L^q(U)$, $|\alpha| \le k$. Moreover,

$$\|\ell\| = \min\{\|g\|_{\bigoplus_{q,|\alpha|\le k} L^q(U)} | g_\alpha \text{ as in } (7.13)\}, \tag{7.14}$$

where the minimizer is unique if $1 < p < \infty$.

In the case $p = 2$ the Riesz[1] representation theorem (for the dual of a Hilbert space; Theorem 2.10 from [**22**]) tells us that the unique minimizer is given by some $g \in H^k(U)$ such that $g_\alpha = \partial_\alpha g$. As another consequence note that a sequence converges weakly in $W^{k,p}(U)$ if and only if all derivatives converge weakly in $L^p(U)$.

Example 7.8. Consider $W^{1,p}(0,1)$, then functions in this space are absolutely continuous and one can consider the linear functional

$$\ell_{x_0}(f) := f(x_0)$$

for given $x_0 \in [0,1]$. Defining

$$g_{x_0}(x) := \frac{1}{\sinh(1)} \begin{cases} \cosh(1-x_0)\cosh(x), & x \le x_0, \\ \cosh(1-x)\cosh(x_0), & x \ge x_0, \end{cases}$$

one verifies

$$\ell_{x_0}(f) = \int_0^1 g_{x_0}(x) f(x) dx + \int_0^1 g_{x_0}'(x) f'(x) dx.$$

This representation is however not unique! To this end, note that for any $h \in W_0^{1,q}[0,1]$

$$\ell(f) = \int_0^1 h'(x) f(x) dx + \int_0^1 h(x) f'(x) dx = h(1)f(1) - h(0)f(0) = 0$$

represents the zero functional. In fact, any representation of the zero functional is of this form (show this).

Moreover, note that

$$|f(x_0)| \le \|g_{x_0}\|_{W^{1,q}} \|f\|_{W^{1,p}}.$$

Since we have $\|g_{x_0}\|_\infty = g_{x_0}(x_0)$, which attains its maximum at the boundary points, we infer

$$\|f\|_\infty \le \coth(1)\|f\|_{W^{1,p}}.$$

[1] Frigyes Riesz (1880–1956), Hungarian mathematician

In particular, we have a continuous embedding $W^{1,p}(0,1) \hookrightarrow C[0,1]$. Moreover, for $1 < p \le \infty$, there is even a continuous embedding into the space of Hölder continuous functions $W^{1,p}(0,1) \hookrightarrow C^{0,\gamma}[0,1]$ with exponent $\gamma := 1-\frac{1}{p}$ (Problem 4.31). $\diamond$

Since functions from $W^{k,p}$ might not even be continuous, it will be convenient to know that they still can be well approximated by nice functions. To this end we next show that smooth functions are dense in $W^{k,p}$. A first naive approach would be to extend $f \in W^{k,p}(U)$ to all of $\mathbb{R}^n$ by setting it 0 outside U and consider $f_\varepsilon := \phi_\varepsilon * f$, where ϕ is the standard Friedrichs[2] mollifier. The problem with this approach is that we generically create a nondifferentiable singularity at the boundary and hence this only works as long as we stay away from the boundary.

Lemma 7.8 (Friedrichs)**.** *Let $f \in W^{k,p}(U)$ and set $f_\varepsilon := \phi_\varepsilon * f$, where ϕ is the standard mollifier. Then for every $\varepsilon_0 > 0$ we have $f_\varepsilon \to f$ in $W^{k,p}(U_{\varepsilon_0})$ if $1 \le p < \infty$, where $U_\varepsilon := \{x \in U | \operatorname{dist}(x, \mathbb{R}^n \setminus U) > \varepsilon\}$. If $p = \infty$ we have $\partial_\alpha f_\varepsilon \to \partial_\alpha f$ a.e. for all $|\alpha| \le k$.*

Proof. Just observe that all derivatives converge in L^p for $1 \le p < \infty$ since $\partial_\alpha f_\varepsilon = (\partial_\alpha \phi_\varepsilon) * f = \phi_\varepsilon * (\partial_\alpha f)$. Here the first equality is Lemma 3.20 (ii) and the second equality only holds (by definition of the weak derivative) on U_ε since in this case $\operatorname{supp}(\phi_\varepsilon(x - .)) = B_\varepsilon(x) \subseteq U$. So if we fix $\varepsilon_0 > 0$, then $f_\varepsilon \to f$ in $W^{k,p}(U_{\varepsilon_0})$. In the case $p = \infty$ the claim follows since $L^\infty_{loc} \subseteq L^1_{loc}$ after passing to a subsequence. That selecting a subsequence is superfluous follows from Problem 3.31. $\square$

Note that, by Lemma 7.15, if $f \in W^{k,\infty}(U)$ then $\partial_\alpha f$ is locally Lipschitz[3] continuous for all $|\alpha| \le k-1$. Hence $\partial_\alpha f_\varepsilon \to \partial_\alpha f$ locally uniformly for all $|\alpha| \le k-1$.

So in particular, we get convergence in $W^{k,p}(U)$ if f has compact support. To adapt this approach to work on all of U we will use a partition of unity.

Theorem 7.9 (Meyers[4]–Serrin[5])**.** *Let $U \subseteq \mathbb{R}^n$ be open and $1 \le p < \infty$. Then $C^\infty(U) \cap W^{k,p}(U)$ is dense in $W^{k,p}(U)$.*

Proof. The idea is to use a partition of unity to decompose f into pieces which are supported on layers close to the boundary and decrease the mollification parameter ε as we get closer to the boundary. To this end we start with the sets $U_j := \{x \in U | \operatorname{dist}(x, \mathbb{R}^n \setminus U) > \frac{1}{j}\}$ and we set $U_j := \emptyset$ for

[2]Kurt Otto Friedrichs (1901–1982), German American mathematician
[3]Rudolf Lipschitz (1832–1903), German mathematician
[4]Norman George Meyers (*1930), American mathematician
[5]James Serrin (1926–2012), American mathematician

$j \leq 0$. Now consider $\tilde{\zeta}_j := \phi_{\varepsilon_j} * \chi_{V_j}$, where $V_j := U_{j+1} \setminus U_{j-1}$ and ε_j chosen sufficiently small such that $\operatorname{supp}(\tilde{\zeta}_j) \subset U_{j+2} \setminus U_{j-2}$. Since the sets V_j cover U, the function $\tilde{\zeta} := \sum_j \tilde{\zeta}_j$ is positive on U and since for $x \in V_k$ only terms with $|j-k| \leq 2$ contribute, we also have $\tilde{\zeta} \in C^\infty(U)$. Hence considering $\zeta_j := \tilde{\zeta}_j/\tilde{\zeta} \in C_c^\infty(U)$ we have $\sum_j \zeta_j = 1$.

Now let $f \in W^{k,p}(U)$ be given and fix $\delta > 0$. By the previous lemma we can choose $\varepsilon_i > 0$ sufficiently small such that $f_j := \phi_{\varepsilon_j} * (\zeta_j f)$ has still support inside $U_{j+2} \setminus U_{j-2}$ and satisfies

$$\|f_j - \zeta_j f\|_{W^{k,p}} \leq \frac{\delta}{2^{j+1}}.$$

Then $f_\delta = \sum_j f_j \in C^\infty(U)$ since again only terms with $|j-k| \leq 2$ contribute. Moreover, for every set $V \subset\subset U$ we have

$$\|f_\delta - f\|_{W^{k,p}(V)} = \|\sum_j (f_j - \zeta_j f)\|_{W^{k,p}(V)} \leq \sum_j \|f_j - \zeta_j f\|_{W^{k,p}(V)} \leq \delta$$

and letting $V \nearrow U$ we get $f_\delta \in W^{k,p}(U)$ as well as $\|f_\delta - f\|_{W^{k,p}(U)} \leq \delta$. □

Historically this theorem had a significant impact since it showed that the two competing ways of defining Sobolev spaces, namely as the set of functions which have weak derivatives in L^p on one side and the closure of smooth functions with respect to the $W^{k,p}$ norm on the other side, actually agree.

Example 7.9. The example $f(x) := |x| \in W^{1,\infty}(-1,1)$ shows that the theorem fails in the case $p = \infty$ since $f'(x) = \operatorname{sign}(x)$ cannot be approximated uniformly by smooth functions. ⋄

For L^p we know that smooth functions with compact support are dense. This is no longer true in general for $W^{k,p}$ since convergence of derivatives enforces that the vanishing of boundary values is preserved in the limit. However, making this precise requires some additional effort. So for now we will just give the closure of $C_c^\infty(U)$ in $W^{k,p}(U)$ a special name $W_0^{k,p}(U)$ as well as $H_0^k(U) := W_0^{k,2}(U)$. It is easy to see that $C_c^k(U) \subseteq W_0^{k,p}(U)$ for every $1 \leq p \leq \infty$ and $W_c^{k,p}(U) \subseteq W_0^{k,p}(U)$ for every $1 \leq p < \infty$ (mollify to get a sequence in $C_c^\infty(U)$ which converges in $W^{k,p}(U)$). In the case $p = \infty$ we have $W_0^{k,\infty}(U) \subseteq C_0^k(U)$ (with equality for nice domains, see Problem 7.23). Moreover, note

Lemma 7.10. *We have $W_0^{k,p}(\mathbb{R}^n) = W^{k,p}(\mathbb{R}^n)$ for $1 \leq p < \infty$.*

Proof. We choose some cutoff function $\zeta_m \in C_c^\infty(\mathbb{R}^n)$, such that $\zeta_m(x) = 1$ for $|x| \leq m$, $\zeta_m(x) = 0$ for $|x| \geq m+1$, and $\|\partial_\alpha \zeta_m\|_\infty \leq C_\alpha$. For example, choose a function $h \in C^\infty(\mathbb{R})$ such $h(x) = 1$ for $x \leq 0$, $h(x) = 0$ for $x \geq 1$

and let $\zeta_m(x) := h(|x| - m)$. Now note that for $f \in L^p(\mathbb{R}^n)$ dominated convergence implies $\zeta_m f \to f$ in L^p and $(\partial_\alpha \zeta_m) f \to 0$ in L^p for $|\alpha| \geq 1$.

Fix $f \in W^{k,p}(\mathbb{R}^n)$ and consider $f_m := f\zeta_m \in W_c^{k,p}(\mathbb{R}^n)$. Then using Leibniz rule we see that $\partial_\alpha f_m \to \partial_\alpha f$ in $L^p(\mathbb{R}^n)$ for $|\alpha| \leq k$ and hence $f_m \to f$ in $W^{k,p}(\mathbb{R}^n)$. Thus $W_c^{k,p}(\mathbb{R}^n)$ is dense in $W^{k,p}(\mathbb{R}^n)$ and by mollification $C_c^\infty(\mathbb{R}^n)$ is dense in $W^{k,p}(\mathbb{R}^n)$. □

Next we collect some basic properties of weak derivatives.

Lemma 7.11. *Let $U \subseteq \mathbb{R}^n$ be open and $1 \leq p \leq \infty$.*

(i) *The operator $\partial_\alpha : W^{k,p}(U) \to W^{k-|\alpha|,p}(U)$ is a bounded linear map and $\partial_\beta \partial_\alpha f = \partial_\alpha \partial_\beta f = \partial_{\alpha+\beta} f$ for $f \in W^{k,p}$ and all multi-indices α, β with $|\alpha| + |\beta| \leq k$.*

(ii) *We have*

$$\int_U g(\partial_\alpha f) d^n x = (-1)^{|\alpha|} \int_U (\partial_\alpha g) f d^n x, \quad g \in W_0^{k,q}(U),\ f \in W^{k,p}(U), \tag{7.15}$$

for all $|\alpha| \leq k$, $\frac{1}{p} + \frac{1}{q} = 1$.

(iii) *Suppose $f \in W^{1,p}(U)$ and $g \in W^{1,q}(U)$ with $\frac{1}{r} := \frac{1}{p} + \frac{1}{q} \leq 1$. Then $f \cdot g \in W^{1,r}(U)$ and we have the **product rule***

$$\partial_j (f \cdot g) = (\partial_j f) g + f(\partial_j g), \qquad 1 \leq j \leq n. \tag{7.16}$$

The same claim holds with $q = p = r$ if $f \in W^{1,p}(U) \cap L^\infty(U)$.

(iv) *Suppose $\eta \in C^1(\mathbb{R}^m)$ has bounded derivatives and satisfies $\eta(0) = 0$ if $|U| = \infty$. Then the map $f \mapsto \eta \circ f$ is a continuous map $W^{1,p}(U, \mathbb{R}^m) \to W^{1,p}(U)$ and we have the **chain rule** $\partial_j(\eta \circ f) = \sum_k (\partial_k \eta)(f) \partial_j f_k$. If $\eta(0) = 0$, then composition with η will also map $W_0^{1,p}(U, \mathbb{R}^m) \to W_0^{1,p}(U)$.*

(v) *Let $\psi : U \to V$ be a C^1 diffeomorphism such that both ψ and ψ^{-1} have bounded derivatives. Then we have a bijective bounded linear map $W^{1,p}(V) \to W^{1,p}(U)$, $f \mapsto f \circ \psi$ and we have the **change of variables formula** $\partial_j(f \circ \psi) = \sum_k (\partial_k f)(\psi) \partial_j \psi_k$.*

(vi) *Let U be connected and suppose $f \in W^{1,p}(U)$ satisfies $\partial_j f = 0$ for $1 \leq j \leq n$. Then f is constant.*

Proof. (i) Problem 7.13.

(ii) Take limits in (7.9) using Hölder's inequality. If $g \in W_c^{k,q}(U)$ only the case $q = \infty$ is of interest which follows from dominated convergence.

(iii) First of all note that if $\phi, \varphi \in C_c^\infty(U)$, then $\phi\varphi \in C_c^\infty(U)$ and hence using the ordinary product rule for smooth functions and rearranging (7.9) with $\varphi \to \phi\varphi$ shows $\phi f \in W_c^{1,p}(U)$. Hence (7.15) with $g \to g\varphi \in W_c^{1,q}$ shows

$$\int_U gf(\partial_j\varphi)d^n x = -\int_U \big((\partial_j f)g + f(\partial_j g)\big)\varphi d^n x,$$

that is, the weak derivatives of $f \cdot g$ are given by the product rule and that they are in $L^r(U)$ follows from the generalized Hölder inequality (3.19).

(iv) Since the claim is trivially true for constant η, we can subtract $\eta(0)$ and can assume $\eta(0) = 0$ without loss of generality. Moreover, by assumption $|\nabla\eta| \le L$ and hence we have $|\eta(x) - \eta(y)| \le L|x-y|$ by the mean value theorem. Hence we see $\|\eta \circ f - \eta \circ g\|_p \le L\|f-g\|_p$ which shows that composition with η is a continuous map $L^p(U,\mathbb{R}^n) \to L^p(U)$. Also note that the proposed derivative will be in $L^p(U)$ provided $f \in W^{1,p}(U,\mathbb{R}^n)$.

Now suppose $f_n \to f$ in $W^{1,p}(U,\mathbb{R}^m)$ for some $V \subset U$. Then

$$\begin{aligned}\|(\nabla\eta)(f)\cdot\partial_j f - (\nabla\eta)(f_n)\cdot\partial_j f_n\|_{L^p} &\le L\|\partial_j f - \partial_j f_n\|_{L^p}\\ &+ \|((\nabla\eta)(f) - (\nabla\eta)(f_n))\cdot\partial_j f\|_{L^p},\end{aligned}$$

where the first norm tends to zero by assumption and the second by dominated convergence after passing to a subsequence which converges a.e. at least for $1 \le p < \infty$. In the case $p = \infty$ this holds since $\nabla\eta$ is uniformly continuous on bounded sets.

Now for $p < \infty$ we can choose f_n to be smooth (by Theorem 7.9). In this case the derivative of $\eta \circ f_n$ can be computed using the chain rule and the above argument shows that this formula remains true in the limit, that is, the proposed derivative is indeed the weak derivative. Since $W^{1,\infty}(U) \subset W^{1,1}_{loc}(U)$ this also covers the case $p = \infty$ by restricting to bounded sets.

Finally, the above argument also shows that if $f_n \to f$ in $W^{1,p}(U,\mathbb{R}^n)$ then every subsequence f_{n_j} has another subsequence f_{m_j} for which $\eta \circ f_{m_j} \to \eta \circ f$ in $W^{1,p}(U)$. This implies that $\eta \circ f_n \to \eta \circ f$.

For the last claim observe that composition with η maps $C_c^1(U,\mathbb{R}^n) \subset W_0^{1,p}(U,\mathbb{R}^n) \to C_c^1(U) \subset W_0^{1,p}(U)$ and hence the claim follows by density of these subspaces.

(v) If J_ψ denotes the Jacobi determinant, then using $|J_\psi| \ge C$ the change of variables formula implies

$$\int_U |f\circ\psi|^p d^n x \le \frac{1}{C}\int_U |f\circ\psi|^p |J_\psi| d^n x = \frac{1}{C}\int_V |f|^p d^n y,$$

which shows that composition with ψ is a homeomorphism between $L^p(U)$ and $L^p(V)$ for $1 \le p < \infty$. In the case $p = \infty$ we have $\|f\circ\psi\|_\infty = \|f\|_\infty$ and the claim is also true. To compute the weak derivative recall $L^p \subset L^1_{loc}$. Now let ϕ_ε be the standard mollifier and consider $f_\varepsilon := \phi_\varepsilon * f$. Then, using this fact, one computes

$$\begin{aligned}\int_U (f\circ\psi)\partial_j\varphi\, d^n x &= \lim_{\varepsilon\downarrow 0}\int_U (f_\varepsilon\circ\psi)(\partial_j\varphi) d^n x\\ &= \lim_{\varepsilon\downarrow 0}\int_U \sum_k ((\partial_k f_\varepsilon)\circ\psi)(\partial_j\psi_k)\varphi\, d^n x\\ &= \int_U \sum_k ((\partial_k f)\circ\psi)(\partial_j\psi_k)\varphi\, d^n x.\end{aligned}$$

This establishes the claim.

(vi) This is just a reformulation of Lemma 3.24. □

Of course item (iv) can be applied to complex-valued functions upon observing that taking real and imaginary parts is a bounded (real) linear map $W^{1,p}(U)\to W^{1,p}(U,\mathbb{R}^2)$, $f\mapsto(\mathrm{Re}(f),\mathrm{Im}(f))$. However, the important case of taking absolute values is not covered by (iv).

Lemma 7.12. *For $f\in W^{1,p}(U)$, $1\le p\le\infty$, we have $|f|\in W^{1,p}(U)$ with*

$$\partial_j|f|(x) = \begin{cases}\frac{\mathrm{Re}(f(x))}{|f(x)|}\partial_j\mathrm{Re}(f(x)) + \frac{\mathrm{Im}(f(x))}{|f(x)|}\partial_j\mathrm{Im}(f(x)), & f(x)\ne 0,\\ 0, & f(x)=0.\end{cases} \tag{7.17}$$

In particular, $\big|\partial_j|f|(x)\big|\le|\partial_j f(x)|$. For $1\le p<\infty$ this map is continuous on $W^{1,p}(U)$.

Furthermore, if f is real-valued we also have $f_\pm := \max(0,\pm f)\in W^{1,p}(U)$ with

$$\partial_j f_\pm(x) = \begin{cases}\pm\partial_j f(x), & \pm f(x)>0,\\ 0, & \text{else},\end{cases}\qquad \partial_j|f|(x) = \begin{cases}\partial_j f(x), & f(x)>0,\\ -\partial_j f(x), & f(x)<0,\\ 0, & \text{else}.\end{cases}$$

Moreover, if $f\in W_0^{1,p}(U)$, then $|f|\in W_0^{1,p}(U)$ for $1\le p<\infty$.

Proof. In order to reduce it to (iv) from the previous lemma we will take $f_1 := \mathrm{Re}(f)$, $f_2 := \mathrm{Im}(f)$ and approximate the absolute value of f by $\eta_\varepsilon(f_1,f_2)$ with $\eta_\varepsilon(x,y)=\sqrt{x^2+y^2+\varepsilon^2}-\varepsilon$.

We start by noting that $\|\nabla\eta_\varepsilon\|_\infty\le 1$ and hence we can apply the chain rule (Lemma 7.11 (iv)) with η_ε to see

$$\int_U \varphi(x)\frac{f_1(x)\partial_j f_1(x)+f_2(x)\partial_j f_2(x))}{\sqrt{|f(x)|^2+\varepsilon^2}} d^n x = -\int_U \partial_j\varphi(x)\eta_\varepsilon(f_1(x),f_2(x)) d^n x$$

Letting $\varepsilon\to 0$ (using dominated convergence) shows

$$\int_U \varphi(x)\frac{f_1(x)\partial_j f_1(x)+f_2(x)\partial_j f_2(x))}{|f(x)|} d^n x = -\int_U \partial_j\varphi(x)|f(x)| d^n x$$

(with the expression for the derivative understood as being 0 if $f(x) = 0$) and establishes the first part. The estimate for the derivative follows from $\mathrm{Re}(z_1)\mathrm{Re}(z_2) + \mathrm{Im}(z_1)\mathrm{Im}(z_2) = |z_1||z_2|\cos(\arg(z_2/z_1))$ for $z_1, z_2 \in \mathbb{C}$.

The second part follows from $f_\pm(x) = \frac{|f(x)| \pm f(x)}{2}$ and linearity of the weak derivative.

Moreover, using $\nabla f = \nabla f_+ - \nabla f_-$ shows that $\nabla f = 0$ for a.e. x with $f(x) = 0$. Hence if we have a sequence $f_n \to f$ in $W^{1,p}(U)$ we can choose a subsequence such that both f and ∇f converge pointwise a.e. Then, by the above formulas and the preceding remark, the same is true for $|f|$ and $\nabla|f|$. Thus dominated convergence shows $|f_n| \to |f|$ in $W^{1,p}(U)$ for $1 \le p < \infty$.

The claim for $f \in W_0^{1,p}(U)$ follows since if $f \in W_c^{1,p}(U)$ then $|f| \in W_c^{1,p}(U)$ and the claim follows by density. □

As byproduct of the proof we note:

Corollary 7.13. *For every $m \in \mathbb{R}$ and every $f \in W^{1,p}(U)$ we have $\nabla f = 0$ for a.e. x with $f(x) = m$.*

Of course this implies that item (iv) from Lemma 7.11 continues to hold if η is only piecewise C^1 with bounded derivative. To see this observe that by linearity it suffices to consider the case where η has only one kink. But then η can be written as the sum of a C^1 function and a multiple of a translated absolute value.

Example 7.10. The example $f_\varepsilon(x) := x - \varepsilon \in W^{1,\infty}(-1,1)$ shows that taking absolute values is not continuous in $W^{1,\infty}(-1,1)$ since $|f_\varepsilon|'(x) = \mathrm{sign}(x-\varepsilon)$ does not converge uniformly to $|f_0|'(x) = \mathrm{sign}(x)$. ⋄

Finally we look at situations where it is not a priori known that the function has a weak derivative. We will offer two variants. The first variant (ii) shows that an estimate on $T_a f - f$ is sufficient, where $T_a f(x) := f(x-a)$ is the translation operator from (3.21). Note that the estimate (ii) below should be thought of as an estimate for the difference quotient

$$D_j^\varepsilon f := \frac{T_{\varepsilon\delta^j} f - f}{-\varepsilon} \tag{7.18}$$

in the direction of the j'th coordinate axis. Our second variant employs duality and requires that the integral in (iii) below gives rise to a bounded functional. Both characterizations fail in the case $p = 1$.

Lemma 7.14. *For $f \in L^p(U)$ consider*

(i) *$f \in W^{1,p}(U)$ with $\|\nabla f\|_p \le C$.*

(ii) *There exists a constant C such that*

$$\|T_a f - f\|_{L^p(V)} \le C|a| \tag{7.19}$$

for every $V \subset\subset U$ and all $a \in \mathbb{R}^n$ with $|a| < \operatorname{dist}(V, \partial U)$.

(iii) *There exists a constant C with*

$$\left| \int_U f\big(\nabla\varphi\big) d^n x \right| \le C\|\varphi\|_{p'}, \qquad \varphi \in C_c^\infty(U), \tag{7.20}$$

where p' is the dual index, $\frac{1}{p} + \frac{1}{p'} = 1$.

Then we have (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (iii) for $1 \le p \le \infty$ and (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii) for $1 < p \le \infty$.

Proof. (i) $\Rightarrow$ (ii): By Theorem 7.9 we can assume that f is smooth without loss of generality and hence

$$|f(x-a) - f(x)| \le |a| \int_0^1 |\nabla f(x - t\, a)| dt$$

from which the case $p = \infty$ is immediate. In the case $1 \le p < \infty$ we integrate this inequality over V and employ Jensen's inequality to obtain

$$\begin{aligned} \|T_a f - f\|_{L^p(V)}^p &\le |a|^p \int_V \left| \int_0^1 |\nabla f(x - t\, a)| dt \right|^p d^n x \\ &\le |a|^p \int_V \int_0^1 |\nabla f(x - t\, a)|^p dt\, d^n x \\ &\le |a|^p \int_0^1 \int_V |\nabla f(x - t\, a)|^p d^n x\, dt \le |a|^p \|\nabla f\|_{L^p(U)}^p. \end{aligned}$$

(ii) $\Rightarrow$ (iii): Fix φ and choose some $V \subset\subset U$ with $\operatorname{supp}(\varphi) \subset V$. Using

$$\int_U (T_a f - f)\varphi d^n x = \int_U f\,(T_{-a}\varphi - \varphi) d^n x$$

for $|a| < \operatorname{dist}(V, \partial U)$ we obtain from (ii)

$$\left| \int_U f\,(T_{-a}\varphi - \varphi) d^n x \right| = \left| \int_V (T_a f - f)\varphi d^n x \right| \le C|a| \|\varphi\|_{p'}.$$

Choosing $a = \varepsilon\delta^j$ and taking $\varepsilon \to 0$ we get

$$\left| \int_U f\,(\partial_j \varphi) d^n x \right| \le C\|\varphi\|_{p'},$$

which implies (iii) with C replaced by $\sqrt{n}C$.

(iii) $\Rightarrow$ (i) for $p \ne 1$: Item (iii) implies that $\ell_j(\varphi) := \int_U f\big(\partial_j\varphi\big) d^n x$ is a densely defined bounded linear functional on $L^{p'}(U)$. Hence by Theorem 6.1 there is some $g_j \in L^p(U)$ (with $\|g_j\|_p \le C$) such that $\ell_j(\varphi) = -\int_U g_j \varphi d^n x$, that is $\partial_j f = g_j$.

This establishes the lemma in the case $p \ne 1$. The direction (iii) $\Rightarrow$ (ii) without the assumption $p \ne 1$ is left as an exercise (Problem 7.18). $\square$

Example 7.11. Consider $f(x) = \mathrm{sign}(x)$ on $U = \mathbb{R}$. We already know from Example 7.5 that f does not have a weak derivative. However, items (ii) and (iii) hold with $C = 2$.

The problem in the case $p = 1$ is that, by the Riesz–Markov representation theorem (Theorem 6.10), every bounded linear functional on $C_0(U)$ is given by a complex measure. Hence in this case the weak derivatives of f are in general complex measures rather than functions, that is, there exist Borel measures μ_j such that

$$\int_U f(\partial_j\varphi) d^n x = -\int_U \varphi d\mu_j(x).$$

The optimal constant in (iii)

$$V_U(f) := \sup_{\varphi\in C_c^\infty(U), \|\varphi\|_\infty\le 1} \left| \int_U f(\nabla\varphi) d^n x \right| \tag{7.21}$$

is known as the **total variation** of f. It is a semi-norm since $V_U(f) = 0$ if and only if f is constant on every connected component by Lemma 3.24. Accordingly the class of functions satisfying (ii) or (iii),

$$BV(U) := \{f \in L^1(U) | V_U(f) < \infty\}, \tag{7.22}$$

is known as the functions of **bounded variation**. In terms of the measures μ_j it is given by

$$V_U(f)^2 = \sum_{j=1}^n |\mu_j|(U)^2.$$

$\diamond$

In the case $p = \infty$ we can also characterize $W^{1,\infty}$ as follows:

Lemma 7.15. *We have $C_b^{0,1}(U) \subseteq W^{1,\infty}(U)$ with the embedding being continuous. Conversely, if U is convex then we have equality and the embedding is a homeomorphism.*

Proof. If $f \in C_b^{0,1}(U)$, then Lemma 7.14 implies $f \in W^{1,\infty}$ with $\|\nabla f\|_\infty \le [f]_1$.

Conversely, let $f \in W^{1,\infty}(U)$ and $f_\varepsilon = \phi_\varepsilon * f$ with ϕ the standard mollifier. Then by the mean value theorem

$$|f_\varepsilon(x) - f_\varepsilon(y)| \le \|\nabla f_\varepsilon\|_\infty |x-y| \le \|\nabla f\|_\infty |x-y|,$$

where the last inequality holds for ε sufficiently small. In fact, note that for $\varepsilon < \mathrm{dist}(x, \partial U)$ we have $\partial_j f_\varepsilon(x) = (\phi_\varepsilon * \partial_j f)(x)$ and the claim follows from Young's inequality (3.25). Now if x, y are Lebesgue points of f (cf. Lemma 3.21), then we can take the limit $\varepsilon \to 0$ to conclude

$$|f(x) - f(y)| \le \|\nabla f\|_\infty |x-y| \qquad \text{a.e. } x, y \in U.$$

In particular, f is uniformly continuous on a dense subset and hence has a (Lipschitz) continuous extension to all of U. □

Problem 7.7. *Consider $f(x) = \sqrt{x}$, $U = (0,1)$. Compute the weak derivative. For which p is $f \in W^{1,p}(U)$?*

Problem 7.8. *Show that the derivative of the function from Example 7.4 is not integrable.*

Problem 7.9. *Consider the Hilbert space $H^1(0,1)$. Compute the orthogonal complement of the following subspaces:*
a) $H^1_0(0,1)$ b) $\{f \in H^1(0,1) | \int_0^1 f(x)dx = 0\}$

Problem* 7.10. *Show that f is weakly differentiable in the interval (a,b) if and only if $f(x) = f(c) + \int_c^x h(t)dt$ is absolutely continuous and $f' = h$ in this case. (Hint: Lemma 7.11 (vi).)*

Problem* 7.11. *Consider $U := B_1(0) \subset \mathbb{R}^n$ and $f(x) = \tilde{f}(|x|)$ with $\tilde{f} \in C^1(0,1]$. Then $f \in W^{1,p}_{loc}(B_1(0) \setminus \{0\})$ and*

$$\partial_j f(x) = \tilde{f}'(|x|)\frac{x_j}{|x|}.$$

Show that if $\limsup_{r\to 0} r^{n-1}|\tilde{f}(r)| < \infty$, then $f \in W^{1,p}(B_1(0))$ if and only if $\tilde{f}, \tilde{f}' \in L^p((0,1), r^{n-1}dr)$.

Conclude that for $f(x) := |x|^{-\gamma}$, $\gamma > 0$, we have $f \in W^{1,p}(B_1(0))$ with

$$\partial_j f(x) = -\frac{\gamma x_j}{|x|^{\gamma+2}}$$

provided $\gamma < \frac{n-p}{p}$. (Hint: Use integration by parts on a domain which excludes $B_\varepsilon(0)$ and let $\varepsilon \to 0$.)

Problem* 7.12. *Show that for $\varphi \in C_c^\infty(\mathbb{R}^n)$ we have*

$$\varphi(0) = -\frac{1}{S_n}\int_{\mathbb{R}^n} \frac{x}{|x|^n} \cdot \nabla\varphi(x) d^n x.$$

Hence this weak derivative cannot be interpreted as a function. (Hint: Start with $\varphi(0) = -\int_0^\infty \big(\frac{d}{dr}\varphi(r\omega)\big)dr = -\int_0^\infty \nabla\varphi(r\omega)\cdot\omega\, dr$ and integrate with respect to ω over the unit sphere S^{n-1}; cf. Lemma 2.18.)

Problem* 7.13. *Show Lemma 7.11 (i).*

Problem* 7.14. *Suppose $f \in W^{k,p}(U)$ and $h \in C_b^k(U)$. Then $h \cdot f \in W^{k,p}(U)$ and we have **Leibniz' rule***

$$\partial_\alpha(h \cdot f) = \sum_{\beta\le\alpha} \binom{\alpha}{\beta} (\partial_\beta h)(\partial_{\alpha-\beta} f), \tag{7.23}$$

where $\binom{\alpha}{\beta} := \frac{\alpha!}{\beta!(\alpha-\beta)!}$, $\alpha! := \prod_{j=1}^m (\alpha_j!)$, and $\beta \le \alpha$ means $\beta_j \le \alpha_j$ for $1 \le j \le m$.

Problem 7.15. *Let $\psi : U \to V$ be a C^k diffeomorphism such that all derivatives of both ψ and ψ^{-1} are bounded. Then we have a bijective bounded liner map $W^{k,p}(V) \to W^{k,p}(U)$, $f \mapsto f \circ \psi$.*

Problem 7.16. *Suppose for each $x \in U$ there is an open neighborhood $V(x) \subseteq U$ such that $f \in W^{k,p}(V(x))$. Then $f \in W^{k,p}_{loc}(U)$. Moreover, if $\|f\|_{W^{k,p}(V)} \le C$ for every $V \subset\subset U$, then $f \in W^{k,p}(U)$.*

Problem 7.17. *Suppose $1 < p \le \infty$. Show that if $f_n \in W^{1,p}(U)$ is a sequence such that $f_n \to f$ in L^p and $\|\nabla f_n\|_p \le C$, then $f \in W^{1,p}(U)$. (Hint: Since $L^p(U)$ is the dual of the separable Banach space $L^{p'}(U)$ with $\frac{1}{p} + \frac{1}{p'} = 1$, we can extract a weak-$*$ convergent subsequence from any bounded sequence (Lemma 4.36 from* **[22]***).)*

Problem 7.18. *Establish the direction (iii) $\Rightarrow$ (ii) in Lemma 7.14 for arbitrary $1 \le p \le \infty$. (Hint: Problem 3.37.)*

Problem 7.19. *Show that $W^{k,p}(U) \cap W^{j,q}(U)$ (with $1 \le p, q \le \infty$, $j, k \in \mathbb{N}_0$) together with the norm $\|f\|_{W^{k,p} \cap W^{j,q}} := \|f\|_{W^{k,p}} + \|f\|_{W^{j,q}}$ is a Banach space.*

Problem 7.20 (Radial Sobolev spaces). *Consider $B := B_R(0) \subseteq \mathbb{R}^n$ (with the case $R = \infty$ allowed). Show that the subset $W^{k,p}_{\mathrm{rad}}(B)$ of radial functions from $W^{k,p}(B)$ is closed.*

For a radial function f define $\tilde{f}$ via $f(x) = \tilde{f}(r)$, where $r = |x|$. Show that if $f \in W^{1,p}_{\mathrm{rad}}(B)$ then $\tilde{f} \in W^{1,p}((0,R), r^{n-1}dr)$ with the embedding being continuous and the derivative given by

$$\tilde{f}'(r) = \sum_{j=1}^{n} \frac{x_j}{r} (\partial_j f)(x), \qquad r = |x|.$$

Here $W^{1,p}((0,R), r^{n-1}dr)$ is the set of all functions $f \in AC(0,R)$ for which $f, f' \in L^p((0,R), r^{n-1}dr)$.

7.3. Extension and trace operators

To proceed further we will need to be able to extend a given function beyond its original domain U. As already pointed out before, simply setting it equal to zero on $\mathbb{R}^n \setminus U$ will in general create a nondifferentiable singularity along the boundary. Moreover, considering $U = (-1,0) \cup (0,1)$ we have $f(x) := \operatorname{sign}(x) \in W^{1,p}(U)$ but it is not possible to extend f to $\mathbb{R}$ such that the extension is in $W^{1,p}(\mathbb{R})$.

Of course such problems do not arise if $f \in W^{1,p}_0(U)$ since we can simply extend f to a function on $\bar{f} \in W^{1,p}(\mathbb{R}^n)$ by setting $\bar{f}(x) = 0$ for $x \in \mathbb{R}^n \setminus U$ (Problem 7.21).

We will say that a domain $U \subseteq \mathbb{R}^n$ has the **extension property** if for all $1 \le p \le \infty$ there is an extension operator $E : W^{1,p}(U) \to W^{1,p}(\mathbb{R}^n)$ such that

- E is bounded, i.e., $\|Ef\|_{W^{1,p}(\mathbb{R}^n)} \le C_{U,p}\|f\|_{W^{1,p}(U)}$ and
- $Ef\big|_U = f$.

We begin by showing that if the boundary is a hyperplane, we can do the extension by a simple reflection. To this end consider the reflection $x^\star := (x_1, \dots, x_{n-1}, -x_n)$ which is an involution on $\mathbb{R}^n$. For a domain U which is symmetric with respect to reflection, that is, $U^\star = U$, write $U_\pm := \{x \in U | \pm x_n > 0\}$ and a function f defined on U_+ can be extended to $U_- \cup U_+$ using

$$f^\star(x) := \begin{cases} f(x), & x \in U_+, \\ f(x^\star), & x \in U_-. \end{cases} \tag{7.24}$$

Note that $f^\star$ extends to a continuous function in $C(\overline{U})$ provided $f \in C(\overline{U_+})$.

Lemma 7.16. *Let $U \subseteq \mathbb{R}^n$ be symmetric with respect to reflection and $1 \le p \le \infty$. If $f \in W^{1,p}(U_+)$ then the symmetric extension $f^\star \in W^{1,p}(U)$ satisfies $\|f^\star\|_{W^{1,p}(U)} = 2^{1/p}\|f\|_{W^{1,p}(U_+)}$. Moreover,*

$$(\partial_j f^\star) = \begin{cases} (\partial_j f)^\star, & 1 \le j < n, \\ \operatorname{sign}(x_n)(\partial_n f)^\star, & j = n. \end{cases} \tag{7.25}$$

Proof. It suffices to compute the weak derivatives. We start with $1 \le j < n$ and

$$\int_U f^\star \partial_j \varphi d^n x = \int_{U_+} f\, \partial_j \varphi^{\#} d^n x,$$

where $\varphi^{\#}(x) = \varphi(x) + \varphi(x^\star)$. Since $\varphi^{\#}$ is not compactly supported in U_+ we use a cutoff function $\eta_\varepsilon(x) = \eta(x_n/\varepsilon)$, where $\eta \in C^\infty(\mathbb{R}, [0,1])$ satisfies $\eta(r) = 0$ for $r \le \frac{1}{2}$ and $\eta(r) = 1$ for $r \ge 1$ (e.g., integrate and shift the standard mollifier to obtain such a function). Then

$$\begin{aligned} \int_U f^\star \partial_j \varphi\, d^n x &= \lim_{\varepsilon\to 0} \int_{U_+} f\, \partial_j(\eta_\varepsilon \varphi^{\#}) d^n x = -\lim_{\varepsilon \to 0} \int_{U_+} (\partial_j f)\eta_\varepsilon \varphi^{\#} d^n x \\ &= -\int_{U_+} (\partial_j f)\varphi^{\#} d^n x = -\int_U (\partial_j f)^\star \varphi\, d^n x \end{aligned}$$

for $1 \le j < n$. For $j = n$ we proceed similarly,

$$\int_U f^\star \partial_n \varphi\, d^n x = \int_{U_+} f\, \partial_n \varphi^\sharp d^n x,$$

where $\varphi^\sharp(x) = \varphi(x) - \varphi(x^\star)$. Note that $\varphi^\sharp(x_1, \dots, x_{n-1}, 0) = 0$ and hence $|\varphi^\sharp(x)| \le L\, x_n$ on U_+. Using this last estimate we have $|(\partial_n \eta_\varepsilon)\varphi^\sharp| \le C$ and hence we obtain as before

$$\int_U f^\star \partial_n \varphi \, d^n x = \lim_{\varepsilon \to 0} \int_{U_+} f \, \partial_n(\eta_\varepsilon \varphi^\sharp) d^n x = - \lim_{\varepsilon \to 0} \int_{U_+} (\partial_n f) \eta_\varepsilon \varphi^\sharp d^n x$$
$$= - \int_{U_+} (\partial_n f) \varphi^\sharp d^n x = - \int_U \operatorname{sign}(x_n)(\partial_n f)^\star \varphi \, d^n x,$$

which finishes the proof. □

Corollary 7.17. $\mathbb{R}^n_+$ *has the extension property. In fact, any rectangle (not necessarily bounded)* Q *has the extension property.*

Proof. Given a rectangle use the above lemma to extend it along every hyperplane bounding the rectangle. Finally, use a smooth cut-off function (e.g. mollify the characteristic function of a slightly larger rectangle). □

While this already covers some interesting domains, note that it fails if we look for example at the exterior of a rectangle. So our next result shows (maybe not too surprising), that it is the boundary which will play the crucial role. To this end we recall that U is said to have a C^1 boundary if around any point $x^0 \in \partial U$ we can find a C^1 diffeomorphism ψ which straightens out the boundary (cf. Section 2.4). As a preparation we note:

Lemma 7.18. *Suppose* U *has a bounded* C^1 *boundary. Then there is a finite number of open sets* $\{U_j\}_{j=0}^m$ *and corresponding functions* $\zeta_j \in C^\infty(\mathbb{R}^n)$ *with* $\operatorname{supp}(\zeta_j) \subset U_j$ *such that* $\sum_{j=0}^m \zeta_j(x) = 1$ *for all* $x \in U$, $U_0 \subseteq U$, $\{U_j\}_{j=1}^m$ *are bounded and cover* ∂U, *and for each* U_j, $1 \le j \le m$, *there is a* C^1 *diffeomorphism* $\psi_j : U_j \to Q_j$, *where* Q_j *is a rectangle which is symmetric with respect to reflection.*

Proof. Since near each $x \in \partial U$ we can straighten out the boundary, there is a corresponding open neighborhood U_x and a C^1 diffeomorphism $\psi_x : U_x \to Q_x$, where Q_x is a rectangle which is symmetric with respect to reflection. Moreover, there is also a corresponding radius $r(x)$ such that $\bar{B}_{r(x)}(x) \subset U_x$. By compactness of ∂U there are finitely many points $\{x_j\}_{j=1}^m$ such that the corresponding balls $B_{r(x_j)}(x_j)$ cover the boundary. Take $U_j := U_{x_j}$. Choose nonnegative functions $\tilde{\zeta}_j \in C_c^\infty(U_j)$ such that $\tilde{\zeta}_j > 0$ on $\bar{B}_{r(x_j)}(x_j)$ (e.g. mollify the characteristic function of $B_{r(x_j)}(x_j)$). Let $V := \bigcup_{j=1}^m B_{r(x_j)}(x_j)$ and $U_0 := U$. Choose a nonnegative function $\tilde{\zeta}_0 \in C^\infty(\mathbb{R}^n)$ supported inside U such that $\tilde{\zeta}_0 > 0$ on $\overline{U} \setminus V$ and a nonnegative function $\tilde{\zeta}_{m+1} \in C^\infty(\mathbb{R}^n)$ supported on $\mathbb{R}^n \setminus \overline{U}$ such that $\tilde{\zeta}_{m+1} > 0$ on $\mathbb{R}^n \setminus (U \cup V)$ (e.g. again by mollification of the corresponding characteristic functions). Then $\zeta := \sum_{j=0}^{m+1} \tilde{\zeta}_j \in C^\infty(\mathbb{R}^n)$ is positive on $\mathbb{R}^n$ and $\zeta_j := \tilde{\zeta}_j / \zeta$ are the functions we are looking for. □

Now we are ready to show:

Lemma 7.19. *Suppose U has a bounded C^1 boundary, then U has the extension property. Moreover, the extension of a continuous function can be chosen continuous and if U is bounded, the extension can be chosen with compact support.*

Proof. Choose functions ζ_j as in Lemma 7.18 and split $f \in W^{1,p}(U)$ according to $\sum_j f_j$, where $f_j := \zeta_j f$. Then f_0 can be extended to $\mathbb{R}^n$ by setting it equal to 0 outside U. Moreover, f_j can be mapped to $Q_{j,+}$ using ψ_j and extended to Q_j using the symmetric extension. Note that this extension has compact support and so has the pull back $\bar{f}_j$ to U_j; in particular, it can be extended to $\mathbb{R}^n$ by setting it equal to 0 outside U_j. By construction we have $\|\bar{f}_j\|_{W^{1,p}(U_j)} \le C_j \|f_j\|_{W^{1,p}(U_j)}$ and the product rule implies $\|f_j\|_{W^{1,p}(U_j)} \le \tilde{C}_j \|f\|_{W^{1,p}(U)}$. Hence $\bar{f} := \sum_j \bar{f}_j$ is the required extension.

The last claim follows since the symmetric extension of a continuous function is continuous. □

As a first application note that by mollifying an extension we see that we can approximate by functions which are smooth up to the boundary. Moreover, using a suitable cutoff function we can also assume that the approximating functions have compact support.

Corollary 7.20. *Suppose U has the extension property, then $C_c^\infty(\mathbb{R}^n)$ is dense in $W^{1,p}(U)$ for $1 \le p < \infty$.*

Proof. Simply mollify an extension. If U is bounded the extension will have compact support and its mollification will be in $C_c^\infty(\mathbb{R}^n)$. If U is unbounded, multiply the mollification with a cutoff function to get a function from $C_c^\infty(\mathbb{R}^n)$ as in the proof of Lemma 7.10. □

Corollary 7.21. *Suppose U has the extension property, then $W^{1,\infty}(U) = C_b^{0,1}(U)$ with equivalent norms.*

Proof. Since U has the extension property, we can extend f to $W^{1,\infty}(\mathbb{R}^n)$ and hence f is Lipschitz continuous by Lemma 7.15. □

Note that it is sometimes also of interest to look at the corresponding extension problem for $W^{k,p}(U)$ with $k \ge 1$. It can be shown that there is an extension operator $E: W^{k,p}(U) \to W^{k,p}(\mathbb{R}^n)$ provided the boundary satisfies a local Lipschitz condition (see Theorem VI.5 in **[19]** for details).

Next we show that functions in $W^{1,p}$ have boundary values in L^p. This might be surprising since a function from $W^{1,p}(U)$ is only defined almost everywhere and the boundary ∂U is a set of measure zero. Please recall that for U with a C^1 boundary there is a corresponding surface measure dS and by $L^p(\partial U)$ we will always understand $L^p(\partial U, dS)$.

Theorem 7.22. *Suppose U has a bounded C^1 boundary, then there exists a bounded **trace operator***

$$T : W^{1,p}(U) \to L^p(\partial U) \tag{7.26}$$

which satisfies $Tf = f\big|_{\partial U}$ for $f \in C(\overline{U}) \cap W^{1,p}(U)$. Moreover, we have $|Tf| = T|f|$ and for real-valued functions also $(Tf)_\pm = Tf_\pm$.

Proof. In the case $p = \infty$ functions from $W^{1,\infty}(U)$ are Lipschitz continuous by Corollary 7.21 and hence continuous up to the boundary. So there is nothing to do.

Thus we can focus on the case $1 \le p < \infty$. As a preparation we note that by Corollary 7.20 the set $C(\overline{U}) \cap W^{1,p}(U)$ is dense by the previous lemma and that the Gauss–Green theorem continues to hold for $u \in C(\overline{U}, \mathbb{R}^n) \cap W^{1,1}(U, \mathbb{R}^n)$ if U is bounded. To see this choose $u \in C(\overline{U}, \mathbb{R}^n) \cap W^{1,1}(U, \mathbb{R}^n)$ and extend it to a function $\bar{u} \in C_c(\mathbb{R}^n, \mathbb{R}^n) \cap W^{1,1}(\mathbb{R}^n, \mathbb{R}^n)$. Then the Gauss–Green theorem holds for the mollification $u_\varepsilon := \phi_\varepsilon * \bar{u}$ and since we have $u_\varepsilon \to u$ uniformly on $\overline{U}$ as well as $\partial_j u_\varepsilon \to \partial_j u$ in $L^1(U, \mathbb{R}^n)$ the Gauss–Green theorem remains true in the limit $\varepsilon \to 0$.

Now take $f \in C(\overline{U}) \cap W^{1,p}(U)$. As in the proof of Lemma 7.19, using a partition of unity and straightening out the boundary, we can reduce it to the case where $f \in C_c^1(\mathbb{R}^n)$ has compact support $\operatorname{supp}(f) \subset Q$ such that $\partial U \cap \operatorname{supp}(f) \subset \partial \mathbb{R}^n_+$. Then using the Gauss–Green theorem and assuming f real-valued without loss of generality we have (cf. Problem 7.22)

$$\begin{aligned}\int_{\partial U} |f|^p d^{n-1}x &= -\int_{Q_+} (|f|^p)_{x_n} d^n x = -p \int_{Q_+} \operatorname{sign}(f)|f|^{p-1}(\partial_n f) d^n x \\ &\le p\|f\|_p^{p-1} \|\nabla f\|_p,\end{aligned}$$

where we have used Hölders inequality in the last step. Hence the trace operator defined on $C(\overline{U}) \cap W^{1,p}(U)$ is bounded and since the latter set is dense, there is a unique extension to all of $W^{1,p}(U)$.

To see the last claim observe that if $f_n \in C(\overline{U}) \cap W^{1,p}(U) \to f$, then $|f_n| \in C(\overline{U}) \cap W^{1,p}(U) \to |f|$ by Lemma 7.12. Hence $|Tf| = \lim_{n\to\infty} |Tf_n| = \lim_{n\to\infty} T|f_n| = T|f|$. □

Of course this result can also be applied to derivatives:

Corollary 7.23. *Suppose U has a bounded C^1 boundary, then there exists a bounded trace operator*

$$T : W^{k,p}(U) \to W^{k-1,p}(\partial U) \tag{7.27}$$

which satisfies $Tf = f\big|_{\partial U}$ for $f \in C^{k-1}(\overline{U}) \cap W^{k,p}(U)$.

As an application we can extend the Gauss–Green theorem and integration by parts to $W^{1,p}$ vector fields.

Lemma 7.24. *Let U be a bounded C^1 domain in $\mathbb{R}^n$ and $u \in W^{1,p}(U,\mathbb{R}^n)$ a vector field. Then the Gauss–Green formula* (2.65) *holds if the boundary values of u are understood as traces as in the previous theorem. Moreover, the integration by parts formula* (2.67) *also holds for $f \in W^{1,p}(U)$, $g \in W^{1,q}(U)$ with $\frac{1}{p}+\frac{1}{q}=1$.*

Proof. Since U has the extension property, we can extend u to $W^{1,p}(\mathbb{R}^n,\mathbb{R}^n)$. Consider the mollification $u_\varepsilon := \phi_\varepsilon * u$ and apply the Gauss–Green theorem to u_ε. Now let $\varepsilon \to 0$ and observe that the left-hand side converges since $\partial_j u_\varepsilon \to \partial_j u$ in $L^p \subset L^1$. Similarly the right-hand side converges by continuity of the trace operator. The integration by parts formula follows from the Gauss–Green theorem applied to the product fg and employing the product rule. □

Finally we identify the kernel of the trace operator.

Lemma 7.25. *If U is a bounded C^1 domain in $\mathbb{R}^n$, then the kernel of the trace operator is given by $\mathrm{Ker}(T) = W_0^{1,p}(U)$ for $1 \le p < \infty$.*

Proof. Clearly $W_0^{1,p}(U) \subseteq \mathrm{Ker}(T)$. Conversely it suffices to show that $f \in \mathrm{Ker}(T)$ can be approximated by functions from $C_c^\infty(U)$. Using a partition of unity as in the proof of Lemma 7.19 we can assume $U = Q_+$, where Q is a rectangle which is symmetric with respect to reflection. Setting

$$\bar{f}(x) = \begin{cases} f(x), & x \in Q_+, \\ 0, & x \in Q_-, \end{cases}$$

integration by parts using the previous lemma shows

$$\int_Q \bar{f}\partial_j\varphi\, d^n x = \int_{Q_+} f\partial_j\varphi\, d^n x = -\int_{Q_+} (\partial_j f)\varphi\, d^n x, \quad \varphi \in C_c^\infty(Q),$$

that $\bar{f} \in W^{1,p}(Q)$ with $\partial_j\bar{f}(x) = \partial_j f(x)$ for $x \in Q_+$ and $\partial_j \bar{f}(x) = 0$ else. Now consider $f_\varepsilon(x) = (\phi_{\varepsilon/2} * \bar{f})(x - \varepsilon\delta^n) \in C_c^\infty(Q_+)$ with ϕ the standard mollifier. This is the required sequence by Lemma 3.21 and Problem 3.19. □

Problem* 7.21. *Show that $f \in W_0^{k,p}(U)$ can be extended to a function $\bar{f} \in W_0^{k,p}(\mathbb{R}^n)$ by setting it equal to zero outside U. In this case the weak derivatives of $\bar{f}$ are obtained by setting the weak derivatives of f equal to zero outside U.*

Problem* 7.22. *Suppose $\gamma \ge 1$. Show that $f \in W^{1,p}(U)$ implies $|f|^\gamma \in W^{1,p/\gamma}(U)$ with $\partial_j|f|^\gamma = \gamma|f|^{\gamma-1}\partial_j|f|$. (Hint: Lemma 7.12.)*

Problem 7.23. *Suppose U has a bounded C^1 boundary. Show that $W_0^{1,\infty}(U) = C_0^1(U)$. (Hint: Use Lemma 7.18 to reduce it to the case of a straight boundary. Near the straight boundary use a cutoff η_ε as in the proof of Lemma 7.16.)*

Problem 7.24. *Let $1 \le p < \infty$ and U bounded. Show that $Tf = f\big|_{\partial U}$ defined on $C(\overline{U}) \subseteq L^p(U) \to L^p(\partial U)$ is unbounded (and hence has no meaningful extension to $L^p(U)$). (Hint: Take a sequence which equals 1 on the boundary and converges to 0 in the interior.)*

Problem 7.25. *Suppose $u \in H_0^1(B_1(0))$ satisfies $u(x) = -u(x^\star)$, where $x^\star := (x_1, \dots, x_{n-1}, -x_n)$. Show that $u \in H_0^1(B_1(0) \cap \mathbb{R}^n_+)$, where $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n | x_n > 0\}$.*

Problem 7.26. *Consider the punctured ball $U := B_1(0) \setminus \{0\}$. Show that $W_0^{1,p}(U) = W_0^{1,p}(B_1(0))$ and $W^{1,p}(U) = W^{1,p}(B_1(0))$ for $p < n$.*

Problem 7.27. *Let $1 \le p < \infty$ and U bounded. Show that $Tf = f\big|_{\partial U}$ defined on $C(\overline{U}) \subseteq L^p(U) \to L^p(\partial U)$ is unbounded (and hence has no meaningful extension to $L^p(U)$). (Hint: Take a sequence which equals 1 on the boundary and converges to 0 in the interior.)*

Problem 7.28. *Show that $C_0(\overline{U}) \cap W^{1,p}(U) \subseteq W_0^{1,p}(U)$, $1 \le p < \infty$. (Hint: Of course, if U has a nice boundary, this is immediate using traces. For the general case use an approximation based on Lemma 7.12.)*

7.4. Embedding theorems

We have already seen that functions in $W^{1,p}$ are not necessarily continuous (unless $n = 1$). This raises the question in what sense a function from $W^{1,p}$ is better than a function from L^p? For example, is it in L^q for some q other than p? In this respect it is instructive to look at an example which should be understood as a benchmark for the results to follow.

Example 7.12. Let $U := B_1(0)$ and consider $f(x) := |x|^{-\gamma}$. Then by Problem 7.11

$$\partial_j f(x) = -\gamma \frac{x_j}{|x|} |x|^{-\gamma-1},$$

where the factor $\frac{|x_j|}{|x|} \le 1$ is bounded. Hence by Example 2.16 we have $f \in W^{1,p}(U)$ provided $\gamma < \frac{n}{p} - 1$. Since we have $f \in L^q(U)$ provided $\gamma < \frac{n}{q}$ the optimal index for which $f \in L^{p^*}(U)$ is $p^* := \frac{np}{n-p}$ provided $n > p$. If $n < p$, then we have $-\gamma > 1 - \frac{n}{p} > 0$ and hence f is continuous. In fact it will be Hölder continuous of exponent $1 - \frac{n}{p}$.

Of course we can also take higher derivatives into account. To this end, using induction, it is straightforward to verify that

$$\partial_\alpha f(x) = \frac{P_\alpha(x)}{|x|^{|\alpha|}} |x|^{-\gamma-|\alpha|},$$

where P_α is a homogenous polynomial of degree $|\alpha|$. In particular, note that the factor $P_\alpha(x)|x|^{-|\alpha|}$ is bounded. Hence the optimal index for which $f \in L^{p^*}(U)$ provided $f \in W^{k,p}(U)$ is $p_k^* := \frac{np}{n-kp}$ for $n > pk$. For $n < pk$ we will have $f \in C^{k-l-1}$, where $l \in \lfloor \frac{n}{p} \rfloor$ with the highest derivative being Hölder continuous of exponent $1 - \frac{n}{p} + l$. $\diamond$

Theorem 7.26 (Gagliardo[6]–Nirenberg–Sobolev). *Suppose $1 \le p < n$ and $U \subseteq \mathbb{R}^n$ is open. Then there is a continuous embedding $W_0^{1,p}(U) \hookrightarrow L^q(U)$ for all $p \le q \le p^*$, where $\frac{1}{p^*} := \frac{1}{p} - \frac{1}{n}$. Moreover,*

$$\|f\|_{p^*} \le \frac{p(n-1)}{(n-p)} \prod_{j=1}^n \|\partial_j f\|_p^{1/n} \le \frac{p(n-1)}{n(n-p)} \sum_{j=1}^n \|\partial_j f\|_p. \tag{7.28}$$

Proof. It suffices to prove the case $q = p^*$ since the rest follows from interpolation (Problem 3.13). Moreover, by density it suffices to prove the inequality for $f \in C_c^\infty(\mathbb{R}^n)$. In this respect note that if you have a sequence $f_n \in C_c^\infty(U)$ which converges to some f in $W_0^{1,p}(U)$, then by (7.28) this sequence will also converge in $L^{p^*}(U)$ and by considering pointwise convergent subsequences both limits agree.

We start with the case $p = 1$ and observe

$$|f(x)| = \left| \int_{-\infty}^{x_1} \partial_1 f(r, \tilde{x}_1) dr \right| \le \int_{-\infty}^{\infty} |\partial_1 f(r, \tilde{x}_1)| dr,$$

where we denote by $\tilde{x}_j := (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$ the vector obtained from x with the j'th component dropped. Denote by $f_1(\tilde{x}_1)$ the right-hand side of the above inequality and apply the same reasoning to the other coordinate directions to obtain

$$|f(x)|^n \le \prod_{j=1}^n f_j(\tilde{x}_j).$$

Now we claim that if $f_j \in L^1(\mathbb{R}^{n-1})$, then

$$\Big\| \prod_{j=1}^n f_j(\tilde{x}_j)^{\frac{1}{n-1}} \Big\|_{L^1(\mathbb{R}^n)} \le \prod_{j=1}^n \|f_j\|_{L^1(\mathbb{R}^{n-1})}^{\frac{1}{n-1}}.$$

For $n = 2$ this is just Fubini and hence we can use induction. To this end fix the last coordinate x_{n+1} and apply Hölder's inequality and the induction

[6] Emilio Gagliardo (1930–2008), Italian Mathematician

hypothesis to obtain

$$\int_{\mathbb{R}^n} \prod_{j=1}^{n+1} |f_j(\tilde{x}_j)|^{\frac{1}{n}} d^n x \le \|f_{n+1}^{1/n}\|_{L^n(\mathbb{R}^n)} \Big\| \prod_{j=1}^{n} |f_j(\tilde{x}_j)|^{\frac{1}{n}} \Big\|_{L^{n/(n-1)}(\mathbb{R}^n)}$$
$$= \|f_{n+1}\|_{L^1(\mathbb{R}^n)}^{1/n} \Big\| \prod_{j=1}^{n} |f_j(\tilde{x}_j)|^{\frac{1}{n-1}} \Big\|_{L^1(\mathbb{R}^n)}^{1-\frac{1}{n}} \le \|f_{n+1}\|_{L^1(\mathbb{R}^n)}^{1/n} \prod_{j=1}^{n} \|f_j(\tilde{x}_j)\|_{L^1(\mathbb{R}^{n-1})}^{1/n}.$$

Now integrate this inequality with respect to the missing variable x_{n+1} and use the iterated Hölder inequality (Problem 3.11 with $r = 1$ and $p_j = n$) to obtain the claim.

Moreover, applying this to our situation where $\|f_j\|_{L^1(\mathbb{R}^{n-1})} = \|\partial_j f\|_1$ we obtain

$$\|f\|_{n/(n-1)}^{n/(n-1)} \le \prod_{j=1}^{n} \|\partial_j f\|_1^{\frac{1}{n-1}}$$

which is precisely (7.28) for the case $p = 1$ (the second inequality in (7.28) is just the inequality of arithmetic and geometric means). To see the case of general p let $f \in C_c^\infty(\mathbb{R}^n)$ and apply the case $p = 1$ to $f \to |f|^\gamma$ for $\gamma > 1$ to be determined and recall Problem 7.22. Then

$$\left(\int_{\mathbb{R}^n} |f|^{\frac{\gamma n}{n-1}} d^n x \right)^{\frac{n-1}{n}} \le \prod_{j=1}^{n} \left(\int_{\mathbb{R}^n} \big|\partial_j |f|^\gamma \big| d^n x \right)^{1/n} \tag{7.29}$$
$$= \gamma \prod_{j=1}^{n} \left(\int_{\mathbb{R}^n} |f|^{\gamma-1} |\partial_j f| d^n x \right)^{1/n} \le \gamma \| |f|^{\gamma-1} \|_{p/(p-1)} \prod_{j=1}^{n} \|\partial_j f\|_p^{1/n},$$

where we have used Hölder in the last step. Now we choose $\gamma := \frac{p(n-1)}{n-p} > 1$ such that $\frac{\gamma n}{n-1} = \frac{(\gamma-1)p}{p-1} = p^*$, which gives the general case. □

Note that a simple scaling argument (Problem 7.30) shows that (7.28) can only hold for p^*. Furthermore, using an extension operator this result also extends to $W^{1,p}(U)$:

Corollary 7.27. *Suppose U has the extension property and $1 \le p < n$, then there is a continuous embedding $W^{1,p}(U) \hookrightarrow L^q(U)$ for every $p \le q \le p^*$, where $\frac{1}{p^*} = \frac{1}{p} - \frac{1}{n}$.*

Proof. Let $\tilde{f} = Ef \in W^{1,p}(\mathbb{R}^n)$ be an extension of $f \in W^{1,p}(U)$. Then $\|f\|_{L^q(U)} \le \|\tilde{f}\|_{L^q(\mathbb{R}^n)} \le C_p \|f\|_{W^{1,p}(\mathbb{R}^n)} \le C_p C_{U,p} \|f\|_{W^{1,p}(U)}$, where we have used $W_0^{1,p}(\mathbb{R}^n) = W^{1,p}(\mathbb{R}^n)$ (Lemma 7.10). □

Note that involving the extension operator implies that we need the full $W^{1,p}$ norm to bound the L^{p^*} norm. A constant function shows that indeed an inequality involving only the derivatives on the right-hand side cannot hold on bounded domains (cf. also Theorem 7.38).

In the borderline case $p = n$ one has $p^* = \infty$, however, the example in Problem 7.31 shows that functions in $W^{1,n}$ can be unbounded if $n > 1$ (for $n = 1$ we have already seen in Example 7.8 that we have a continuous embedding $W^{1,1}(a,b) \hookrightarrow C[a,b]$). Nevertheless, we have at least the following result:

Lemma 7.28. *Suppose $p = n$ and $U \subseteq \mathbb{R}^n$ is open. Then there is a continuous embedding $W_0^{1,n}(U) \hookrightarrow L^q(U)$ for every $n \le q < \infty$.*

Proof. As before it suffices to establish $\|f\|_q \le C\|f\|_{W^{1,n}}$ for $f \in C_c^\infty(\mathbb{R}^n)$. To this end we employ (7.29) with $p = n$ implying

$$\|f\|_{\gamma n/(n-1)}^\gamma \le \gamma \|f\|_{(\gamma-1)n/(n-1)}^{\gamma-1} \prod_{j=1}^n \|\partial_j f\|_n^{1/n} \le \frac{\gamma}{n} \|f\|_{(\gamma-1)n/(n-1)}^{\gamma-1} \sum_{j=1}^n \|\partial_j f\|_n.$$

Using Young's inequality (Problem 7.33), $\alpha^{(\gamma-1)/\gamma}\beta^{1/\gamma} \le \frac{\gamma-1}{\gamma}\alpha + \frac{1}{\gamma}\beta$ for nonnegative numbers $\alpha, \beta \ge 0$, this gives

$$\|f\|_{\gamma n/(n-1)} \le C\Big(\|f\|_{(\gamma-1)n/(n-1)} + \sum_{j=1}^n \|\partial_j f\|_n\Big).$$

Now choosing $\gamma = n$ we get $\|f\|_{n^2/(n-1)} \le C\|f\|_{W^{1,n}}$ and by interpolation (Problem 3.13) the claim holds for $q \in [n, n\frac{n}{n-1}]$. So we can choose $\gamma = n+1$ to get the claim for $q \in [n, (n+1)\frac{n}{n-1}]$ and iterating this procedure finally gives the claim for $q \in [n, (n+k)\frac{n}{n-1}]$, which establishes the result. □

Corollary 7.29. *Suppose U has the extension property and $p = n$, then there is a continuous embedding $W^{1,n}(U) \hookrightarrow L^q(U)$ for every $n \le q < \infty$.*

In the case $p > n$ functions from $W^{1,p}$ will be continuous (in the sense that there is a continuous representative). In fact, they will even be bounded Hölder continuous functions and hence are continuous up to the boundary (cf. Theorem 7.2 and the discussion after this theorem).

Theorem 7.30 (Morrey)**.** *Suppose $n < p \le \infty$ and $U \subseteq \mathbb{R}^n$ is open. There is a continuous embedding $W_0^{1,p}(U) \hookrightarrow C_0^{0,\gamma}(\overline{U})$, where $\gamma = 1 - \frac{n}{p}$. Here $C_0^{0,\gamma}(\overline{U}) := C_b^{0,\gamma}(\overline{U}) \cap C_0(U)$ is the space of Hölder continuous functions vanishing at the boundary.*

Proof. In the case $p = \infty$ there is nothing to do, since $W_0^{1,\infty}(U) \subseteq C_0^1(U)$ and $[f]_1 \le \|\nabla f\|_\infty$. Hence we can assume $n < p < \infty$. Moreover, as before, by density we can assume $f \in C_c^\infty(\mathbb{R}^n)$.

We begin by considering a cube Q of side length r containing 0. Then, for $x \in Q$ and $\bar{f} = r^{-n}\int_Q f(x)d^n x$ we have

$$\bar{f} - f(0) = r^{-n}\int_Q \big(f(x) - f(0)\big)d^n x = r^{-n}\int_Q\int_0^1 \frac{d}{dt}f(tx)dt\, d^n x$$

and hence

$$\begin{aligned}|\bar{f} - f(0)| &\le r^{-n}\int_Q\int_0^1 |\nabla f(tx)||x| dt\, d^n x \le r^{1-n}\int_Q\int_0^1 |\nabla f(tx)| dt\, d^n x\\ &= r^{1-n}\int_0^1\int_{tQ} |\nabla f(y)|\frac{d^n y}{t^n}\, dt \le r^{1-n}\int_0^1 \|\nabla f\|_{L^p(tQ)}\frac{|tQ|^{1-1/p}}{t^n}\, dt\\ &\le \frac{r^\gamma}{\gamma}\|\nabla f\|_{L^p(Q)}\end{aligned}$$

where we have used Hölder's inequality in the fourth step. By a translation this gives

$$|\bar{f} - f(x)| \le \frac{r^\gamma}{\gamma}\|\nabla f\|_{L^p(Q)}$$

for any cube Q of side length r containing x and combining the corresponding estimates for two points we obtain

$$|f(x) - f(y)| \le \frac{2\|\nabla f\|_{L^p(Q)}}{\gamma}|x-y|^\gamma \tag{7.30}$$

for any cube containing both x and y (note that we can choose the side length of Q to be $r = \max_{1\le j\le n}|x_j - y_j| \le |x-y|$). Since we can of course replace $L^p(Q)$ by $L^p(\mathbb{R}^n)$ we get Hölder continuity of f. Moreover, taking a cube of side length $r = 1$ containing x we get (using again Hölder)

$$|f(x)| \le |\bar{f}| + \frac{2\|\nabla f\|_{L^p(Q)}}{\gamma} \le \|f\|_{L^p(Q)} + \frac{2\|\nabla f\|_{L^p(Q)}}{\gamma} \le C\|f\|_{W^{1,p}(\mathbb{R}^n)}$$

establishing the theorem. □

Corollary 7.31. *Suppose U has the extension property and $n < p \le \infty$, then there is a continuous embedding $W^{1,p}(U) \hookrightarrow C_b^{0,\gamma}(\overline{U})$, where $\gamma = 1 - \frac{n}{p}$.*

Proof. Let $\tilde{f} = Ef \in W^{1,p}(\mathbb{R}^n)$ be an extension of $f \in W^{1,p}(U)$. Then $\|f\|_{C_b^{0,\gamma}(\overline{U})} \le \|\tilde{f}\|_{C_b^{0,\gamma}(\mathbb{R}^n)} \le C_p\|f\|_{W^{1,p}(\mathbb{R}^n)} \le C_p C_{U,p}\|f\|_{W^{1,p}(U)}$, where we have used $W_0^{1,p}(\mathbb{R}^n) = W^{1,p}(\mathbb{R}^n)$ and Morrey's theorem if $p < \infty$ and Lemma 7.15 if $p = \infty$. □

Example 7.13. The example from Problem 7.32 shows that for a domain with a cusp, functions from $W^{1,p}$ might be unbounded (and hence in particular not in $C^{1,\gamma}$) even for $p > n$. ⋄

Example 7.14. For $p = \infty$ this embedding is surjective in case of a convex domain (Lemma 7.15) or a domain with the extension property (Corollary 7.21). However, for $n < p < \infty$ this is not the case. To see this consider the Takagi function (Problem 4.27) which is in $C^{1,\gamma}[0,1]$ for every $\gamma < 1$ but not absolutely continuous (not even of bounded variation) and hence not in $W^{1,p}(0,1)$ for any $1 \le p \le \infty$. Note that this example immediately extends to higher dimensions by considering $f(x) = b(x_1)$ on the unit cube. ⋄

As a consequence of the proof we also get that for $n < p$ Sobolev functions are differentiable a.e.

Lemma 7.32. *Suppose $n < p \le \infty$ and $U \subseteq \mathbb{R}^n$ is open. Then $f \in W^{1,p}_{loc}(U)$ is differentiable a.e. and the a.e. derivative equals the weak derivative.*

Proof. Since $W^{1,\infty}_{loc} \subseteq W^{1,p}_{loc}$ for any $p < \infty$ we can assume $n < p < \infty$. Let $x \in U$ be an L^p Lebesgue point of the gradient, that is,

$$\lim_{r \downarrow 0} \frac{1}{|Q_r(x)|} \int_{Q_r(x)} |\nabla f(x) - \nabla f(y)|^p d^n y = 0,$$

where $Q_r(x)$ is a cube of side length r containing x. Now let $y \in Q_r(x)$ and $r = |y - x|$ (by shrinking the cube w.l.o.g.). Then replacing $f(y) \to f(y) - f(x) - \nabla f(x) \cdot (y - x)$ in (7.30) we obtain

$$\big|f(y) - f(x) - \nabla f(x) \cdot (y - x)\big| \le \frac{2}{\gamma} |x - y|^\gamma \left(\int_{Q_r(x)} |\nabla f(x) - \nabla f(z)|^p d^n z \right)^{1/p}$$
$$= \frac{2}{\gamma} |x - y| \left(\frac{1}{|Q_r(x)|} \int_{Q_r(x)} |\nabla f(x) - \nabla f(z)|^p d^n z \right)^{1/p}$$

and, since x is an L^p Lebesgue point of the gradient, the right-hand side is $o(|x - y|)$, that is, f is differentiable at x and its gradient equals its weak gradient. □

Note that since by Lemma 7.15 every locally Lipschitz continuous function is locally $W^{1,\infty}$, we obtain as an immediate consequence:

Theorem 7.33 (Rademacher)**.** *Every locally Lipschitz continuous function is differentiable almost everywhere.*

So far we have only looked at first order derivatives. However, we can also cover the case of higher order derivatives by repeatedly applying the above results to the fact that $\partial_j f \in W^{k-1,p}(U)$ for $f \in W^{k,p}(U)$.

Theorem 7.34. *Suppose $U \subseteq \mathbb{R}^n$ is open and $1 \le p \le \infty$. There are continuous embeddings*

$$W_0^{k,p}(U) \hookrightarrow L^q(U), \quad q \in [p, p_k^*] \text{ if } \frac{1}{p_k^*} = \frac{1}{p} - \frac{k}{n} > 0,$$
$$W_0^{k,p}(U) \hookrightarrow L^q(U), \quad q \in [p, \infty) \text{ if } \frac{1}{p} = \frac{k}{n},$$
$$W_0^{k,p}(U) \hookrightarrow C_0^{k-l-1,\gamma}(\overline{U}), \quad l = \lfloor \frac{n}{p} \rfloor, \begin{cases} \gamma = 1 - \frac{n}{p} + l, & \frac{n}{p} \notin \mathbb{N}_0, \\ \gamma \in [0,1), & \frac{n}{p} \in \mathbb{N}_0, \end{cases} \text{ if } \frac{1}{p} < \frac{k}{n}.$$

If in addition $U \subseteq \mathbb{R}^n$ has the extension property, then there are continuous embeddings

$$W^{k,p}(U) \hookrightarrow L^q(U), \quad q \in [p, p_k^*] \text{ if } \frac{1}{p_k^*} = \frac{1}{p} - \frac{k}{n} > 0,$$
$$W^{k,p}(U) \hookrightarrow L^q(U), \quad q \in [p, \infty) \text{ if } \frac{1}{p} = \frac{k}{n},$$
$$W^{k,p}(U) \hookrightarrow C_b^{k-l-1,\gamma}(\overline{U}), \quad l = \lfloor \frac{n}{p} \rfloor, \begin{cases} \gamma = 1 - \frac{n}{p} + l, & \frac{n}{p} \notin \mathbb{N}_0, \\ \gamma \in [0,1), & \frac{n}{p} \in \mathbb{N}_0, \end{cases} \text{ if } \frac{1}{p} < \frac{k}{n}.$$

Proof. If $\frac{1}{p} > \frac{k}{n}$ we apply Theorem 7.26 to successively conclude $\|\partial^\alpha f\|_{L^{p_j^*}} \le C\|f\|_{W_0^{k,p}}$ for $|\alpha| \le k - j$ for $j = 1, \dots, k$. If $\frac{1}{p} = \frac{k}{n}$ we proceed in the same way but use Lemma 7.28 in the last step. If $\frac{1}{p} < \frac{k}{n}$ we first apply Theorem 7.26 l times as before. If $\frac{n}{p}$ is not an integer we then apply Theorem 7.30 to conclude $\|\partial^\alpha f\|_{C_0^{0,\gamma}} \le C\|f\|_{W_0^{k,p}}$ for $|\alpha| \le k - l - 1$. If $\frac{n}{p}$ is an integer, we apply Theorem 7.26 $l-1$ times and then Lemma 7.28 once to conclude $\|\partial^\alpha f\|_{L^q} \le C\|f\|_{W_0^{k,p}}$ for any $q \in [p,\infty)$ for $|\alpha| \le k - l$. Hence we can apply Theorem 7.30 to conclude $\|\partial^\alpha f\|_{C_0^{0,\gamma}} \le C\|f\|_{W_0^{k,p}}$ for any $\gamma \in [0,1)$ for $|\alpha| \le k - l - 1$.

The second part follows analogously using the corresponding results for domains with the extension property. □

Note that for $p = 1$ we have a slightly stronger result $W_0^{n,1}(U) \hookrightarrow C_0(U)$ in the borderline case $k = n$ — see Problem 7.38.

Moreover, for $q \in [p, p^*)$ the embedding is even compact (it fails for $q = p^*$ — see Problem 7.34).

Theorem 7.35 (Rellich–Kondrachov[7]). *Suppose $U \subseteq \mathbb{R}^n$ is open and bounded and $1 \le p \le \infty$. Then there are compact embeddings*

$$W_0^{1,p}(U) \hookrightarrow L^q(U), \quad q \in [p, p^*), \quad \frac{1}{p^*} = \frac{1}{p} - \frac{1}{n}, \text{ if } p \le n,$$
$$W_0^{1,p}(U) \hookrightarrow C_0(\overline{U}), \text{ if } p > n.$$

If in addition $U \subseteq \mathbb{R}^n$ has the extension property, then there are compact embeddings

$$W^{1,p}(U) \hookrightarrow L^q(U), \quad q \in [p, p^*), \quad \frac{1}{p^*} = \frac{1}{p} - \frac{1}{n}, \text{ if } p \le n,$$
$$W^{1,p}(U) \hookrightarrow C(\overline{U}), \text{ if } p > n.$$

Proof. The case $p > n$ follows from $W_0^{1,p}(U) \hookrightarrow C_0^{0,\gamma}(\overline{U}) \hookrightarrow C_0(\overline{U})$, where the first embedding is continuous by Theorem 7.30 and the second is compact by Theorem 7.3. Similarly in the second case using $W^{1,p}(U) \hookrightarrow C_b^{0,\gamma}(\overline{U}) \hookrightarrow C(\overline{U})$, where the first embedding is continuous by Corollary 7.31.

Next consider the case $p < n$. Let $F \subset W^{1,p}(U)$ (or $F \subset W_0^{1,p}(U)$) be a bounded subset. Using an extension operator (or Problem 7.21) we can assume that $F \subset W^{1,p}(\mathbb{R}^n)$ with $\mathrm{supp}(f) \subseteq V$ for all $f \in F$ and some fixed set V. By Lemma 7.14 (applied on $\mathbb{R}^n$) we have

$$\|T_a f - f\|_p \le |a| \|\nabla f\|_p$$

and using the interpolation inequality from Problem 3.13 and Theorem 7.34 we have

$$\|T_a f - f\|_q \le \|T_a f - f\|_p^{1-\theta} \|T_a f - f\|_{p^*}^{\theta} \le |a|^{1-\theta} \|\nabla f\|_p^{1-\theta} C^\theta \|f\|_{1,p}^{\theta},$$

where $\frac{1}{q} = \frac{1-\theta}{p} + \frac{\theta}{p^*}$, $\theta \in [0,1]$. Hence F is relatively compact by Theorem 3.16. In the case $p = n$, we can replace p^* by any value larger than q. $\square$

Since for bounded U the embedding $C(\overline{U}) \hookrightarrow L^p(U)$ is continuous, we obtain:

Corollary 7.36. *Under the assumptions of the above theorem the embeddings $W_0^{k+1,p}(U) \hookrightarrow W_0^{k,p}(U)$ and $W^{k+1,p}(U) \hookrightarrow W^{k,p}(U)$ are compact.*

Proof. By the Rellich–Kondrachov theorem the embedding $W^{1,p}(U) \hookrightarrow L^p(U)$ is compact. Hence, given a bounded sequence in $W^{k+1,p}(U)$ we can find a subsequence for which all partial derivatives of order up to k converge in $L^p(U)$. Hence this sequence converges in $W^{k,p}(U)$ by Corollary 7.6. $\square$

[7] Vladimir Iosifovich Kondrashov (1909–1971), Russian mathematicia

Example 7.15. The Rellich–Kondrachov theorem fails for $q = p^*$. To see this choose a nonzero function $f \in W_0^{1,p}(U)$ with compact support in some small ball. Now consider $f_\varepsilon(x) := \varepsilon^{-n/p^*} f(x/\varepsilon)$. Then $\|f_\varepsilon\|_{1,p} \le \|f\|_{1,p}$ and $\|f_\varepsilon\|_{p^*} = \|f\|_{p^*}$. If f_ε had a convergent subsequence in $L^{p^*}(U)$, this subsequence must converge to 0 since $f_\varepsilon(x) \to 0$ a.e., a contradiction. $\diamond$

Example 7.16. The Rellich–Kondrachov theorem fails on $\mathbb{R}^n$. To see this choose $\varphi \in C_c^\infty(\mathbb{R}^n)$ with support in $B_{1/2}(0)$ and consider $F = \{\varphi_k := \varphi(. - k\delta^1) | k \in \mathbb{N}\}$. Now note that both the $W^{1,p}$ as well as the L^q norm of φ_k are independent of k and two different functions have disjoint supports. So there is no way to extract a convergent subsequence and an extra condition is needed. $\diamond$

Theorem 7.37. *Let $1 \le p \le n$. A set $F \subseteq W^{1,p}(\mathbb{R}^n)$ is relatively compact in $L^q(\mathbb{R}^n)$ for every $q \in [p, p^*)$ if F is bounded and for every $\varepsilon > 0$ there is some $r > 0$ such that $\|(1 - \chi_{B_r(0)})f\|_p < \varepsilon$ for all $f \in F$.*

Proof. Condition (i) of Theorem 3.16 is verified literally as in the previous theorem. Similarly condition (ii) follows from interpolation since $\|(1 - \chi_{B_r(0)})f\|_q \le \|(1-\chi_{B_r(0)})f\|_p^{1-\theta} \|(1-\chi_{B_r(0)})f\|_{p^*}^\theta \le \|(1-\chi_{B_r(0)})f\|_p^{1-\theta} \|f\|_{p^*}^\theta$. $\square$

Note that this extra condition might come for free in case of radial symmetry (Problem 7.35).

As a consequence of Theorem 7.35 we also can get an important inequality.

Theorem 7.38 (Poincaré inequality)**.** *Let $U \subset \mathbb{R}^n$ be open and bounded. Then for $f \in W_0^{1,p}(U)$, $1 \le p \le \infty$, we have*

$$\|f\|_p \le C \|\nabla f\|_p. \tag{7.31}$$

If in addition U is a connected subset with the extension property, then for $f \in W^{1,p}(U)$, $1 \le p \le \infty$, we have

$$\|f - (f)_U\|_p \le C \|\nabla f\|_p, \tag{7.32}$$

where $(f)_U := \frac{1}{|U|} \int_U f\, d^n x$ is the average of f over U.

Proof. We begin with the second case and argue by contradiction. If the claim were wrong we could find a sequence of functions $f_m \in W^{1,p}(U)$ such that $\|f_m - (f_m)_U\|_p > m \|\nabla f_m\|_p$. hence the function $g_m := \|f_m - (f_m)_U\|_p^{-1}(f_m - (f_m)_U)$ satisfies $\|g_m\|_p = 1$, $(g_m)_U = 0$, and $\|\nabla g_m\|_p < \frac{1}{m}$. In particular, the sequence is bounded and by Corollary 7.36 we can assume $g_m \to g$ in $L^p(U)$ without loss of generality. Moreover, $\|g\|_p = 1$, $(g)_U = 0$, and

$$\int_U g\partial_j\varphi d^n x = \lim_{m\to\infty}\int_U g_m\partial_j\varphi d^n x = -\lim_{m\to\infty}\int_U (\partial_j g_m)\varphi d^n x = 0.$$

That is, $\partial_j g = 0$ and since U is connected, g must be constant on U by Lemma 7.11 (vi). Moreover, $(g)_U = 0$ implies $g = 0$ contradicting $\|g\|_p = 1$.

To see the first case we proceed similarly to find a sequence $g_m := \|f_m\|_p^{-1} f_m$ producing a limit such that $\|g\|_p = 1$ and $\partial_j g = 0$. Now take a ball $B := B_r(0)$ containing U such that $B\setminus U$ has positive Lebesgue measure. Observe that we can extend f_m to $\bar f_m \in W^{1,p}(B)$ by setting it equal to 0 outside U which will give a corresponding sequence $\bar g_m := \|f_m\|_p^{-1}\bar f_m$ and a corresponding limit $\bar g$. Since B is connected we again get that $\bar g$ is constant on B and since $\bar g$ vanishes on $B\setminus U$ it must vanish on all of B contradicting $\|\bar g\|_p = 1$. □

Example 7.17. Using the Poincaré inequality we can shed some further light on the case $f\in W^{1,n}(\mathbb{R}^n)$ from Lemma 7.28. First note that a simple scaling shows that the constant C_r for a ball of radius r in the Poincaré inequality is given by $C_r = C_1 r$. Hence using Poincaré's and Hölder's inequalities we obtain

$$\begin{aligned}\frac{1}{|B_r|}\int_{B_r(x)} |f(y)-(f)_{B_r(x)}|d^n y &\le C_1 r\int_{B_r(x)} |\nabla f(y)|\frac{d^n y}{|B_r|}\\ &\le C_1 r\left(\int_{B_r(x)} |\nabla f(y)|^n\frac{d^n y}{|B_r|}\right)^{1/n} \le \frac{C_1}{V_n^{1/n}}\|\nabla f\|_n.\end{aligned}$$

Locally integrable functions for which the left-hand side is bounded are called functions of **bounded mean oscillation** $\mathrm{BMO}(\mathbb{R}^n)$ and one sets

$$\|f\|_{\mathrm{BMO}} = \sup_{x,r}\frac{1}{|B_r|}\int_{B_r(x)} |f(y)-(f)_{B_r(x)}|d^n y.$$

It is straightforward to verify that this is a semi-norm and $\|f\|_{\mathrm{BMO}} = 0$ if and only if f is constant. ⋄

Finally it is often important to know when $W^{1,p}(U)$ is an algebra: By the product rule we have $\partial_j(fg) = (\partial_j f)g + f(\partial_j g)$ and for this to be in L^p we need that f, g are bounded which follows from Morrey's inequality if $n < p$. Working a bit harder one can even show:

Theorem 7.39. *Suppose $U\subseteq\mathbb{R}^n$ is open. If $\frac{1}{p} < \frac{k}{n}$, then $W_0^{k,p}(U)$ is a Banach algebra with*

$$\|fg\|_{k,p} \le C\|f\|_{k,p}\|g\|_{k,p}. \tag{7.33}$$

If U is bounded and has the extension property, the result also holds for $W^{k,p}(U)$.

Proof. First of all it suffices to show the inequality for the case when f and g are $C^\infty \cap W^{k,p}$. Moreover, by Leibniz' rule (Problem 7.14) it suffices to estimate $\|(\partial_\alpha f)(\partial_\beta g)\|_p$ for $|\alpha| + |\beta| \le k$. To this end we will use the generalized Hölder inequality (Problem 3.10) and hence we need to find $1 \le q_\alpha, q_\beta \le \infty$ with $\frac{1}{p} = \frac{1}{q_\alpha} + \frac{1}{q_\beta}$ such that $W^{m-|\alpha|,p} \hookrightarrow L^{q_\alpha}$ and $W^{m-|\beta|,p} \hookrightarrow L^{q_\beta}$.

Let l be the largest integer such that $\frac{1}{p} < \frac{k-l}{n}$. Then Theorem 7.34 allows us to choose $q_\alpha = \infty$, $q_\beta = p$ for $|\alpha| \le l$ and similarly $q_\alpha = p$, $q_\beta = \infty$ for $|\beta| \le l$. Otherwise, that is if $\frac{1}{p} \ge \frac{k-|\alpha|}{n}$ and $\frac{1}{p} \ge \frac{k-|\beta|}{n}$ then Theorem 7.34 imposes the restrictions $\frac{1}{q_\alpha} \ge \frac{1}{p} - \frac{k-|\alpha|}{n}$ and $\frac{1}{q_\beta} \ge \frac{1}{p} - \frac{k-|\beta|}{n}$. Hence $\frac{1}{q_\alpha} + \frac{1}{q_\beta} \ge \frac{1}{p} - \left(\frac{k}{n} - \frac{1}{p}\right)$ and we can find the required indices. □

Problem 7.29. *Show that for $f \in H^1_0((a,b))$ we have*

$$\|f\|_\infty^2 \le 2\|f\|_2 \|f'\|_2.$$

Show that the inequality continues to hold if $f \in H^1(\mathbb{R})$ or $f \in H^1((0,\infty))$. (Hint: Start by differentiating $|f(x)|^2$.)

Problem* 7.30. *Show that the inequality $\|f\|_q \le C\|\nabla f\|_p$ for $f \in W^{1,p}(\mathbb{R}^n)$ can only hold for $q = \frac{np}{n-p}$. (Hint: Consider $f_\lambda(x) = f(\lambda x)$.)*

Problem* 7.31. *Show that $f(x) := \log\log(1 + \frac{1}{|x|})$ is in $W^{1,n}(B_1(0))$ if $n > 1$. (Hint: Problem 7.11.)*

Problem* 7.32. *Consider $U := \{(x,y) \in \mathbb{R}^2 | 0 < x, y < 1,\ x^\beta < y\}$ and $f(x,y) := y^{-\alpha}$ with $\alpha, \beta > 0$. Show $f \in W^{1,p}(U)$ for $p < \frac{1+\beta}{(1+\alpha)\beta}$. Now observe that for $0 < \beta < 1$ and $\alpha < \frac{1-\beta}{2\beta}$ we have $2 < \frac{1+\beta}{(1+\alpha)\beta}$.*

Problem* 7.33. *Prove* ***Young's inequality***

$$\alpha^{1/p}\beta^{1/q} \le \frac{1}{p}\alpha + \frac{1}{q}\beta, \qquad \frac{1}{p} + \frac{1}{q} = 1, \quad \alpha, \beta \ge 0.$$

Show that equality occurs precisely if $\alpha = \beta$. (Hint: Take logarithms on both sides.)

Problem* 7.34. *Let $U = B_1(0) \subset \mathbb{R}^n$ and consider*

$$u_m(x) = \begin{cases} m^{\frac{n}{p}-1}(1 - m|x|), & |x| < \frac{1}{m}, \\ 0, & \text{else.} \end{cases}$$

Show that u_m is bounded in $W^{1,p}(U)$ for $1 \le p < n$ but has no convergent subsequence in $L^{p^*}(U)$. (Hint: The beta integral from Problem 2.27 might be useful.)

Problem 7.35 (Strauss lemma). *Show the Strauss inequality*

$$\|r^{(n-1)/2}\tilde{f}(r)\|_\infty \le 2S_n^{-1}\|f\|_2\|\partial f\|_2$$

for a radial function $f \in H^1_{\mathrm{rad}}(\mathbb{R}^n)$ (cf. Problem 7.20). Use this to show

$$\|(1-\chi_{B_r(0)})f\|_p \le \frac{C}{r^{(n-1)(p-2)/(2p)}}\|f\|_{H^1}, \quad p \ge 2,$$

and conclude that $H^1_{\mathrm{rad}}(\mathbb{R}^n)$ is compactly embedded into $L^p(\mathbb{R}^n)$ for $p \in (2, \frac{2n}{n-2})$ if $n \ge 2$.

Problem 7.36 (Ehrling's lemma)**.** *Let X, Y, and Z be Banach spaces. Assume X is compactly embedded into Y and Y is continuously embedded into Z. Show that for every $\varepsilon > 0$ there exists some $C(\varepsilon)$ such that*

$$\|x\|_Y \le \varepsilon\|x\|_X + C(\varepsilon)\|x\|_Z.$$

Problem 7.37. *Suppose $U \subseteq \mathbb{R}^n$ is bounded and has the extension property. Show that there exists a constant C such that*

$$\|f\|_{k,p} \le C\left(\sum_{|\alpha|=k}\|\partial_\alpha f\|_p + \|f\|_p\right)$$

(Hint: Problem 7.36 and Corollary 7.36.)

Problem 7.38. *Let $U \subseteq \mathbb{R}^n$. Show that there is a bounded embedding $W^{n,1}_0(U) \hookrightarrow C_0(U)$ satisfying*

$$\|f\|_\infty \le \|\partial_{(1,\dots,1)}f\|_1.$$

Problem 7.39. *Let U be a bounded domain with a C^1 boundary and $1 \le p < \infty$. Show that for every $a > 0$ there is a constant C such that*

$$\int_U |f|^p d^n x \le C\left(\int_U |\nabla f|^p d^n x + a\int_{\partial U}|f|^p dS\right), \qquad f \in W^{1,p}(U).$$

Problem 7.40. *Show that item (iv) From Lemma 7.11 holds for Lipschitz continuous η. (Hint: Lemma 7.32.)*

7.5. Lipschitz domains

It turns out that Lemma 7.32 is the key to extending several results from C^1 to Lipschitz domains. Of course the first step is to understand a change of variables in case the transformation $\psi : U \to V$ is **bi-Lipschitz**, that is, bijective such that both the map and its inverse are Lipschitz. In other words, there is a constant C such that

$$\frac{1}{C} \le \frac{|\psi(x)-\psi(y)|}{|x-y|} \le C, \qquad x \ne y \in U. \tag{7.34}$$

By Lemma 7.32 the Jacobi matrix $\frac{\partial \psi}{\partial x}$ of a Lipschitz function exits a.e. and (since it is defined as a limit) is measurable. In particular, the same is true for the Jacobi determinant

$$J_\psi := \det(\frac{\partial \psi}{\partial x}), \tag{7.35}$$

which is bounded from above and below by positive constants. With this in mind we obtain:

Theorem 7.40 (change of variables). *Let $U, V \subseteq \mathbb{R}^n$ and suppose $\psi : U \to V$ is bi-Lipschitz. Then*

$$(\psi^{-1})_\star d^n x = |J_\psi(x)| d^n x. \tag{7.36}$$

In particular,

$$\int_U f(\psi(x))|J_\psi(x)| d^n x = \int_V f(y) d^n y \tag{7.37}$$

whenever f is nonnegative or integrable over V.

Proof. We literally follow the proof of Theorem 2.17 and we will just point out the differences. It suffices to show

$$\int_{\psi(R)} d^n y = \int_R |J_\psi(x)| d^n x$$

for every bounded open rectangle $R \subseteq U$. By Theorem 1.3 it will then follow for characteristic functions and thus for arbitrary functions by the very definition of the integral.

To this end we consider the integral

$$I_\varepsilon := \int_{\psi(R)} \int_R |J_\psi(\psi^{-1}(y))| \varphi_\varepsilon(\psi(z) - y) d^n z\, d^n y$$

Here $\varphi := V_n^{-1} \chi_{B_1(0)}$ and $\varphi_\varepsilon(y) := \varepsilon^{-n} \varphi(\varepsilon^{-1} y)$, where V_n is the volume of the unit ball (cf. below), such that $\int \varphi_\varepsilon(x) d^n x = 1$.

To begin with we consider the inner integral

$$h_\varepsilon(y) := \int_R \varphi_\varepsilon(\psi(z) - y) d^n z.$$

For $\varepsilon < \varepsilon_0$ the integrand is nonzero only for $z \in K = \psi^{-1}(\overline{B_{\varepsilon_0}(y)})$, where K is some compact set containing $x = \psi^{-1}(y)$. Using the affine change of coordinates $z = x + \varepsilon w$ we obtain

$$h_\varepsilon(y) = \int_{W_\varepsilon(x)} \varphi\left(\frac{\psi(x + \varepsilon w) - \psi(x)}{\varepsilon}\right) d^n w, \qquad W_\varepsilon(x) = \frac{1}{\varepsilon}(K - x).$$

By

$$\left|\frac{\psi(x + \varepsilon w) - \psi(x)}{\varepsilon}\right| \geq \frac{1}{C}|w|,$$

the integrand is nonzero only for $w \in B_C(0)$. Hence, as $\varepsilon \to 0$ the domain $W_\varepsilon(x)$ will eventually cover all of $B_C(0)$ and dominated convergence implies

$$\lim_{\varepsilon\downarrow 0} h_\varepsilon(y) = \int_{B_C(0)} \varphi(\frac{\partial\psi}{\partial x}x)w)d^n w = |J_\psi(x)|^{-1}.$$

for all x where ψ is differentiable. Consequently, $\lim_{\varepsilon\downarrow 0} I_\varepsilon = |\psi(R)|$ again by dominated convergence. Now we use Fubini to interchange the order of integration

$$I_\varepsilon = \int_R \int_{\psi(R)} |J_\psi(\psi^{-1}(y))|\varphi_\varepsilon(\psi(z) - y)d^n y\, d^n z.$$

The we have

$$\lim_{\varepsilon\downarrow 0} \int_{\psi(R)} |J_\psi(\psi^{-1}(y))|\varphi_\varepsilon(\psi(z) - y)d^n y = |J_\psi(\psi^{-1}(\psi(z)))| = |J_\psi(z)|$$

at every Lebesgue point of $J_\psi(\psi^{-1}(y))$ (Problem 3.31) and hence dominated convergence shows $\lim_{\varepsilon\downarrow 0} I_\varepsilon = \int_R |J_\psi(z)|d^n z$. □

With this result at our disposal we can now show that item (v) from Lemma 7.11 holds for bi-Lipschitz maps.

Lemma 7.41. *Let $\psi : U \to V$ be a bi-Lipschitz. Then $f \in W^{1,p}(V)$ if and only if $f \circ \psi \in W^{1,p}(U)$ and we have the **change of variables formula** $\partial_j(f \circ \psi) = \sum_k (\partial_k f)(\psi)\partial_j \psi_k$. Moreover, $\|f \circ \psi\|_{W^{1,p}} \le C\|f\|_{W^{1,p}}$.*

Proof. If ψ is Lipschitz, then, by Lemma 7.32, the derivatives $\partial_j\psi$ as well as $\partial_j\psi^{-1}$ exist a.e. and are bounded. In particular, the Jacobi determinant exists a.e. and satisfies $|J_\psi| \ge C$. Thus we can conclude that composition with ψ is a homeomorphism between $L^p(U)$ and $L^p(V)$ for $1 \le p \le \infty$ as in the proof of Lemma 7.11.

Furthermore, as in the proof of Lemma 7.11 consider f_ε and note that $f_\varepsilon \circ \psi$ is Lipschitz and hence is in $W^{1,\infty}(U)$ by Lemma 7.15. In particular, it has a weak derivative which can be computed a.e. using the chain rule. Hence the rest follows as in the proof of Lemma 7.32. □

Of course once we have this result, it suffices to observe that Lemma 7.18 holds if U has a bounded Lipschitz boundary and the maps ψ_j are bi-Lipschitz instead of C^1 diffeomorphism, to conclude that Lemma 7.19 extends to such domains. That is, domains with a bounded Lipschitz boundary have the extension property. Similarly, Theorem 7.22 extends to this situation and hence such domains have a well-defined trace operator.

Finally let me remark, that also the Gauss–Green theorem extends to Lipschitz domains.

Theorem 7.42 (Gauss–Green). *If U is a bounded Lipschitz domain in $\mathbb{R}^n$ and $u \in W^{p,1}(U, \mathbb{R}^n)$ is a vector field, then*

$$\int_U (\operatorname{div} u) d^n x = \int_{\partial U} u \cdot \nu \, dS, \tag{7.38}$$

where the derivatives are understood as weak derivatives and the boundary values are understood in the sense of traces.

Proof. It suffices to first proof the case where $u \in C^1(U, \mathbb{R}^n)$ since the extension to $W^{1,p}$ then follows as in the proof of Lemma 7.24. But in this case one can again follow the proof from Theorem 2.19 upon observing that the Leibniz integral rule from Problem 2.35 holds pointwise at every point where g is differentiable and the fact that the fundamental theorem of calculus holds for Lipschitz functions (Theorem 4.29 below together with the fact that every Lipschitz function is absolutely continuous by Example 4.13). □

Of course this implies that we also have integration by parts as well as the other consequences of this theorem.

7.6. Applications to elliptic equations

The purpose of this section is to show that Sobolev spaces provide a convenient framework for treating partial differential equations by functional analytic methods. To focus on the main ideas we will start by looking at the **Laplace equation**

$$\begin{aligned} -\Delta u(x) &= f(x), & x &\in U, \\ u(x) &= 0, & x &\in \partial U, \end{aligned} \tag{7.39}$$

on a bounded domain $U \subseteq \mathbb{R}^n$ with Dirichlet boundary conditions. Here $\Delta u = \sum_{j=1}^n \partial_j^2 u$ as usual. If we regard the derivatives as weak derivatives, then our equation reads

$$\int_U (\Delta\varphi)(x) u(x) d^n x = \int_U \varphi(x) f(x) d^n x, \qquad \varphi \in C_c^\infty(U), \tag{7.40}$$

or, after an integration by parts, we can also write it in the more symmetric form

$$\sum_{j=1}^n \int_U (\partial_j \varphi)(\partial_j u) d^n x = \int_U \varphi(x) f(x) d^n x, \qquad \varphi \in C_c^\infty(U). \tag{7.41}$$

Now recall that by the Poincaré inequality (Theorem 7.38) we have a scalar product

$$\langle u, v \rangle := \sum_{j=1}^n \int_U (\partial_j u)(\partial_j v) d^m x \tag{7.42}$$

on $H_0^1(U,\mathbb{R})$ whose associated norm is equivalent to the usual one. Moreover, using the fact that $C_c^\infty(U,\mathbb{R}) \subset H_0^1(U,\mathbb{R})$ is dense we see that we can write our last form as

$$\langle v, u\rangle = \langle v, f\rangle_2, \qquad v \in H_0^1(U,\mathbb{R}), \tag{7.43}$$

where $\langle u, v\rangle_2 := \int_U u(x)v(x)d^nx$ denotes the scalar product in $L^2(U,\mathbb{R})$.

We will call a solution $u \in H_0^1(U,\mathbb{R})$ of (7.43) a **weak solution** of the Dirichlet problem (7.39). If a solution is, in addition, in $H^2(U,\mathbb{R})$, it is called a **strong solution**. In this case we can undo our integration by parts and conclude that u solves (7.39), where the derivatives are understood as weak derivatives and the boundary condition is understood in the sense of traces (at least for U with sufficiently smooth boundary; see Lemma 7.25).

Finally note that (7.43) can be understood as

$$\langle v, u\rangle = \langle Jv, f\rangle_2, \qquad v \in H_0^1(U,\mathbb{R}), \tag{7.44}$$

where $J : H_0^1(U,\mathbb{R}) \hookrightarrow L^2(U,\mathbb{R})$ is the natural embedding. Since this embedding is bounded (in fact, even compact; we will come back to this later), we can use the adjoint operator to write this as

$$\langle v, u\rangle = \langle v, J^*f\rangle, \qquad v \in H_0^1(U,\mathbb{R}), \tag{7.45}$$

which shows that the weak problem (7.43) has a unique solution $u = J^*f \in H_0^1(U,\mathbb{R})$ for every $f \in L^2(U,\mathbb{R})$. Also note the estimate $\|u\| = \|J^*f\| \le C\|f\|_2$, where C is the optimal constant from the Poincaré inequality (since $\|J^*\| = \|J\| = C$).

Now what about strong solutions? To this end we regard (7.39) as an operator equation

$$Lu = f, \tag{7.46}$$

where

$$Lu := -\Delta u, \qquad u \in \mathfrak{D}(L) := H_0^1(U,\mathbb{R}) \cap H^2(U,\mathbb{R}). \tag{7.47}$$

Then uniqueness of weak solutions implies that $\bar{L} := (JJ^*)^{-1}$ is a well-defined operator $\bar{L} : \mathfrak{D}(\bar{L}) \subset L^2(U,\mathbb{R}) \to L^2(U,\mathbb{R})$ which coincides with L when restricted to $\mathfrak{D}(L)$ (in particular, $\mathfrak{D}(L) \subseteq \mathfrak{D}(\bar{L}) = \operatorname{Ran}(JJ^*) \subset \operatorname{Ran}(J) = H_0^1(U,\mathbb{R})$). Since JJ^* is self-adjoint, $\bar{L}$ is also self-adjoint, known as the **Friedrichs extension** of the Dirichlet Laplacian (see [**21**, Sect. 2.3]). Choosing $v \in C_c^\infty(U,\mathbb{R})$ in (7.43) shows that

$$\mathfrak{D}(\bar{L}) = \{u \in H_0^1(U,\mathbb{R}) | \Delta u \in L^2(U,\mathbb{R})\}, \tag{7.48}$$

where Δu is understood as a weak derivative (this does not mean that the second derivatives exist individually, it is only this particular combination of second derivatives which is required to exist). When we also have $\mathfrak{D}(\bar{L}) \subseteq \mathfrak{D}(L)$, that is, when every weak solution is also a strong solution, is a tricky question which we will not answer here.

Instead, we point out that since the embedding J is not only continuous, but even compact by the Rellich–Kondrachov theorem (Theorem 7.35), we can apply the spectral theorem for compact operators (since JJ^* is self-adjoint Theorem 3.7 from [**22**] will do):

Theorem 7.43. *The operator $\bar{L}$ has a sequence of discrete real eigenvalues $0 < \lambda_0 < \lambda_1 < \cdots$ converging to ∞. The corresponding normalized eigenfunctions form an orthonormal basis for $L^2(U, \mathbb{R}^n)$.*

Observe that the inverse of the lowest eigenvalue λ_0^{-1} is the optimal constant for the Poincarè inequality.

Finally we remark that our consideration extend easily to the Dirichlet problem for second order **elliptic equations** of the form

$$Lu(x) := -\sum_{i,j=1}^{n} \partial_i A_{ij}(x) \partial_j u(x) + c(x) u(x) \tag{7.49}$$

with $A_{i,j}, c \in L^\infty(U, \mathbb{R})$ with

$$a_0 = \inf_{e \in S^n, x \in U} e_i A_{ij}(x) e_j > 0, \quad c_0 = \inf_{x \in U} c(x) \geq 0. \tag{7.50}$$

As domain we choose $\mathfrak{D}(L) = \{u \in H_0^1(U, \mathbb{R}) | A_{ij} \partial_j u \in H^1(U, \mathbb{R}), 1 \leq i, j \leq n\}$. Then the ellipticity condition $a_0 > 0$ ensures that the symmetric bilinear form

$$a(u, v) := \sum_{i,j=1}^{n} \int_U \left(A_{ij}(x) (\partial_j u(x)) (\partial_i v(x)) + c(x) u(x) v(x) \right) d^n x \tag{7.51}$$

gives rise to a norm which is equivalent to the usual norm on $H_0^1(U, \mathbb{R})$ and hence we can proceed as before.

Problem 7.41. *Compute J^* for $U := (0, 1) \subset \mathbb{R}$.*

Problem 7.42. *Consider the elliptic Dirichlet problem associated with*

$$Lu(x) := -\sum_{i,j=1}^{n} \partial_i A_{ij}(x) \partial_j u(x) + \sum_{j=1}^{n} b_j(x) \partial_j u(x) + c(x) u(x),$$

where $A_{i,j}, b_j, c \in L^\infty(U, \mathbb{R})$ with

$$a_0 = \inf_{|e|=1, x \in U} \sum_{i,j} e_i A_{ij}(x) e_j, \quad b_0 = \sup_{x \in U} |b(x)|, \quad c_0 = \inf_{x \in U} c(x).$$

*Show that $Lu = f$ has a unique weak solution in $H_0^1(U, \mathbb{R})$ provided $4a_0 c_0 > b_0^2$. (Hint: Choose $H_0^1(U, \mathbb{R})$ equipped with (7.42) as underlying Hilbert space. On this Hilbert space there is a corresponding bilinear form $a(u, v)$, however, if $b \neq 0$ this form is not symmetric. To overcome this problem use the Lax–Milgram theorem (Theorem 2.17 from [**22**]). Again there will be a problem* *with the term corresponding to b when establishing coercivity. However, note that this term can be controlled using the other two.)*

Chapter 8

The Fourier transform

8.1. The Fourier transform on L^1 and L^2

For $f \in L^1(\mathbb{R}^n)$ we define its **Fourier transform**[1] via

$$\mathcal{F}(f)(p) \equiv \hat{f}(p) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}px} f(x) d^n x. \tag{8.1}$$

Here $px = p_1 x_1 + \cdots + p_n x_n$ is the usual scalar product in $\mathbb{R}^n$ and we will use $|x| = \sqrt{x_1^2 + \cdots + x_n^2}$ for the Euclidean norm.

Lemma 8.1. *The Fourier transform is a bounded map from $L^1(\mathbb{R}^n)$ into $C_b(\mathbb{R}^n)$ satisfying*

$$\|\hat{f}\|_\infty \le (2\pi)^{-n/2} \|f\|_1. \tag{8.2}$$

Proof. Since $|\mathrm{e}^{-\mathrm{i}px}| = 1$ the estimate (8.2) 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. □

Note that if f is nonnegative we have equality: $\|\hat{f}\|_\infty = (2\pi)^{-n/2} \|f\|_1 = \hat{f}(0)$.

The following simple properties are left as an exercise.

[1] Joseph Fourier (1768–1830), French mathematician and physicist

Lemma 8.2. *Let $f \in L^1(\mathbb{R}^n)$. Then*

$$(f(x+a))^\wedge(p) = \mathrm{e}^{\mathrm{i}ap}\hat{f}(p), \qquad a \in \mathbb{R}^n, \tag{8.3}$$

$$(f(Mx))^\wedge(p) = |\det(M)|^{-1}\hat{f}((M^{-1})^T p), \qquad M \in \mathrm{GL}(\mathbb{R}^n), \tag{8.4}$$

$$(\mathrm{e}^{\mathrm{i}xa}f(x))^\wedge(p) = \hat{f}(p-a), \qquad a \in \mathbb{R}^n, \tag{8.5}$$

$$(f(\lambda x))^\wedge(p) = \frac{1}{\lambda^n}\hat{f}(\frac{p}{\lambda}), \qquad \lambda > 0, \tag{8.6}$$

$$(f(-x))^\wedge(p) = \hat{f}(-p). \tag{8.7}$$

Next we look at the connection with differentiation.

Lemma 8.3. *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*

$$(\partial_j f)^\wedge(p) = \mathrm{i}p_j\hat{f}(p). \tag{8.8}$$

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

$$(x_j f(x))^\wedge(p) = \mathrm{i}\partial_j\hat{f}(p). \tag{8.9}$$

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 2.13. In particular, $\hat{f}(p)$ is differentiable. □

This result immediately extends to higher derivatives. Roughly speaking this lemma shows that the decay of a function is related to the smoothness of its Fourier transform and the smoothness of a function is related to the decay of its Fourier transform.

Next, 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

$$\partial_\alpha f := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}, \quad x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \quad |\alpha| := \alpha_1 + \cdots + \alpha_n. \tag{8.10}$$

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**

$$\mathcal{S}(\mathbb{R}^n) := \{f \in C^\infty(\mathbb{R}^n) | \sup_x |x^\alpha (\partial_\beta f)(x)| < \infty, \ \forall \alpha, \beta \in \mathbb{N}_0^n\} \tag{8.11}$$

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. Also, by Leibniz' rule, the product of two Schwartz functions is again a Schwartz function.

Lemma 8.4. *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*

$$(\partial_\alpha f)^\wedge(p) = (\mathrm{i}p)^\alpha \hat{f}(p), \qquad (x^\alpha f(x))^\wedge(p) = \mathrm{i}^{|\alpha|} \partial_\alpha \hat{f}(p). \tag{8.12}$$

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 (8.2). 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**.

In particular, this allows us to conclude that the Fourier transform of an integrable function will vanish at ∞. Recall that we denote the space of all continuous functions $f : \mathbb{R}^n \to \mathbb{C}$ which vanish at ∞ by $C_0(\mathbb{R}^n)$.

Corollary 8.5 (Riemann-Lebesgue)**.** *The Fourier transform maps $L^1(\mathbb{R}^n)$ into $C_0(\mathbb{R}^n)$.*

Proof. First of all recall that $C_0(\mathbb{R}^n)$ equipped with the sup norm is a Banach space and that $\mathcal{S}(\mathbb{R}^n)$ is dense (Problem 7.4). By the previous lemma we have $\hat{f} \in C_0(\mathbb{R}^n)$ 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 (8.2) shows that the Fourier transform extends to a continuous map from $L^1(\mathbb{R}^n)$ into $C_0(\mathbb{R}^n)$. □

Next we will turn to the inversion of the Fourier transform. As a preparation we will need the Fourier transform of a **Gaussian**.

Lemma 8.6. *We have* $\mathrm{e}^{-z|x|^2/2} \in \mathcal{S}(\mathbb{R}^n)$ *for* $\mathrm{Re}(z) > 0$ *and*

$$\mathcal{F}(\mathrm{e}^{-z|x|^2/2})(p) = \frac{1}{z^{n/2}}\mathrm{e}^{-|p|^2/(2z)}. \tag{8.13}$$

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 $n = 1$ (Problem 8.3).

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 2.23)

$$c = \hat{\phi}_z(0) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp(-zx^2/2)dx = \frac{1}{\sqrt{z}}$$

at least for $z > 0$. However, since the integral is holomorphic for $\mathrm{Re}(z) > 0$ by Problem 2.18, 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 show

Theorem 8.7. *The Fourier transform is a bounded injective map from* $L^1(\mathbb{R}^n)$ *into* $C_0(\mathbb{R}^n)$*. Its inverse is given by*

$$f(x) = \lim_{\varepsilon\downarrow 0} \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}px-\varepsilon|p|^2/2}\hat{f}(p)d^n p, \tag{8.14}$$

where the limit has to be understood in L^1*. Moreover,* (8.14) *holds at every Lebesgue point (cf. Theorem 4.6) and hence in particular at every point of continuity.*

Proof. Abbreviate $\phi_\varepsilon(x) := (2\pi)^{-n/2}\exp(-\varepsilon|x|^2/2)$. Then the right-hand side is given by

$$\int_{\mathbb{R}^n} \phi_\varepsilon(p)\mathrm{e}^{\mathrm{i}px}\hat{f}(p)d^n p = \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \phi_\varepsilon(p)\mathrm{e}^{\mathrm{i}px}f(y)\mathrm{e}^{-\mathrm{i}py}d^n y d^n p$$

and, invoking Fubini and Lemma 8.2, we further see that this is equal to

$$= \int_{\mathbb{R}^n} (\phi_\varepsilon(p)\mathrm{e}^{\mathrm{i}px})^\wedge(y)f(y)d^n y = \int_{\mathbb{R}^n} \frac{1}{\varepsilon^{n/2}}\phi_{1/\varepsilon}(y-x)f(y)d^n y.$$

But the last integral converges to f in $L^1(\mathbb{R}^n)$ by Lemma 3.21. Moreover, it is straightforward to see that it converges at every point of continuity. The case of Lebesgue points follows from Problem 9.15. □

Of course when $\hat{f} \in L^1(\mathbb{R}^n)$, the limit is superfluous and we obtain

Corollary 8.8. *Suppose $f, \hat{f} \in L^1(\mathbb{R}^n)$. Then*

$$(\hat{f})^\vee = f, \tag{8.15}$$

where

$$\check{f}(p) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}px} f(x) d^n x = \hat{f}(-p). \tag{8.16}$$

In particular, $\mathcal{F} : F^1(\mathbb{R}^n) \to F^1(\mathbb{R}^n)$ is a bijection, where

$$F^1(\mathbb{R}^n) := \{f \in L^1(\mathbb{R}^n) | \hat{f} \in L^1(\mathbb{R}^n)\}. \tag{8.17}$$

Moreover, $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ is a bijection.

Observe that we have $F^1(\mathbb{R}^n) \subset L^1(\mathbb{R}^n) \cap C_0(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$ for any $p \in [1, \infty]$ (cf. also Problem 8.2) and choosing f continuous (8.15) will hold pointwise.

However, note that $\mathcal{F} : L^1(\mathbb{R}^n) \to C_0(\mathbb{R}^n)$ is not onto (cf. Problem 8.7). Nevertheless the inverse Fourier transform $\mathcal{F}^{-1}$ is a closed map from $\mathrm{Ran}(\mathcal{F}) \to L^1(\mathbb{R}^n)$ by Lemma 8.1 from **[22]**.

Lemma 8.9. *Suppose $f \in F^1(\mathbb{R}^n)$. Then $f, \hat{f} \in L^2(\mathbb{R}^n)$ and*

$$\|f\|_2^2 = \|\hat{f}\|_2^2 \le (2\pi)^{-n/2} \|f\|_1 \|\hat{f}\|_1 \tag{8.18}$$

holds.

Proof. This follows from Fubini's theorem since

$$\begin{aligned}\int_{\mathbb{R}^n} |\hat{f}(p)|^2 d^n p &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x)^* \hat{f}(p) \mathrm{e}^{\mathrm{i}px} d^n p\, d^n x \\ &= \int_{\mathbb{R}^n} |f(x)|^2 d^n x\end{aligned}$$

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 1.16 from **[22]**, 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 8.10 (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 (8.1) whenever the right-hand side is integrable.

Lemma 8.11. *Let $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, then (8.1) continues to hold, where $\mathcal{F}$ now denotes the extension of the Fourier transform from $\mathcal{S}(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$. Similarly, for $f \in L^2(\mathbb{R}^n)$ with $\hat{f} \in L^1(\mathbb{R}^n)$ Theorem 8.7 continues to hold.*

Proof. If f has compact support, then by Lemma 3.20 its mollification $\phi_\varepsilon * f \in C_c^\infty(\mathbb{R}^n)$ converges to f both in L^1 and L^2. Hence the claim holds for every f 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.

For the second claim note that by Theorem 8.7 $(\mathcal{F}^{-1}f)(x) = (\mathcal{F}f)(-x)$ at least for $f \in \mathcal{S}(\mathbb{R}^n)$ which remains true when taking limits. □

In particular,

$$\hat{f}(p) = \lim_{m\to\infty} \frac{1}{(2\pi)^{n/2}} \int_{|x|\le m} \mathrm{e}^{-\mathrm{i}px} f(x) d^n x, \tag{8.19}$$

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 8.12. *The* ***convolution***

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y) g(x-y) d^n y = \int_{\mathbb{R}^n} f(x-y) g(y) d^n y \tag{8.20}$$

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***

$$\|f * g\|_1 \le \|f\|_1 \|g\|_1. \tag{8.21}$$

Moreover, its Fourier transform is given by

$$(f * g)^\wedge = (2\pi)^{n/2} \hat{f}\hat{g}. \tag{8.22}$$

Proof. The fact that $f * g$ is in L^1 together with Young's inequality follows by applying Fubini's theorem to $h(x,y) = f(x-y)g(y)$ (in fact we have shown a more general version in Lemma 3.20). For the last claim we compute

$$\begin{aligned}
(f * g)^\wedge(p) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}px} \left(\int_{\mathbb{R}^n} f(y) g(x-y) d^n y \right) d^n x \\
&= \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}py} f(y) \left(\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}p(x-y)} g(x-y) d^n x \right) d^n y \\
&= \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}py} f(y) \hat{g}(p) d^n y = (2\pi)^{n/2} \hat{f}(p)\hat{g}(p),
\end{aligned}$$

where we have again used Fubini's theorem. □

Example 8.1. The image of the integrable functions under the Fourier transform is known as the **Wiener algebra**

$$\mathcal{A}(\mathbb{R}^n) := \{\hat{f} | f \in L^1(\mathbb{R}^n)\}.$$

By construction, this is just the range of the Fourier transform $\mathcal{F}(L^1(\mathbb{R}^n)) \subseteq C_0(\mathbb{R}^n)$ and the Fourier transform is a bijection $\mathcal{F}(L^1(\mathbb{R}^n)) \to \mathcal{A}(\mathbb{R}^n)$. If we equip it with the norm $\|\hat{f}\|_{\mathcal{A}} := (2\pi)^{-n/2}\|f\|_1$ we get a Banach space isomorphic to $L^1(\mathbb{R}^n)$ and the norm satisfies $\|f\|_\infty \le \|f\|_{\mathcal{A}}$ by (8.2). Moreover, Lemma 8.12 shows that the product of two functions $f, g \in \mathcal{A}(\mathbb{R}^n)$ is again in the Wiener algebra with

$$\|fg\|_{\mathcal{A}} \le \|f\|_{\mathcal{A}}\|g\|_{\mathcal{A}}.$$

In other words, the Wiener algebra 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 8.13. *Let $g \in L^1(\mathbb{R}^n)$ and $f \in L^2(\mathbb{R}^n)$. Then $f * g \in L^2(\mathbb{R}^n)$ with*

$$\|f * g\|_2 \le \|f\|_2\|g\|_1. \tag{8.23}$$

and

$$(f * g)^\wedge = (2\pi)^{n/2}\hat{f}\hat{g}. \tag{8.24}$$

Proof. The fact that $f * g$ is in L^2 together with the inequality follows as before (or see Lemma 3.20). Hence taking a cutoff $f_m := f\chi_{B_m(0)}$ such that $f_r \in L^1 \cap L^2$ and $f_r \to f$ in L^2 establishes the claim upon taking the limit in $(f_m * g)^\wedge = (2\pi)^{n/2}\hat{f}_m\hat{g}$. □

Corollary 8.14. *The convolution of two $\mathcal{S}(\mathbb{R}^n)$ functions as well as their product is in $\mathcal{S}(\mathbb{R}^n)$ and*

$$(f * g)^\wedge = (2\pi)^{n/2}\hat{f}\hat{g}, \qquad (fg)^\wedge = (2\pi)^{-n/2}\hat{f} * \hat{g} \tag{8.25}$$

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 formula 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 8.15. *The convolution of two $L^2(\mathbb{R}^n)$ functions is in* $\mathrm{Ran}(\mathcal{F}) \subset C_0(\mathbb{R}^n)$ *and we have $\|f * g\|_\infty \le \|f\|_2\|g\|_2$ as well as*

$$(fg)^\wedge = (2\pi)^{-n/2}\hat{f} * \hat{g}, \qquad (f * g)^\wedge = (2\pi)^{n/2}\hat{f}\hat{g} \tag{8.26}$$

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 $f_m, g_m \in \mathcal{S}(\mathbb{R}^n)$ 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 $C_0(\mathbb{R}^n)$ and on $L^2(\mathbb{R}^n)$ we obtain

$$(fg)^\wedge = \lim_{m\to\infty} (f_m g_m)^\wedge = (2\pi)^{-n/2} \lim_{m\to\infty} \hat{f}_m * \hat{g}_m = (2\pi)^{-n/2} \hat{f} * \hat{g}.$$

Similarly,

$$(f * g)^\wedge = \lim_{m\to\infty} (f_m * g_m)^\wedge = (2\pi)^{n/2} \lim_{m\to\infty} \hat{f}_m \hat{g}_m = (2\pi)^{n/2} \hat{f}\hat{g}$$

from which that last claim follows since $\mathcal{F} : \mathrm{Ran}(\mathcal{F}) \to L^1(\mathbb{R}^n)$ is closed as it is the inverse of a bounded operator (Lemma 8.1 from [**22**]). □

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

$$\|(x_j - x^0) f(x)\|_2^2 = \int_{\mathbb{R}^n} (x_j - x^0)^2 |f(x)|^2 d^n x \tag{8.27}$$

which will be small if f is well concentrated around x^0 in the j'th coordinate direction.

Theorem 8.16 (Heisenberg uncertainty principle). *Suppose $f \in \mathcal{S}(\mathbb{R}^n)$. Then for any $x^0, p^0 \in \mathbb{R}$ we have*

$$\|(x_j - x^0) f(x)\|_2 \|(p_j - p^0) \hat{f}(p)\|_2 \ge \frac{\|f\|_2^2}{2}. \tag{8.28}$$

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 8.2. 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,

$$\|f\|_2^2 \le 2 \|x_j f(x)\|_2 \|\partial_j f(x)\|_2 = 2 \|x_j f(x)\|_2 \|p_j \hat{f}(p)\|_2$$

the claim follows. □

The name stems 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 (8.28) 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 8.17. *Suppose that $f \in L^1(\mathbb{R}^n)$ has compact support, then f is real analytic. In particular, f can only vanish on a discrete set unless it vanishes identically.*

Proof. By Theorem 8.7 we have

$$f(x) = \frac{1}{(2\pi)^{n/2}} \int_B \mathrm{e}^{\mathrm{i}xp} \hat{f}(p) d^n p, \qquad A := \operatorname{supp}(\hat{f}).$$

Moreover, by the multinomial theorem

$$\mathrm{e}^{\mathrm{i}xp} = \sum_{k=0}^{\infty} \mathrm{i}^k \frac{(xp)^k}{k!} = \sum_{\alpha} \mathrm{i}^{|\alpha|} \frac{x^\alpha p^\alpha}{\alpha!}$$

and hence

$$f(x) = \sum_{\alpha} \frac{c_\alpha}{\alpha!} x^\alpha, \qquad c_\alpha := \frac{\mathrm{i}^{|\alpha|}}{(2\pi)^{n/2}} \int_A p^\alpha \hat{f}(p) d^n p.$$

In particular, $\partial_\alpha f(0) = c_\alpha$. Now let y be an accumulation points of zeros of f. We can assume $y = 0$ after translating f. Then $\partial_\alpha f(0) = 0$ for all multi-indices α and the previous formula shows $f = 0$. □

Problem 8.1. *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 8.2. *Show that $F^1(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$ with*

$$\|f\|_p \le (2\pi)^{\frac{n}{2}(1-\frac{1}{p})} \|f\|_1^{\frac{1}{p}} \|\hat{f}\|_1^{1-\frac{1}{p}}.$$

Moreover, show that $\mathcal{S}(\mathbb{R}^n) \subset F^1(\mathbb{R}^n)$ and conclude that $F^1(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $p \in [1, \infty)$. (Hint: Use $x^p \le x$ for $0 \le x \le 1$ to show $\|f\|_p \le \|f\|_\infty^{1-1/p} \|f\|_1^{1/p}$.)

Problem* 8.3. *Suppose $f_j \in L^1(\mathbb{R})$, $j = 1, \dots, n$ and set $f(x) = \prod_{j=1}^n f_j(x_j)$. Show that $f \in L^1(\mathbb{R}^n)$ with $\|f\|_1 = \prod_{j=1}^n \|f_j\|_1$ and $\hat{f}(p) = \prod_{j=1}^n \hat{f}_j(p_j)$.*

Problem 8.4. *Compute the Fourier transform of the following functions $f : \mathbb{R} \to \mathbb{C}$:*

(i) $f(x) = \chi_{(-1,1)}(x)$. (ii) $f(x) = \frac{\mathrm{e}^{-k|x|}}{k}$, $\operatorname{Re}(k) > 0$.

Problem 8.5. *Show*

$$\int_0^\infty \frac{\sin(x)^2}{x^2} dx = \frac{\pi}{2}.$$

(Hint: Problem 8.4 (i).)

Problem 8.6. *Suppose $f \in L^1(\mathbb{R}^n)$. If f is continuous at 0 and $\hat{f}(p) \geq 0$ then*

$$f(0) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{f}(p) d^n p.$$

Use this to show the Plancherel identity for $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ by applying it to $F := f * \tilde{f}$, where $\tilde{f}(x) = f(-x)^*$.

Problem* 8.7. *Show that $\mathcal{F} : L^1(\mathbb{R}^n) \to C_0(\mathbb{R}^n)$ 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)*: Consider $\phi_z(x) = \exp(-zx^2/2)$ for $z = \lambda + \mathrm{i}\omega$ with $\lambda > 0$.)*

Problem 8.8 (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 8.2 and the fact that a subspace is total if and only if its orthogonal complement is zero.)*

Problem 8.9 (Rellich)**.** *Let C be a positive constant and let $f_j : \mathbb{R}^n \to \mathbb{R}$ be two measurable functions such that $f_j(x) \geq 1$ with $\lim_{|x|\to\infty} f_j(x) = \infty$ for $j = 0, 1$. Then the set $\{f \in L^2(\mathbb{R}^n) | \, \|f_0 f\|_2 + \|f_1 \hat{f}\|_2 \leq C\}$ is compact in $L^2(\mathbb{R}^n)$. (Hint: Example 3.3.)*

Problem 8.10. *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 $|\mathrm{Im}(p)| < k$.*

8.2. Some further topics

A function $f : \mathbb{R}^n \to \mathbb{C}$ is called **spherically symmetric** (or **radial**) if it is invariant under rotations; that is, $f(Ox) = f(x)$ for all $O \in \mathrm{SO}(\mathbb{R}^n)$ (equivalently, f depends only on the distance to the origin $|x|$). By Lemma 8.2 we have $(f \circ O)^\wedge = \hat{f} \circ O$ (recall $|\det(O)| = 1$) and hence the Fourier transform of a spherically symmetric function is again spherically symmetric. In fact, in this case the integral can be evaluated using spherical coordinates. The final result can be expressed in terms of the the **Hankel transform**

$$\mathcal{H}_\nu(f)(r) = \int_0^\infty f(s) J_\nu(sr) s \, ds, \qquad \nu \geq -\frac{1}{2}, \tag{8.29}$$

with

$$J_\nu(z) := \sum_{j=0}^\infty \frac{(-1)^j}{j! \Gamma(\nu + j + 1)} \left(\frac{z}{2}\right)^{2j+\nu} \tag{8.30}$$

the **Bessel function** of order $\nu \in \mathbb{C}$ ([**14**, (10.2.2)]; see Problem 8.11).

Example 8.2. Note that for half integer order the Bessel functions reduce to trigonometric functions. In fact, we have (Problem 8.13)

$$J_{-1/2}(z) = \sqrt{\frac{2}{\pi z}} \cos(z).$$

and using Problem 8.11 (i) one gets

$$J_{\nu+n}(z) = (-1)^n z^{\nu+n} \left(\frac{1}{z}\frac{d}{dz}\right)^n z^{-\nu} J_\nu(z).$$

In particular,

$$J_{1/2}(z) = \sqrt{\frac{2}{\pi z}} \sin(z), \quad J_{3/2}(z)\sqrt{\frac{2}{\pi z}} \left(-\cos(z) + \frac{1}{z}\sin(z)\right),$$

etc. ◇

It will be convenient to use the following modified version

$$\tilde{\mathcal{H}}_\nu(f)(r) = \mathcal{H}_\nu(f(s)(s/r)^\nu)(r) = \int_0^\infty f(s)\tilde{J}_\nu(sr)s^{2\nu+1}ds, \tag{8.31}$$

where $\tilde{J}_\nu(z) := z^{-\nu}J_\nu(z)$. Note that the integral representation from Problem 8.12 implies that $\tilde{J}_\nu(r)$ is bounded for $r > 0$, $\nu \geq -\frac{1}{2}$, explicitly, $|\tilde{J}_\nu(r)| \leq (\sqrt{\pi}\Gamma(\nu+\frac{1}{2})2^\nu)^{-1}$. In particular, the modified Hankel transform is a bounded linear transform $L^1((0,\infty), r^{2\nu+1}dr) \to C_b([0,\infty))$. Moreover, using the connection with the Fourier transform established below, one sees that it is unitary on $L^2((0,\infty), r^{2\nu+1}dr)$ at least for $\nu = \frac{n}{2} - 1$, $n \in \mathbb{N}$. In the latter case the integral has to be understood as a limit $\lim_{m\to\infty}\int_0^m$ like in the case of the Fourier transform.

Theorem 8.18. *Let $f(x) = F(|x|)$ be a radial function with $f \in L^1(\mathbb{R}^n)$ or $f \in L^2(\mathbb{R}^n)$. Then the Fourier transform can be expressed in terms of the Hankel transform:*

$$\hat{f}(p) = F_n(|p|), \qquad F_n(r) := \tilde{\mathcal{H}}_{\frac{n}{2}-1}(F)(r). \tag{8.32}$$

Proof. The case $n = 1$ is part of Problem 8.13 and hence we can assume $n \geq 2$ without loss of generality. Since $\hat{f}(p)$ is radial, we can choose $p = |p|\delta_n$ in the direction of the last coordinate axis for the purpose of evaluating the integral. Then, using spherical coordinates we obtain

$$\begin{aligned}\hat{f}(p) &= \frac{S_{n-1}}{(2\pi)^{n/2}} \int_0^\infty \left(\int_0^\pi \mathrm{e}^{-\mathrm{i}|p|r\cos(\theta)} \sin(\theta)^{n-2} d\theta\right) F(r) r^{n-1} dr \\ &= \int_0^\infty J_{n/2-1}(r|p|)F(r)(r/|p|)^{n/2-1} r dr = \tilde{\mathcal{H}}_{\frac{n}{2}-1}(F)(|p|)\end{aligned}$$

by Problem 8.12. □

Corollary 8.19. *The Hankel transform $\mathcal{H}_\nu$ is unitary on $L^2((0,\infty), r\,dr)$ for $\nu = \frac{n}{2}-1$, $n \in \mathbb{N}$.*

Another useful fact is that the radial Fourier transform for different odd or even dimensions can be computed recursively:

Lemma 8.20. *Let $F \in L^1((0,\infty), r^{\nu+1}dr) \cap L^1((0,\infty), r^{\nu+2}dr)$. Then*

$$\tilde{\mathcal{H}}_{\nu+1}(F)(r) = -\frac{1}{r}\frac{d}{dr}\tilde{\mathcal{H}}_\nu(F)(r). \tag{8.33}$$

Similarly, if $F \in L^2((0,\infty), r^{\nu+1}dr) \cap L^2((0,\infty), r^{\nu+2}dr)$ and $\nu = \frac{n}{2}-1$, $n \in \mathbb{N}$.

Proof. Note that $\tilde{J}_\nu(z) = z^{-\nu}J_\nu(z)$ is an entire function which satisfies $\tilde{J}_\nu'(z) = -z\tilde{J}_{\nu+1}(z)$ by item (i) from Problem 8.11. Hence

$$-\frac{1}{r}\frac{d}{dr}\tilde{\mathcal{H}}_\nu(F)(r) = \frac{1}{r}\int_0^\infty sr\tilde{J}_\nu'(sr)sF(s)s^{2\nu+1}ds = \tilde{\mathcal{H}}_{\nu+1}(F)(r),$$

where interchanging differentiation and integration is permissible since $\tilde{J}_\nu$ is bounded on the real line. The L^2 case follows by approximation. □

Example 8.3. Let us compute the Fourier transform of the characteristic function of the unit ball in $\mathbb{R}^3$. In one dimension we have $\hat{\chi}_{(-1,1)}(p) = \sqrt{\frac{\pi}{2}}\frac{\sin(p)}{p}$ and hence

$$\hat{\chi}_{B_1(0)}(p) = -\frac{1}{r}\frac{d}{dr}\hat{\chi}_{(-1,1)}(r)\Big|_{r=|p|} = \sqrt{\frac{\pi}{2}}\left(\frac{1}{|p|^3}\sin(|p|) - \frac{1}{|p|^2}\cos(|p|)\right).$$

⋄

Example 8.4. From Problem 8.4 we know that the Fourier transform of $f(x) = \mathrm{e}^{-t|x|} \in L^1(\mathbb{R})$ for $\mathrm{Re}(t) > 0$ is given by

$$\hat{f}(p) = \sqrt{\frac{2}{\pi}}\frac{t}{t^2+x^2}$$

Hence, applying Lemma 8.20 recursively we obtain

$$\hat{f}(p) = \frac{\Gamma(\frac{n+1}{2})2^{n/2}}{\sqrt{\pi}}\frac{t}{(t^2+x^2)^{(n+1)/2}}$$

in $\mathbb{R}^n$ for odd n. In fact, this formula holds for all $n \in \mathbb{N}$ (Problem 8.14). ⋄

In probability theory, the Fourier transform of a measure is known as the **characteristic function** and plays an important role. Let us be a bit more general and define the Fourier transform of a complex measure μ by

$$\hat{\mu}(p) := \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n}\mathrm{e}^{-\mathrm{i}px}d\mu(x). \tag{8.34}$$

Note that in probability theory the characteristic function is defined as $\varphi_\mu(p) := (2\pi)^{-n/2}\hat{\mu}(-p)$.

Lemma 8.21. *The Fourier transform is a bounded injective map from $\mathcal{M}(\mathbb{R}^n)$ into $C_b(\mathbb{R}^n)$ satisfying*

$$\|\hat{\mu}\|_\infty \le (2\pi)^{-n/2}|\mu|(\mathbb{R}^n). \tag{8.35}$$

Proof. The estimate is immediate from (4.25) and dominated convergence implies that $\hat{\mu}(p)$ is continuous. Moreover, for $\varphi \in \mathcal{S}(\mathbb{R}^n)$ Fubini implies

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

and hence injectivity follows from Lemma 3.23. □

Example 8.5. The Fourier transform of the Dirac measure δ_{x_0} is

$$\hat{\delta}_{x_0}(p) = \frac{1}{(2\pi)^{n/2}} \mathrm{e}^{-\mathrm{i}px_0}. \qquad \diamond$$

Theorem 8.22. *Consider the Fourier transform of a complex measure on $\mathbb{R}$. Then*

$$\lim_{r\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-r}^{r} \frac{\mathrm{e}^{\mathrm{i}bp} - \mathrm{e}^{\mathrm{i}ap}}{\mathrm{i}p} \hat{\mu}(p) dp = \frac{\mu([a,b]) + \mu((a,b))}{2}. \tag{8.36}$$

In particular,

$$\mu((a,b]) = \lim_{\varepsilon\downarrow 0} \lim_{r\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-r}^{r} \frac{\mathrm{e}^{\mathrm{i}(b+\varepsilon)p} - \mathrm{e}^{\mathrm{i}(a+\varepsilon)p}}{\mathrm{i}p} \hat{\mu}(p) dp. \tag{8.37}$$

Proof. First of all note that the integrand is bounded since

$$\left| \frac{\mathrm{e}^{\mathrm{i}bp} - \mathrm{e}^{\mathrm{i}ap}}{\mathrm{i}p} \right| = \left| \int_a^b \mathrm{e}^{\mathrm{i}t} dt \right| \le b - a.$$

Hence we can use Fubini to write

$$\frac{1}{2\pi} \int_{\mathbb{R}} \int_{-r}^{r} \frac{\mathrm{e}^{\mathrm{i}bp} - \mathrm{e}^{\mathrm{i}ap}}{\mathrm{i}p} \mathrm{e}^{-\mathrm{i}px} dp\, d\mu(x).$$

Now the inner integral gives

$$\begin{aligned} \frac{1}{2\pi} \int_{-r}^{r} \frac{\mathrm{e}^{\mathrm{i}bp} - \mathrm{e}^{\mathrm{i}ap}}{\mathrm{i}p} \mathrm{e}^{-\mathrm{i}px} dp &= \frac{1}{2\pi} \int_{-r}^{r} \frac{\sin((b-x)p) - \sin((a-x)p)}{p} dp \\ &= \frac{1}{\pi} \big(\mathrm{Si}((b-x)r) - \mathrm{Si}((a-x)r) \big), \end{aligned}$$

where $\mathrm{Si}(z) = \int_0^z \frac{\sin(x)}{x} dx$ is the sine integral. In the limit $r \to \infty$ we obtain (using the Dirichlet integral — Problem 8.31)

$$\begin{aligned} \frac{1}{\pi} \big(\mathrm{Si}((b-x)r) - \mathrm{Si}((a-x)r) \big) &\to \frac{1}{2} \big(\mathrm{sign}((b-x)r) - \mathrm{sign}((a-x)r) \big) \\ &= \frac{\chi_{(a,b)}(x) + \chi_{[a,b]}(x)}{2} \end{aligned}$$

from which the claim follows after invoking dominated convergence. □

Corollary 8.23. *For the Fourier transform of a complex measure on $\mathbb{R}^n$ we have*

$$\mu((a,b]) = \lim_{\varepsilon\downarrow 0} \lim_{r\to\infty} \frac{1}{\sqrt{(2\pi)^{n/2}}} \int_{[-r,r]^n} \left(\prod_{j-1}^{n} \frac{\mathrm{e}^{\mathrm{i}(b_j+\varepsilon)p_j} - \mathrm{e}^{\mathrm{i}(a_j+\varepsilon)p_j}}{\mathrm{i}p_j} \right) \hat{\mu}(p) dp. \tag{8.38}$$

Finally we look at the connection with weak convergence.

Lemma 8.24. *Let μ_m, μ be finite positive measures on $\mathbb{R}^n$ and suppose the Fourier transforms converge pointwise*

$$\hat{\mu}_m(p) \to \hat{\mu}(p), \qquad p \in \mathbb{R}^n.$$

Then $\mu_n \to \mu$ weakly (cf. (5.20)*).*

Proof. First of all note that convergence for $p=0$ is equivalent to $\mu_m(\mathbb{R}^n) \to \mu(\mathbb{R}^n)$. In particular, $\mu_m(\mathbb{R}^n)$ is bounded and it suffices to establish vague convergence by Lemma 5.14.

Furthermore, using the definition of $\hat{f}$ and Fubini shows

$$\int \hat{f} d\mu = \int f(p)\hat{\mu}(p)dp, \qquad f \in L^1(\mathbb{R}^n).$$

Hence choosing $\hat{f} \in \mathcal{S}(\mathbb{R}^n)$ we see that (5.24) holds for $f \in \mathcal{S}(\mathbb{R})$ and vague convergence follows from dominated convergence. □

Next we turn to the case, where the limit is not known to be the Fourier transform of a measure.

Theorem 8.25 (Lévy continuity theorem)**.** *Let μ_m be a sequence of finite positive measures on $\mathbb{R}^n$ and suppose the Fourier transforms converge pointwise*

$$\hat{\mu}_m(p) \to \phi(p), \qquad p \in \mathbb{R}^n.$$

Then ϕ is the Fourier transform of a positive measure μ if and only if ϕ is continuous at 0*. Moreover, in this case $\mu_m \to \mu$ weakly (cf.* (5.20)*).*

Proof. Convergence at $p=0$ implies $\mu_m(\mathbb{R}^n) \to \phi(0)$ and hence $\mu_m(\mathbb{R}^n)$ is bounded. Thus by Theorem 6.11 we can pass to a subsequence and assume that μ_m converges to some measure μ vaguely.

To show that $\mu_m \to \mu$ weakly we will show that the sequence μ_m is tight (cf. Problem 5.15). To make the argument more transparent we look at one dimension first. We start with (using Fubini)

$$\begin{aligned}\frac{1}{2r}\int_{-r}^{r}\hat{\mu}_m(p)dp &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\frac{1}{2r}\int_{-r}^{r}\mathrm{e}^{-\mathrm{i}px}dp\,d\mu_m(x)\\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\frac{\sin(rx)}{rx}d\mu_m(x)\\ &\le \frac{1}{\sqrt{2\pi}}\left(\mu_m([-t,t]) + \frac{1}{rt}\mu_m(\mathbb{R}\setminus[-t,t])\right)\\ &= \frac{1}{\sqrt{2\pi}}\left(\mu_m(\mathbb{R}) - (1-\frac{1}{rt})\mu_m(\mathbb{R}\setminus[-t,t])\right).\end{aligned}$$

Choosing $t = \frac{2}{r}$ we thus obtain

$$\mu_m(\{x\in\mathbb{R}|\,|x| > \frac{r}{2}\}) \le \frac{\sqrt{2\pi}}{r}\int_{-r}^{r}(\hat{\mu}_m(0) - \hat{\mu}_m(p))dp.$$

Hence in the n dimensional case this implies

$$\mu_m(\{x\in\mathbb{R}^n|\,|x_j| > \frac{2}{r}\}) \le \frac{\sqrt{2\pi}}{r}\int_{-r}^{r}(\hat{\mu}_m(0) - \hat{\mu}_m(p_j\delta_j))dp_j,$$

where δ_j is the j'th basis vector in $\mathbb{R}^n$. Moreover, for $K_r := [-\frac{r}{2},\frac{r}{2}]^n$ this implies

$$\mu_m(\mathbb{R}^n\setminus K_r) \le \frac{\sqrt{2\pi}}{r}\sum_{j=1}^{n}\int_{-r}^{r}(\hat{\mu}_m(0) - \hat{\mu}_m(p_j\delta_j))dp_j.$$

Now choose r such that $|\phi(0) - \phi(p_j\delta_j)| \le \frac{\varepsilon}{4\sqrt{2\pi}n}$ for $|p| \le r$. Then

$$\lim_{m\to\infty}\mu_m(\mathbb{R}^n\setminus K_r) \le \frac{\sqrt{2\pi}}{r}\sum_{j=1}^{n}\int_{-r}^{r}(\phi(0) - \phi(p_j\delta_j))dp_j \le \frac{\varepsilon}{2},$$

which shows that the left-hand side is smaller than ε for m sufficiently large. □

Example 8.6. Choosing $d\mu_m(x) = \mathrm{e}^{-x^2/(2m)}\frac{dx}{\sqrt{m}}$ we have $\hat{\mu}_m(p) = \mathrm{e}^{-mp^2/2} \to \chi_{\{0\}}(p)$, which shows that continuity at 0 is important. Note that in this case $\mu_m \to 0$ vaguely. ◇

Another natural question is when a function ϕ is the Fourier transform of a positive measure. To answer this question one calls a function $\phi:\mathbb{R}^n\to\mathbb{C}$ **positive definite** if

$$\sum_{j,k=1}^{m}\phi(p_j - p_k)c_j c_k^* \ge 0 \tag{8.39}$$

for any finite collection of points $p_j\in\mathbb{R}^n$ and constants $c_j\in\mathbb{C}$. In other words, the matrix $\{\phi(p_j - p_k)\}_{1\le j,k\le m}$ is positive semidefinite. This definition should be compared with the definition (3.38) of a positive semidefinite kernel, which shows that ϕ is positive definite is the same as saying that $\phi(p-q)$ is a positive semidefinite kernel.

It is straightforward to check that the Fourier transform of a finite positive measure is positive definite:

$$\sum_{j,k} \hat{\mu}(p_j - p_k) c_j c_k^* = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \Big| \sum_j c_j \mathrm{e}^{-\mathrm{i} p_j x} \Big|^2 d\mu(x) \geq 0. \tag{8.40}$$

In fact, if we suppose $\hat{\mu} \in L^1(\mathbb{R}^n)$, then the integral operator

$$(Kf)(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{\mu}(y-x) f(y) dy$$

is positive, since $\mathcal{F} K \mathcal{F}^{-1}$ is multiplication by $(\hat{\mu})^\vee$ (and by injectivity of $\mathcal{F}$ we also see that in this case μ is absolutely continuous with density $(\hat{\mu})^\vee$). This connection with positive definite kernels is the key to establishing the converse:

Theorem 8.26 (Bochner[2]). *Let ϕ be a positive definite function which is continuous at 0. Then ϕ is the Fourier transform of a finite positive measure.*

Proof. Let us first consider the case when ϕ is integrable. Since ϕ is continuous (Problem 8.17), the kernel $K(p,q) = \phi(p-q)$ is positive definite and hence the associated integral operator $K: L^2 \to L^2$ is also positive, in the sense that $\langle f, Kf \rangle \geq 0$, by Lemma 3.28 (you can also show this directly by extracting the relevant part from the proof). In particular, by Corollary 8.13 it is given by multiplication with $\hat{\phi}$ once we apply the Fourier transform:

$$\langle f, Kf \rangle = (2\pi)^{n/2} \langle \hat{f}, \hat{\phi} \hat{f} \rangle \geq 0.$$

Now let f_ε be a Gaussians such that $\hat{f}_\varepsilon^2$ is an approximating identity to conclude $\hat{\phi} \geq 0$. Moreover, by (8.14) and monotone convergence we have

$$(2\pi)^{n/2} \phi(0) = \lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}^n} \mathrm{e}^{-\varepsilon |y|^2/2} \hat{\phi}(y) d^n y = \int_{\mathbb{R}^n} \hat{\phi}(y) d^n y,$$

which shows $\hat{\phi} \in L^1$. Thus ϕ is the Fourier transform of $\hat{\phi}(-y)$ establishing the claim in the case that ϕ is integrable.

In the general case we consider $\phi_\varepsilon(p) = \phi(p) \mathrm{e}^{-\varepsilon p^2/2}$ which is integrable since ϕ is bounded (Problem 8.17). Moreover, since $\mathrm{e}^{-\varepsilon p^2/2}$ gives rise to a positive definite kernel the same is true for the product (Lemma 3.30 (v) and the following example). Now ϕ_ε is the Fourier transform of a positive measure by the first part and so is its limit by Lévy's continuity theorem. $\square$

[2] Salomon Bochner (1899–1982), Austrian–Hungarian mathematician

Problem 8.11. *Show that $J_\nu(z)$ is a solution of the Bessel differential equation*

$$z^2 u'' + z u' + (z^2 - \nu^2) u = 0.$$

Prove the following properties of the Bessel functions.

(i) $(z^{\pm\nu} J_\nu(z))' = \pm z^{\pm\nu} J_{\nu\mp 1}(z)$.

(ii) $J_{\nu-1}(z) + J_{\nu+1}(z) = \frac{2\nu}{z} J_\nu(z)$.

(iii) $J_{\nu-1}(z) - J_{\nu+1}(z) = 2J_\nu'(z)$.

Problem 8.12. *Show the following integral representations for Bessel functions:*

$$\begin{aligned} J_\nu(z) &= \frac{1}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}\left(\frac{z}{2}\right)^\nu \int_0^\pi \mathrm{e}^{\pm \mathrm{i} z\cos(\theta)} \sin(\theta)^{2\nu} d\theta \\ &= \frac{1}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}\left(\frac{z}{2}\right)^\nu \int_0^\pi \cos(z\cos(\theta)) \sin(\theta)^{2\nu} d\theta \\ &= \frac{2}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}\left(\frac{z}{2}\right)^\nu \int_0^{\pi/2} \cos(z\cos(\theta)) \sin(\theta)^{2\nu} d\theta \\ &= \frac{2}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}\left(\frac{z}{2}\right)^\nu \int_0^{\pi/2} \cos(z\sin(\theta)) \cos(\theta)^{2\nu} d\theta, \end{aligned}$$

for $\mathrm{Re}(\nu) > -\frac{1}{2}$ and $z \in \mathbb{C}$. (Hint: Replace the outer cosine by its power series and use Problem 2.27.)

Problem 8.13. *Verify the expression for $J_{-1/2}(z)$ and conclude that*

$$\tilde{\mathcal{H}}_{-1/2}(f)(r) = \frac{1}{\pi}\int_0^\infty \cos(rs) f(s) ds = \hat{f}(r)$$

if we extend f to all of $\mathbb{R}$ such that $f(r) = f(-r)$. (Hint: (Problem 2.25).)

Problem 8.14. *Show that the formula derived in Example 8.4 holds for all $n \in \mathbb{N}$. (Hint: Use the power series for the Bessel function and Legendre's duplication formula from Problem 2.27.)*

Problem 8.15. *Suppose $\int_{\mathbb{R}^n} (1+|x|)^k d\mu(x) < \infty$ for some $k \in \mathbb{N}$. Show that $\hat{\mu} \in C^k(\mathbb{R}^n)$ and*

$$\partial_\alpha \hat{\mu}(p) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}px} (-\mathrm{i}x)^\alpha d\mu(x).$$

Problem 8.16. *Show $(\mu * \nu)^\wedge = (2\pi)^{n/2}\hat{\mu}\hat{\nu}$ (see Problem 3.34).*

Problem 8.17. *Show that a positive definite function satisfies $\phi(-p) = \phi(p)^*$ and $|\phi(p)| \le \phi(0)$. Moreover,*

$$|\phi(p) - \phi(q)| \le 4\phi(0)|\phi(0) - \phi(p-q)|.$$

(Hint: Look at the cases of two $\{0, p\}$ and three points $\{0, p, q\}$ and use the fact that the determinant of a positive semidefinite matrix is nonnegative.)

8.3. Applications to linear partial differential equations

By virtue of Lemma 8.4 the Fourier transform can be used to map linear partial differential equations with constant coefficients to algebraic equations, thereby providing a mean of solving them. To illustrate this procedure we look at the famous **Poisson equation**, that is, given a function g, find a function f satisfying

$$-\Delta f = g. \tag{8.41}$$

For simplicity, let us start by investigating this problem in the space of Schwartz functions $\mathcal{S}(\mathbb{R}^n)$. Assuming there is a solution we can take the Fourier transform on both sides to obtain

$$|p|^2 \hat{f}(p) = \hat{g}(p) \qquad \Rightarrow \qquad \hat{f}(p) = |p|^{-2}\hat{g}(p). \tag{8.42}$$

Since the right-hand side is integrable for $n \geq 3$ we obtain that our solution is necessarily given by

$$f(x) = (|p|^{-2}\hat{g}(p))^{\vee}(x). \tag{8.43}$$

In fact, this formula still works provided $g(x), |p|^{-2}\hat{g}(p) \in L^1(\mathbb{R}^n)$. Moreover, if we additionally assume $\hat{g} \in L^1(\mathbb{R}^n)$, then $|p|^2\hat{f}(p) = \hat{g}(p) \in L^1(\mathbb{R}^n)$ and Lemma 8.3 implies that $f \in C^2(\mathbb{R}^n)$ as well as that it is indeed a solution. Note that if $n \geq 3$, then $|p|^{-2}\hat{g}(p) \in L^1(\mathbb{R}^n)$ follows automatically from $g, \hat{g} \in L^1(\mathbb{R}^n)$ (show this!).

Moreover, we clearly expect that f should be given by a convolution. However, since $|p|^{-2}$ is not in $L^p(\mathbb{R}^n)$ for any p, the formulas derived so far do not apply.

Lemma 8.27. *Let $0 < \alpha < n$ and suppose $g \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ as well as $|p|^{-\alpha}\hat{g}(p) \in L^1(\mathbb{R}^n)$. Then*

$$(|p|^{-\alpha}\hat{g}(p))^{\vee}(x) = \int_{\mathbb{R}^n} I_\alpha(|x-y|) g(y) d^n y, \tag{8.44}$$

*where the **Riesz potential** is given by*

$$I_\alpha(r) := \frac{\Gamma(\frac{n-\alpha}{2})}{2^\alpha \pi^{n/2}\Gamma(\frac{\alpha}{2})} \frac{1}{r^{n-\alpha}}. \tag{8.45}$$

Proof. Note that, while $|.|^{-\alpha}$ is not in $L^p(\mathbb{R}^n)$ for any p, our assumption $0 < \alpha < n$ ensures that the singularity at zero is integrable.

We set $\phi_t(p) = \exp(-t|p|^2/2)$ and begin with the elementary formula

$$|p|^{-\alpha} = c_\alpha \int_0^\infty \phi_t(p) t^{\alpha/2-1} dt, \quad c_\alpha = \frac{1}{2^{\alpha/2}\Gamma(\alpha/2)},$$

which follows from the definition of the gamma function (Problem 2.24) after a simple scaling. Since $|p|^{-\alpha}\hat{g}(p)$ is integrable we can use Fubini and Lemma 8.6 to obtain

$$\begin{aligned}(|p|^{-\alpha}\hat{g}(p))^\vee(x) &= \frac{c_\alpha}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}xp}\left(\int_0^\infty \phi_t(p)t^{\alpha/2-1}dt\right)\hat{g}(p)d^np\\ &= \frac{c_\alpha}{(2\pi)^{n/2}}\int_0^\infty \left(\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}xp}\hat{\phi}_{1/t}(p)\hat{g}(p)d^np\right)t^{(\alpha-n)/2-1}dt.\end{aligned}$$

Since $\phi, g \in L^1$ we know by Lemma 8.12 that $\hat{\phi}\hat{g} = (2\pi)^{-n/2}(\phi * g)^\wedge$ Moreover, since $\hat{\phi}\hat{g} \in L^1$ Theorem 8.7 gives us $(\hat{\phi}\hat{g})^\vee = (2\pi)^{-n/2}\phi * g$. Thus, we can make a change of variables and use Fubini once again (since $g \in L^\infty$)

$$\begin{aligned}(|p|^{-\alpha}\hat{g}(p))^\vee(x) &= \frac{c_\alpha}{(2\pi)^{n/2}}\int_0^\infty \left(\int_{\mathbb{R}^n} \phi_{1/t}(x-y)g(y)d^ny\right)t^{(\alpha-n)/2-1}dt\\ &= \frac{c_\alpha}{(2\pi)^{n/2}}\int_0^\infty \left(\int_{\mathbb{R}^n} \phi_t(x-y)g(y)d^ny\right)t^{(n-\alpha)/2-1}dt\\ &= \frac{c_\alpha}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \left(\int_0^\infty \phi_t(x-y)t^{(n-\alpha)/2-1}dt\right)g(y)d^ny\\ &= \frac{c_\alpha/c_{n-\alpha}}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} \frac{g(y)}{|x-y|^{n-\alpha}}d^ny\end{aligned}$$

to obtain the desired result. □

Note that the conditions of the above theorem are, for example, satisfied if $g, \hat{g} \in L^1(\mathbb{R}^n)$ which holds, for example, if $g \in \mathcal{S}(\mathbb{R}^n)$. In summary, if $g \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, $|p|^{-2}\hat{g}(p) \in L^1(\mathbb{R}^n)$ and $n \geq 3$, then

$$f = \Phi * g \tag{8.46}$$

is a classical solution of the Poisson equation, where

$$\Phi(x) := \frac{\Gamma(\frac{n}{2}-1)}{4\pi^{n/2}}\frac{1}{|x|^{n-2}}, \qquad n \geq 3, \tag{8.47}$$

is known as the **fundamental solution** of the Laplace equation.

A few words about this formula are in order. First of all, our original formula in Fourier space shows that the multiplication with $|p|^{-2}$ improves the decay of $\hat{g}$ and hence, by virtue of Lemma 8.4, f should have, roughly speaking, two derivatives more than g. However, unless $\hat{g}(0)$ vanishes, multiplication with $|p|^{-2}$ will create a singularity at 0 and hence, again by Lemma 8.4, f will not inherit any decay properties from g. In fact, evaluating the above formula with $g = \chi_{B_1(0)}$ (Problem 8.18) shows that f might not decay better than Φ even for g with compact support.

Moreover, our conditions on g might not be easy to check as it will not be possible to compute $\hat{g}$ explicitly in general. So if one wants to deduce

$\hat{g} \in L^1(\mathbb{R}^n)$ from properties of g, one could use Lemma 8.4 together with the Riemann–Lebesgue lemma to show that this condition holds if $g \in C^k(\mathbb{R}^n)$, $k > n-2$, such that all derivatives are integrable and all derivatives of order less than k vanish at ∞ (Problem 8.19). This seems a rather strong requirement since our solution formula will already make sense under the sole assumption $g \in L^1(\mathbb{R}^n)$. However, as the example $g = \chi_{B_1(0)}$ shows, this *solution* might not be C^2 and hence one needs to weaken the notion of a solution if one wants to include such situations. This will lead us to the concepts of weak derivatives and Sobolev spaces. As a preparation we will develop some further tools which will allow us to investigate continuity properties of the operator $\mathcal{I}_\alpha f = I_\alpha * f$ in the next section.

Before that, let us summarize the procedure in the general case. Suppose we have the following linear partial differential equations with constant coefficients:

$$P(\mathrm{i}\partial) f = g, \qquad P(\mathrm{i}\partial) = \sum_{\alpha \le k} c_\alpha \mathrm{i}^{|\alpha|} \partial_\alpha. \tag{8.48}$$

Then the solution can be found via the procedure

$$\begin{tikzcd} \hat{g} \arrow[r] & P^{-1}\hat{g} \arrow[d, "\mathcal{F}^{-1}"] \\ g \arrow[u, "\mathcal{F}"] & f \end{tikzcd}$$

and is formally given by

$$f(x) = (P(p)^{-1}\hat{g}(p))^\vee(x). \tag{8.49}$$

It remains to investigate the properties of the solution operator. In general, given a locally integrable function $m \in L^1(\mathbb{R}^n, (1+|x|)^{-r} d^n x)$ for some $r > 0$ one might try to define a corresponding operator via

$$A_m g := (m\hat{g})^\vee, \tag{8.50}$$

for $g \in \mathcal{S}(\mathbb{R}^n)$. In this case m is known as a **Fourier multiplier** and it is said to be an L^p-multiplier if A_m can be extended to a bounded operator in $L^p(\mathbb{R}^n)$.

Example 8.7. Any $m \in L^\infty(\mathbb{R}^n)$ is an L^2 multiplier (in fact the converse is also true — Problem 8.20). $\diamond$

Example 8.8. If $m := (2\pi)^{n/2}\hat{\mu}$ is the Fourier transform of a complex Borel measure, then (by Problem 8.16)

$$A_m g = \mu * g$$

and m is an L^p-multiplier for all $1 \le p \le \infty$ by Problem 4.21.

It can be shown that in the case $p = 1$ every multiplier is of this form. ⋄

Another famous example which can be solved in this way is the **Helmholtz equation**

$$-\Delta f + f = g. \tag{8.51}$$

As before we find that if $g(x), (1+|p|^2)^{-1}\hat{g}(p) \in L^1(\mathbb{R}^n)$ then the solution is given by

$$f(x) = ((1+|p|^2)^{-1}\hat{g}(p))^\vee(x). \tag{8.52}$$

Lemma 8.28. *Let $\alpha > 0$. Suppose $g \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ as well as $(1+|p|^2)^{-\alpha/2}\hat{g}(p) \in L^1(\mathbb{R}^n)$. Then*

$$((1+|p|^2)^{-\alpha/2}\hat{g}(p))^\vee(x) = \int_{\mathbb{R}^n} J_\alpha(|x-y|)g(y)d^n y, \tag{8.53}$$

where the ***Bessel potential*** *is given by*

$$J_\alpha(r) := \frac{2}{(4\pi)^{n/2}\Gamma(\frac{\alpha}{2})}\left(\frac{r}{2}\right)^{-(n-\alpha)/2} K_{(n-\alpha)/2}(r), \qquad r > 0, \tag{8.54}$$

with

$$K_\nu(r) = K_{-\nu}(r) = \frac{1}{2}\left(\frac{r}{2}\right)^\nu \int_0^\infty \mathrm{e}^{-t-\frac{r^2}{4t}} \frac{dt}{t^{\nu+1}}, \qquad r > 0, \nu \in \mathbb{R}, \tag{8.55}$$

the modified Bessel function of the second kind of order ν ([**14**, (10.32.10)]*).*

Proof. We proceed as in the previous lemma. We set $\phi_t(p) = \exp(-t|p|^2/2)$ and begin with the elementary formula

$$\frac{\Gamma(\frac{\alpha}{2})}{(1+|p|^2)^{\alpha/2}} = \int_0^\infty t^{\alpha/2-1}\mathrm{e}^{-t(1+|p|^2)}dt.$$

Since $g, \hat{g}(p)$ are integrable we can use Fubini and Lemma 8.6 to obtain

$$\begin{aligned}
(\frac{\hat{g}(p)}{(1+|p|^2)^{\alpha/2}})^\vee(x) &= \frac{\Gamma(\frac{\alpha}{2})^{-1}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}xp}\left(\int_0^\infty t^{\alpha/2-1}\mathrm{e}^{-t(1+|p|^2)}dt\right)\hat{g}(p)d^n p \\
&= \frac{\Gamma(\frac{\alpha}{2})^{-1}}{(4\pi)^{n/2}} \int_0^\infty \left(\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}xp}\hat{\phi}_{1/2t}(p)\hat{g}(p)d^n p\right)\mathrm{e}^{-t}t^{(\alpha-n)/2-1}dt \\
&= \frac{\Gamma(\frac{\alpha}{2})^{-1}}{(4\pi)^{n/2}} \int_0^\infty \left(\int_{\mathbb{R}^n} \phi_{1/2t}(x-y)g(y)d^n y\right)\mathrm{e}^{-t}t^{(\alpha-n)/2-1}dt \\
&= \frac{\Gamma(\frac{\alpha}{2})^{-1}}{(4\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\int_0^\infty \phi_{1/2t}(x-y)\mathrm{e}^{-t}t^{(\alpha-n)/2-1}dt\right)g(y)d^n y
\end{aligned}$$

to obtain the desired result. Using Fubini in the last step is allowed since g is bounded and $J_\alpha(|x|) \in L^1(\mathbb{R}^n)$ (Problem 8.21). □

Note that the first condition $g \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ implies $g \in L^2(\mathbb{R}^n)$ and thus the second condition $(1+|p|^2)^{-\alpha/2}\hat{g}(p) \in L^1(\mathbb{R}^n)$ will be satisfied if $\frac{n}{2} < \alpha$.

In particular, if $g, \hat{g} \in L^1(\mathbb{R}^n)$, then

$$f = J_2 * g \tag{8.56}$$

is a solution of Helmholtz equation. Note that since our multiplier $(1+|p|^2)^{-1}$ does not have a singularity near zero, the solution f will preserve (some) decay properties of g. For example, it will map Schwartz functions to Schwartz functions and thus for every $g \in \mathcal{S}(\mathbb{R}^n)$ there is a unique solution of the Helmholtz equation $f \in \mathcal{S}(\mathbb{R}^n)$. This is also reflected by the fact that the Bessel potential decays much faster than the Riesz potential. Indeed, one can show that [**14**, (10.25.3)]

$$K_\nu(r) = \sqrt{\frac{\pi}{2r}}\mathrm{e}^{-r}(1 + O(r^{-1})) \tag{8.57}$$

as $r \to \infty$. The singularity near zero is of the same type as for I_α since (see [**14**, (10.30.2), (10.30.3)])

$$K_\nu(r) = \begin{cases} \frac{\Gamma(\nu)}{2}\left(\frac{r}{2}\right)^{-\nu} + O(r^{-\nu+2}), & \nu > 0, \\ -\log(\frac{r}{2}) + O(1), & \nu = 0, \end{cases} \tag{8.58}$$

for $r \to 0$.

Problem* 8.18. *Show that for $n = 3$ we have*

$$(\Phi * \chi_{B_1(0)})(x) = \begin{cases} \frac{1}{3|x|}, & |x| \ge 1, \\ \frac{3-|x|^2}{6}, & |x| \le 1. \end{cases}$$

(Hint: Observe that the result depends only on $|x|$. Then choose $x = (0,0,R)$ and evaluate the integral using spherical coordinates.)

Problem* 8.19. *Suppose $g \in C^k(\mathbb{R}^n)$ and $\partial_j^l g \in L^1(\mathbb{R}^n)$ for $j = 1,\dots,n$ and $0 \le l \le k$ as well as $\lim_{|x|\to\infty} \partial_j^l g(x) = 0$ for $j = 1,\dots,n$ and $0 \le l < k$. Then*

$$|\hat{g}(p)| \le \frac{C}{(1+|p|^2)^{k/2}}.$$

Problem* 8.20. *Show that m is an L^2 multiplier if and only if $m \in L^\infty(\mathbb{R}^n)$.*

Problem* 8.21. *Show*

$$\int_0^\infty J_\alpha(r) r^{n-1} dr = \frac{\Gamma(n/2)}{2\pi^{n/2}}, \quad \alpha > 0.$$

Conclude that

$$\|J_\alpha * g\|_p \le \|g\|_p.$$

(Hint: Fubini.)

8.4. Sobolev spaces

We have already introduced Sobolev spaces in Section 7.2. In this section we present an alternate (and in particular independent) approach to Sobolev spaces of index two (the Hilbert space case) on all of $\mathbb{R}^n$.

We begin by introducing the **Sobolev space**

$$H^r(\mathbb{R}^n) := \{f \in L^2(\mathbb{R}^n) | |p|^r \hat{f}(p) \in L^2(\mathbb{R}^n)\}. \tag{8.59}$$

The most important case is when r is an integer, however our definition makes sense for any $r \geq 0$. Moreover, note that $H^r(\mathbb{R}^n)$ becomes a Hilbert space if we introduce the scalar product

$$\langle f, g\rangle_{H^r} := \int_{\mathbb{R}^n} \hat{f}(p)^* \hat{g}(p)(1 + |p|^2)^r d^n p. \tag{8.60}$$

In particular, note that by construction $\mathcal{F}$ maps $H^r(\mathbb{R}^n)$ unitarily onto $L^2(\mathbb{R}^n, \langle p\rangle^r d^n p)$, where (sometimes known as **Japanese bracket**)

$$\langle p\rangle := (1 + |p|^2)^{1/2}, \qquad p \in \mathbb{C}^n. \tag{8.61}$$

Clearly $H^{r+1}(\mathbb{R}^n) \subset H^r(\mathbb{R}^n)$ 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 $L^2(\mathbb{R}^n, \langle p\rangle^r d^n p)$).

The motivation for the definition (8.59) stems from Lemma 8.4 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

$$\partial_\alpha f := ((\mathrm{i}p)^\alpha \hat{f}(p))^\vee, \qquad f \in H^r(\mathbb{R}^n),\ |\alpha| \leq r. \tag{8.62}$$

By Lemma 8.4 this definition coincides with the usual one for every $f \in \mathcal{S}(\mathbb{R}^n)$.

Example 8.9. 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 $f'(x) = -\operatorname{sign}(x)\chi_{[-1,1]}(x)$. ⋄

We also have

$$\begin{aligned}\int_{\mathbb{R}^n} g(x)(\partial_\alpha f)(x) d^n x &= \langle g^*, (\partial_\alpha f)\rangle = \langle \hat{g}(p)^*, (\mathrm{i}p)^\alpha \hat{f}(p)\rangle \\ &= (-1)^{|\alpha|}\langle (\mathrm{i}p)^\alpha \hat{g}(p)^*, \hat{f}(p)\rangle = (-1)^{|\alpha|}\langle \partial_\alpha g^*, f\rangle \\ &= (-1)^{|\alpha|}\int_{\mathbb{R}^n} (\partial_\alpha g)(x) f(x) d^n x, \end{aligned} \tag{8.63}$$

for $f, g \in H^r(\mathbb{R}^n)$. Furthermore, recall that a function $h \in L^1_{loc}(\mathbb{R}^n)$ satisfying

$$\int_{\mathbb{R}^n} \varphi(x) h(x) d^n x = (-1)^{|\alpha|} \int_{\mathbb{R}^n} (\partial_\alpha \varphi)(x) f(x) d^n x, \quad \forall \varphi \in C_c^\infty(\mathbb{R}^n), \tag{8.64}$$

is called the **weak derivative** or the derivative in the sense of distributions of f (by Lemma 3.23 such a function is unique if it exists). Hence, choosing $g = \varphi$ in (8.63), we see that functions in $H^r(\mathbb{R}^n)$ have weak derivatives up to order r, which are in $L^2(\mathbb{R}^n)$. Moreover, the weak derivatives coincide with the derivatives defined in (8.62). Conversely, given (8.64) with $f, h \in L^2(\mathbb{R}^n)$ we can use that $\mathcal{F}$ is unitary to conclude $\int_{\mathbb{R}^n} \hat{\varphi}(p)\hat{h}(p)d^np = \int_{\mathbb{R}^n}(\mathrm{i}p)^\alpha\hat{\varphi}(p)\hat{f}(p)d^np$ for all $\varphi \in C_c^\infty(\mathbb{R}^n)$. By approximation this follows for $\varphi \in H^r(\mathbb{R}^n)$ with $r \ge |\alpha|$ and hence in particular for $\hat{\varphi} \in C_c^\infty(\mathbb{R}^n)$. Consequently $(\mathrm{i}p)^\alpha\hat{f}(p) = \hat{h}(p)$ a.e. implying that $f \in H^r(\mathbb{R}^n)$ if all weak derivatives exist up to order r and 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:

$$\|f\|_{m,2}^2 := \sum_{|\alpha|\le m} \|\partial_\alpha f\|_2^2. \tag{8.65}$$

By $|p^\alpha| \le |p|^{|\alpha|} \le (1+|p|^2)^{m/2}$ it follows that this norm is equivalent to (8.60).

Example 8.10. This definition of a weak derivative is tailored for the method of solving linear constant coefficient partial differential equations as outlined in Section 8.3. While Lemma 8.3 only gives us a sufficient condition on $\hat{f}$ for f to be differentiable, the weak derivatives gives us necessary and sufficient conditions. For example, we see that the Poisson equation (8.41) will have a (unique) solution $f \in H^2(\mathbb{R}^n)$ if and only if $|p|^{-2}\hat{g} \in L^2(\mathbb{R}^n)$. That this is not true for all $g \in L^2(\mathbb{R}^n)$ is connected with the fact that $|p|^{-2}$ is unbounded and hence no L^2 multiplier (cf. Problem 8.20). Consequently the range of Δ when defined on $H^2(\mathbb{R}^n)$ will not be all of $L^2(\mathbb{R}^n)$ and hence the Poisson equation is not solvable within the class $H^2(\mathbb{R}^n)$ for all $g \in L^2(\mathbb{R}^n)$. Nevertheless, we get a unique weak solution under some conditions. Under which conditions this weak solution is also a classical solution can then be investigated separately.

Note that the situation is even simpler for the Helmholtz equation (8.51) since the corresponding multiplier $(1+|p|^2)^{-1}$ does map L^2 to L^2. Hence we get that the Helmholtz equation has a unique solution $f \in H^2(\mathbb{R}^n)$ if and only if $g \in L^2(\mathbb{R}^n)$. Moreover, $f \in H^{r+2}(\mathbb{R}^n)$ if and only if $g \in H^r(\mathbb{R}^n)$. ⋄

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 $\partial_\alpha f(x) \in C_0(\mathbb{R}^n)$ provided $p^\alpha\hat{f}(p) \in L^1(\mathbb{R}^n)$. Moreover, in this situation the derivatives will exist as classical derivatives:

Lemma 8.29. *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 $f \in C_0^k(\mathbb{R}^n)$, the set of functions with* *continuous partial derivatives of order k all of which vanish at ∞. Moreover, the classical and weak derivatives coincide in this case.*

Proof. We begin by observing that by Lemma 8.11

$$f(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}px} \hat{f}(p) d^n p.$$

Now the claim follows as in the proof of Lemma 8.4 by differentiating the integral using Problem 2.13. □

Now we are able to prove the following embedding theorem.

Theorem 8.30 (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 $C_0^k(\mathbb{R}^n)$ with*

$$\|\partial_\alpha f\|_\infty \le C_{n,r} \|f\|_{H^r}, \qquad |\alpha| \le k. \tag{8.66}$$

Proof. 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 $s > \frac{n}{2}$ (see Example 2.16) and $\langle p\rangle^{|\alpha|+s} |\hat{f}(p)| \in L^2(\mathbb{R}^n)$ if $s + |\alpha| \le r$. Hence $\langle p\rangle^{|\alpha|} |\hat{f}(p)| \in L^1(\mathbb{R}^n)$ and the claim follows from the previous lemma. □

In fact, we can even do a bit better.

Lemma 8.31 (Morrey inequality). *Suppose $f \in H^{n/2+\gamma}(\mathbb{R}^n)$ for some $\gamma \in (0,1)$. Then $f \in C_0^{0,\gamma}(\mathbb{R}^n)$, the set of functions which are Hölder continuous with exponent γ and vanish at ∞. Moreover,*

$$|f(x) - f(y)| \le C_{n,\gamma} \|\hat{f}\|_{H^{n/2+\gamma}} |x-y|^\gamma \tag{8.67}$$

in this case.

Proof. We begin with

$$f(x+y) - f(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}px} (\mathrm{e}^{\mathrm{i}py} - 1) \hat{f}(p) d^n p$$

implying

$$|f(x+y) - f(x)| \le \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \frac{|\mathrm{e}^{\mathrm{i}py} - 1|}{\langle p\rangle^{n/2+\gamma}} \langle p\rangle^{n/2+\gamma} |\hat{f}(p)| d^n p.$$

Hence, after applying Cauchy–Schwarz, it remains to estimate (recall (2.42))

$$\begin{aligned}\int_{\mathbb{R}^n} \frac{|\mathrm{e}^{\mathrm{i}py} - 1|^2}{\langle p\rangle^{n+2\gamma}} d^n p &\le S_n \int_0^{1/|y|} \frac{(|y|r)^2}{r^{n+2\gamma}} r^{n-1} dr + S_n \int_{1/|y|}^\infty \frac{4}{r^{n+2\gamma}} r^{n-1} dr \\ &= \frac{S_n}{2(1-\gamma)} |y|^{2\gamma} + \frac{2S_n}{\gamma} |y|^{2\gamma} = \frac{S_n(4-3\gamma)}{2\gamma(1-\gamma)} |y|^{2\gamma},\end{aligned}$$

where $S_n = nV_n$ is the surface area of the unit sphere in $\mathbb{R}^n$. □

Using this lemma we immediately obtain:

Corollary 8.32. *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 $C_0^{k,\gamma}(\mathbb{R}^n)$, the set of functions in $C_0^k(\mathbb{R}^n)$ whose highest derivatives are Hölder continuous of exponent γ.*

In the case $r > \frac{n}{2}$, when Morrey's inequality holds, we even get that $H^r(\mathbb{R}^n)$ is a Banach algebra:

Lemma 8.33. *Suppose $f, g \in H^r(\mathbb{R}^n)$ and $\hat{f}, \hat{g} \in L^1(\mathbb{R}^n)$ for some $r \ge 0$. Then we have $fg \in H^r(\mathbb{R}^n)$ with*

$$\|fg\|_{H^r} \le C_{n,r}\big(\|f\|_{H^r}\|\hat{g}\|_1 + \|\hat{f}\|_1\|g\|_{H^r}\big). \tag{8.68}$$

Moreover, if $r > \frac{n}{2}$, then $\|\hat{f}\|_1 \le c_{n,r}\|f\|_{H^r}$ and we get

$$\|fg\|_{H^r} \le \tilde{C}_{n,r}\|f\|_{H^r}\|g\|_{H^r}. \tag{8.69}$$

Proof. Note that we have

$$\begin{aligned}\langle p\rangle^r &\le (1 + 2|p-q|^2 + 2|q|^2)^{r/2} \le 2^{r/2}(1 + |p-q|^2 + 1 + |q|^2)^{r/2}\\ &\le c_r\langle p-q\rangle^r + \langle q\rangle^r\end{aligned}$$

for $c_r = \max(2^{r-1}, 2^{r/2})$ and $r \ge 0$. Hence by Corollary 8.15 we have

$$\begin{aligned}(2\pi)^{n/2}\langle p\rangle^r|(fg)^\wedge(p)| = \langle p\rangle^r|(\hat{f} * \hat{g})(p)| &\le \int_{\mathbb{R}^n}\langle p\rangle^r|\hat{f}(p-q)||\hat{g}(q)|d^nq\\ &\le c_r\big((|\langle.\rangle^r\hat{f}| * |\hat{g}|)(p) + (|\hat{f}| * |\langle.\rangle^r\hat{g}|)(p)\big)\end{aligned}$$

and the first claim with $C_{n,r} = c_r(2\pi)^{-n/2}$ follows from Young's inequality (3.25). The second claim follows from Cauchy–Schwarz $\|\hat{f}\|_1 \le \|\langle.\rangle^{-r}\|_2\|f\|_{H^r}$ as in the proof of Theorem 8.30. □

Example 8.11. The function $f(x) = \log(|x|)$ is in $H^1(\mathbb{R}^n)$ for $n \ge 3$. In fact, the weak derivatives are given by

$$\partial_j f(x) = \frac{x_j}{|x|^2}. \tag{8.70}$$

However, observe that f is not continuous. ◇

The last example shows that in the case $r < \frac{n}{2}$ functions in H^r are no longer necessarily continuous. In this case we at least get an embedding into some *better* L^p space:

Theorem 8.34 (Sobolev inequality)**.** *Suppose $0 < r < \frac{n}{2}$. Then $H^r(\mathbb{R}^n)$ is continuously embedded into $L^p(\mathbb{R}^n)$ with $p = \frac{2n}{n-2r}$, that is,*

$$\|f\|_p \le \tilde{C}_{n,r}\||.|^r\hat{f}(.)\|_2 \le C_{n,r}\|f\|_{H^r}. \tag{8.71}$$

Proof. We will give a prove based on the Hardy–Littlewood–Sobolev inequality to be proven in Theorem 9.10 below.

It suffices to prove the first inequality. Set $|p|^r\hat{f}(p) = \hat{g}(p) \in L^2$. Moreover, choose some sequence $f_m \in \mathcal{S} \to f \in H^r$. Then, by Lemma 8.27 $f_m = \mathcal{I}_r g_m$, and since the Hardy–Littlewood–Sobolev inequality implies that the map $\mathcal{I}_r : L^2 \to L^p$ is continuous, we have $\|f_m\|_p = \|\mathcal{I}_r g_m\|_p \le \tilde{C}\|g_m\|_2 = \tilde{C}\|\hat{g}_m\|_2 = \tilde{C}\||p|^r\hat{f}_m(p)\|_2$ and the claim follows after taking limits. □

Finally, observe that one can also include polynomial decay (cf. Problem 8.24) by defining the **weighted Sobolev spaces**

$$H^{r,s}(\mathbb{R}^n) = \{f \in H^r(\mathbb{R}^n) | \hat{f} \in H^s(\mathbb{R}^n)\}, \qquad s, r > 0, \tag{8.72}$$

with associated norm

$$\|f\|_{H^{r,s}}^2 := \|f\|_{H^r}^2 + \|\hat{f}\|_{H^s}^2 = \|\langle p\rangle^r \hat{f}(p)\|_2^2 + \|\langle x\rangle^s f(x)\|_2^2. \tag{8.73}$$

Clearly $H^{r,s}(\mathbb{R}^n)$ is a Hilbert space and $\mathcal{S}(\mathbb{R}^n) \subset H^{r,s}(\mathbb{R}^n)$ is dense. Moreover, the Fourier transform $\mathcal{F} : H^{r,s}(\mathbb{R}^n) \to H^{r,s}(\mathbb{R}^n)$ is unitary and $H^{s,r}(\mathbb{R}^n)$ is a Banach algebra if $r > \frac{n}{2}$ (Problem 8.25).

Problem 8.22. *Use dilations $f(x) \mapsto f(\lambda x)$, $\lambda > 0$, to show that $p = \frac{2n}{n-2r}$ is the only index for which the Sobolev inequality $\|f\|_p \le \tilde{C}_{n,r}\||p|^r\hat{f}(p)\|_2$ can hold.*

Problem 8.23. *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 8.24. *Show that the norm in $H^{k,k}$, $k \in \mathbb{N}_0$ is equivalent to the norm*

$$\|f\|^2 := \sum_{|\alpha|+|\beta|=k} \|x^\alpha \partial_\beta f(x)\|_2^2.$$

Problem 8.25. *Show that $H^{r,s}(\mathbb{R}^n)$ is a Banach algebra if $r > \frac{n}{2}$. (Hint: Theorem 8.30 and Lemma 8.33.)*

8.5. Applications to evolution equations

In this section we want to show how to apply these considerations to evolution equations. As a prototypical example we start with the Cauchy problem for the **heat equation**

$$u_t - \Delta u = 0, \qquad u(0) = g. \tag{8.74}$$

It turns out useful to view $u(t, x)$ as a function of t with values in a Banach space X. To this end we let $I \subseteq \mathbb{R}$ be some interval and denote by $C(I, X)$

the set of continuous functions from I to X. Given $t \in I$ we call $u : I \to X$ differentiable at t if the limit

$$\dot{u}(t) = \lim_{\varepsilon \to 0} \frac{u(t+\varepsilon) - u(t)}{\varepsilon} \tag{8.75}$$

exists. The set of functions $u : I \to X$ which are differentiable at all $t \in I$ and for which $\dot{u} \in C(I, X)$ is denoted by $C^1(I, X)$. As usual we set $C^{k+1}(I, X) = \{u \in C^1(I, x) | \dot{u} \in C^k(I, X)\}$. Note that if $U \in \mathscr{L}(X, Y)$ and $u \in C^1(I, X)$, then $Uu \in C^1(I, Y)$ and $\frac{d}{dt} Uu = U\dot{u}$.

A **strongly continuous operator semigroup** (also C_0-semigroup) is a family of operators $T(t) \in \mathscr{L}(X)$, $t \geq 0$, such that

(i) $T(t)g \in C([0, \infty), X)$ for every $g \in X$ (strong continuity) and
(ii) $T(0) = \mathbb{I}$, $T(t+s) = T(t)T(s)$ for every $t, s \geq 0$ (semigroup property).

Given a strongly continuous semigroup we can define its **generator** A as the linear operator

$$Af = \lim_{t \downarrow 0} \frac{1}{t}\big(T(t)f - f\big) \tag{8.76}$$

where the domain $\mathfrak{D}(A)$ is precisely the set of all $f \in X$ for which the above limit exists. The key result is that if A generates a C_0-semigroup $T(t)$, then $u(t) := T(t)g$ will be the unique solution of the corresponding abstract Cauchy problem. More precisely we have (see Lemma 11.7 from [**23**]):

Lemma 8.35. *Let $T(t)$ be a C_0-semigroup with generator A. If $g \in X$ with $u(t) = T(t)g \in \mathfrak{D}(A)$ for $t > 0$ then $u(t) \in C^1((0, \infty), X) \cap C([0, \infty), X)$ and $u(t)$ is the unique solution of the abstract Cauchy problem*

$$\dot{u}(t) = Au(t), \qquad u(0) = g. \tag{8.77}$$

This is, for example, the case if $g \in \mathfrak{D}(A)$ in which case we even have $u(t) \in C^1([0, \infty), X)$.

After these preparations we are ready to return to our original problem (8.74). Let $g \in L^2(\mathbb{R}^n)$ and let $u \in C^1((0, \infty), L^2(\mathbb{R}^n))$ be a solution such that $u(t) \in H^2(\mathbb{R}^n)$ for $t > 0$. Then we can take the Fourier transform to obtain

$$\hat{u}_t + |p|^2 \hat{u} = 0, \qquad \hat{u}(0) = \hat{g}. \tag{8.78}$$

Next, one verifies (Problem 8.26) that the solution (in the sense defined above) of this differential equation is given by

$$\hat{u}(t)(p) = \hat{g}(p) \mathrm{e}^{-|p|^2 t}. \tag{8.79}$$

Accordingly, the solution of our original problem is

$$u(t) = T_H(t)g, \qquad T_H(t) = \mathcal{F}^{-1} \mathrm{e}^{-|p|^2 t} \mathcal{F}. \tag{8.80}$$

Note that $T_H(t) : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a bounded linear operator with $\|T_H(t)\| \le 1$ (since $|\mathrm{e}^{-|p|^2 t}| \le 1$). In fact, for $t > 0$ we even have $T_H(t)g \in H^r(\mathbb{R}^n)$ for any $r \ge 0$ showing that $u(t)$ is smooth even for *rough* initial functions g. In summary,

Theorem 8.36. *The family $T_H(t)$ is a C_0-semigroup on $L^2(\mathbb{R}^n)$ whose generator is Δ, $\mathfrak{D}(\Delta) = H^2(\mathbb{R}^n)$.*

Proof. That $H^2(\mathbb{R}^n) \subseteq \mathfrak{D}(A)$ follows from Problem 8.26. Conversely, let $g \notin H^2(\mathbb{R}^n)$. Then $t^{-1}(\mathrm{e}^{-|p|^2 t} - 1) \to -|p|^2$ uniformly on every compact subset $K \subset \mathbb{R}^n$. Hence $\int_K |p|^2 |\hat{g}(p)|^2 d^n p = \int_K |Ag(x)|^2 d^n x$ which gives a contradiction as K increases. □

Next we want to derive a more explicit formula for our solution. To this end we assume $g \in L^1(\mathbb{R}^n)$ and introduce

$$\Phi_t(x) = \frac{1}{(4\pi t)^{n/2}} \mathrm{e}^{-\frac{|x|^2}{4t}}, \tag{8.81}$$

known as the **fundamental solution** of the heat equation, such that

$$\hat{u}(t) = (2\pi)^{n/2} \hat{g} \hat{\Phi}_t = (\Phi_t * g)^\wedge \tag{8.82}$$

by Lemma 8.6 and Lemma 8.12. Finally, by injectivity of the Fourier transform (Theorem 8.7) we conclude

$$u(t) = \Phi_t * g. \tag{8.83}$$

Moreover, one can check directly that (8.83) defines a solution for arbitrary $g \in L^p(\mathbb{R}^n)$.

Theorem 8.37. *Suppose $g \in L^p(\mathbb{R}^n)$, $1 \le p \le \infty$. Then (8.83) defines a solution for the heat equation which satisfies $u \in C^\infty((0,\infty) \times \mathbb{R}^n)$. The solutions has the following properties:*

(i) *If $1 \le p < \infty$, then $\lim_{t \downarrow 0} u(t) = g$ in L^p. If $p = \infty$ this holds for $g \in C_0(\mathbb{R}^n)$.*

(ii) *If $p = \infty$, then*
$$\|u(t)\|_\infty \le \|g\|_\infty. \tag{8.84}$$
If g is real-valued then so is u and
$$\inf g \le u(t) \le \sup g. \tag{8.85}$$

(iii) *(Mass conservation) If $p = 1$, then*
$$\int_{\mathbb{R}^n} u(t,x) d^n x = \int_{\mathbb{R}^n} g(x) d^n x \tag{8.86}$$
and
$$\|u(t)\|_\infty \le \frac{1}{(4\pi t)^{n/2}} \|g\|_1. \tag{8.87}$$

Proof. That $u \in C^\infty$ follows since $\Phi \in C^\infty$ from Problem 2.13. To see the remaining claims we begin by noting (by Problem 2.23)

$$\int_{\mathbb{R}^n} \Phi_t(x) d^n x = 1. \tag{8.88}$$

Now (i) follows from Lemma 3.21, (ii) is immediate, and (iii) follows from Fubini. □

Note that using Young's inequality (9.10) (to be established below) we even have

$$\|u(t)\|_\infty \le \|\Phi_t\|_q \|g\|_p = \frac{1}{q^{\frac{n}{2q}} (4\pi t)^{\frac{n}{2p}}} \|g\|_p, \qquad \frac{1}{p} + \frac{1}{q} = 1. \tag{8.89}$$

Another closely related equation is the **Schrödinger equation**

$$\mathrm{i} u_t + \Delta u = 0, \qquad u(0) = g. \tag{8.90}$$

As before we obtain that the solution for $g \in H^2(\mathbb{R}^n)$ is given by

$$u(t) = T_S(t) g, \qquad T_S(t) = \mathcal{F}^{-1} \mathrm{e}^{-\mathrm{i}|p|^2 t} \mathcal{F}. \tag{8.91}$$

Note that $T_S(t) : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a unitary operator (since $|\mathrm{e}^{-\mathrm{i}|p|^2 t}| = 1$):

$$\|u(t)\|_2 = \|g\|_2. \tag{8.92}$$

However, while we have $T_S(t) g \in H^r(\mathbb{R}^n)$ whenever $g \in H^r(\mathbb{R}^n)$, unlike the heat equation, the Schrödinger equation does only preserve but not improve the regularity of the initial condition.

Theorem 8.38. *The family $T_S(t)$ is a C_0-group on $L^2(\mathbb{R}^n)$ whose generator is $\mathrm{i}\Delta$, $\mathfrak{D}(\mathrm{i}\Delta) = H^2(\mathbb{R}^n)$.*

As in the case of the heat equation, we would like to express our solution as a convolution with the initial condition. However, now we run into the problem that $\mathrm{e}^{-\mathrm{i}|p|^2 t}$ is not integrable. To overcome this problem we consider

$$f_\varepsilon(p) = \mathrm{e}^{-(\mathrm{i}t + \varepsilon) p^2}, \quad \varepsilon > 0. \tag{8.93}$$

Then, as before we have

$$(f_\varepsilon \hat{g})^\vee(x) = \frac{1}{(4\pi(\mathrm{i}t + \varepsilon))^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\frac{|x-y|^2}{4(\mathrm{i}t+\varepsilon)}} g(y) d^n y \tag{8.94}$$

and hence

$$T_S(t) g(x) = \frac{1}{(4\pi \mathrm{i} t)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}\frac{|x-y|^2}{4t}} g(y) d^n y \tag{8.95}$$

for $t \neq 0$ and $g \in L^2(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$. In fact, letting $\varepsilon \downarrow 0$ the left-hand side converges to $T_S(t) g$ in L^2 and the limit of the right-hand side exists pointwise by dominated convergence and its pointwise limit must thus be equal to its L^2 limit.

Using this explicit form, we can again draw some further consequences. For example, if $g \in L^2(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$, then $u(t) \in C(\mathbb{R}^n)$ for $t \neq 0$ (use dominated convergence and continuity of the exponential) and satisfies

$$\|u(t)\|_\infty \le \frac{1}{|4\pi t|^{n/2}} \|g\|_1. \tag{8.96}$$

Thus we have again spreading of wave functions in this case.

Finally we turn to the **wave equation**

$$u_{tt} - \Delta u = 0, \qquad u(0) = g, \quad u_t(0) = f. \tag{8.97}$$

This equation will fit into our framework once we transform it to a first order system with respect to time:

$$u_t = v, \quad v_t = \Delta u, \qquad u(0) = g, \quad v(0) = f. \tag{8.98}$$

After applying the Fourier transform this system reads

$$\hat{u}_t = \hat{v}, \quad \hat{v}_t = -|p|^2 \hat{u}, \qquad \hat{u}(0) = \hat{g}, \quad \hat{v}(0) = \hat{f}, \tag{8.99}$$

and the solution is given by

$$\begin{aligned} \hat{u}(t,p) &= \cos(t|p|)\hat{g}(p) + \frac{\sin(t|p|)}{|p|} \hat{f}(p), \\ \hat{v}(t,p) &= -\sin(t|p|)|p|\hat{g}(p) + \cos(t|p|)\hat{f}(p). \end{aligned} \tag{8.100}$$

Hence for $(g,f) \in H^2(\mathbb{R}^n) \oplus H^1(\mathbb{R}^n)$ our solution is given by

$$\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = T_W(t) \begin{pmatrix} g \\ f \end{pmatrix}, \quad T_W(t) = \mathcal{F}^{-1} \begin{pmatrix} \cos(t|p|) & \frac{\sin(t|p|)}{|p|} \\ -\sin(t|p|)|p| & \cos(t|p|) \end{pmatrix} \mathcal{F}. \tag{8.101}$$

Theorem 8.39. *The family $T_W(t)$ is a C_0-semigroup on $H^1(\mathbb{R}^n) \oplus L^2(\mathbb{R}^n)$ whose generator is $A = \left(\begin{smallmatrix} 0 & 1 \\ \Delta & 0 \end{smallmatrix}\right)$, $\mathfrak{D}(A) = H^2(\mathbb{R}^n) \oplus H^1(\mathbb{R}^n)$.*

Note that we only get the bounds

$$\|u(t)\|_2 \le \|g\|_2 + |t|\,\|f\|_2, \qquad \|v(t)\|_2 \le \|\partial g\|_2 + \|f\|_2,$$

and, in particular, we do not have a contraction. To get a contraction we can use w defined via $\hat{w}(p) = |p|\hat{v}(p)$ instead of v. Then

$$\begin{pmatrix} u(t) \\ w(t) \end{pmatrix} = \tilde{T}_W(t) \begin{pmatrix} g \\ h \end{pmatrix}, \quad \tilde{T}_W(t) = \mathcal{F}^{-1} \begin{pmatrix} \cos(t|p|) & \sin(t|p|) \\ -\sin(t|p|) & \cos(t|p|) \end{pmatrix} \mathcal{F}, \tag{8.102}$$

where h is defined via $\hat{h} = |p|\hat{f}$. In this case $\tilde{T}_W$ is unitary and thus

$$\|u(t)\|_2^2 + \|w(t)\|_2^2 = \|g\|_2^2 + \|h\|_2^2. \tag{8.103}$$

However, since multiplication with $|p|$ is not surjective from $H^1(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$ the corresponding space is a bit tricky to define in the original x coordinates. In this respect note that $\|w\|_2^2 = \langle w, w\rangle = \langle \hat{w}, \hat{w}\rangle = \langle \hat{v}, |p|^2 \hat{v}\rangle =$

$\sum_j \langle \partial_j v, \partial_j v\rangle$ but this norm is not equivalent to the H^1 norm on $\mathbb{R}^n$ as there is no Poincaré inequality on $\mathbb{R}^n$. This problem does not occur when one considers the **Klein–Gordon equation**

$$u_{tt} - \Delta u + m\, u = 0, \qquad u(0) = g, \quad u_t(0) = f, \tag{8.104}$$

with $m > 0$. The analysis stays the same and all one has to do is replace $|p|$ by $\sqrt{|p|^2 + m}$ throughout.

If $n = 1$ we have $\frac{\sin(t|p|)}{|p|} \in L^2(\mathbb{R})$ and hence we can get an expression in terms of convolutions. In fact, since the inverse Fourier transform of $\frac{\sin(t|p|)}{|p|}$ is $\sqrt{\frac{\pi}{2}}\chi_{[-1,1]}(p/t)$, we obtain

$$u(t,x) = \int_{\mathbb{R}} \frac{1}{2}\chi_{[-t,t]}(x-y) f(y) dy = \frac{1}{2}\int_{x-t}^{x+t} f(y) dy$$

in the case $g = 0$. But the corresponding expression for $f = 0$ is just the time derivative of this expression and thus

$$\begin{aligned} u(t,x) &= \frac{1}{2}\frac{\partial}{\partial t}\int_{x-t}^{x+t} g(y) dy + \frac{1}{2}\int_{x-t}^{x+t} f(y) dy \\ &= \frac{g(x+t) + g(x-t)}{2} + \frac{1}{2}\int_{x-t}^{x+t} f(y) dy, \end{aligned} \tag{8.105}$$

which is known as **d'Alembert's formula**.

To obtain the corresponding formula in $n = 3$ dimensions we use the following observation

$$\frac{\partial}{\partial t}\hat{\varphi}_t(p) = \frac{\sin(t|p|)}{|p|}, \qquad \hat{\varphi}_t(p) := \frac{1 - \cos(t|p|)}{|p|^2}, \tag{8.106}$$

where $\hat{\varphi}_t \in L^2(\mathbb{R}^3)$. Then for $f \in H^r(\mathbb{R}^3)$ we have $U(t) := (2\pi)^{-3/2}\varphi_t * f \in C^1(\mathbb{R}, H^{r+1}(\mathbb{R}^3))$ with $\dot{U} = u$, the solution (in the case $g = 0$) we are looking for. Hence it remains to compute its inverse Fourier transform

$$\varphi_t(x) = \lim_{R\to\infty} \frac{1}{(2\pi)^{3/2}} \int_{B_R(0)} \hat{\varphi}_t(p) \mathrm{e}^{\mathrm{i}px} d^3 p \tag{8.107}$$

using spherical coordinates (without loss of generality we can rotate our coordinate system, such that the third coordinate direction is parallel to x)

$$\varphi_t(x) = \lim_{R\to\infty} \frac{1}{(2\pi)^{3/2}} \int_0^R \int_0^\pi \int_0^{2\pi} \frac{1-\cos(tr)}{r^2} \mathrm{e}^{\mathrm{i}r|x|\cos(\theta)} r^2 \sin(\theta) d\varphi d\theta dr. \tag{8.108}$$

Evaluating the integrals we obtain

$$\begin{aligned}
\varphi_t(x) &= \lim_{R\to\infty} \frac{1}{\sqrt{2\pi}} \int_0^R (1-\cos(tr)) \left(\int_0^\pi \mathrm{e}^{\mathrm{i}r|x|\cos(\theta)} \sin(\theta) d\theta \right) dr \\
&= \lim_{R\to\infty} \sqrt{\frac{2}{\pi}} \int_0^R (1-\cos(tr)) \frac{\sin(r|x|)}{|x| r} dr \\
&= \lim_{R\to\infty} \frac{1}{\sqrt{2\pi}} \int_0^R \left(2\frac{\sin(r|x|)}{|x|r} + \frac{\sin(r(t-|x|))}{|x|r} - \frac{\sin(r(t+|x|))}{|x|r} \right) dr, \\
&= \lim_{R\to\infty} \frac{1}{\sqrt{2\pi}|x|} \left(2\operatorname{Si}(R|x|) + \operatorname{Si}(R(t-|x|)) - \operatorname{Si}(R(t+|x|)) \right),
\end{aligned} \tag{8.109}$$

where

$$\operatorname{Si}(z) = \int_0^z \frac{\sin(x)}{x} dx \tag{8.110}$$

is the **sine integral**. Using $\operatorname{Si}(-x) = -\operatorname{Si}(x)$ for $x \in \mathbb{R}$ and (Problem 8.31)

$$\lim_{x\to\infty} \operatorname{Si}(x) = \frac{\pi}{2} \tag{8.111}$$

we finally obtain (since the pointwise limit must equal the L^2 limit)

$$\varphi_t(x) = \sqrt{\frac{\pi}{2}} \frac{\chi_{[0,|t|]}(|x|)}{|x|}. \tag{8.112}$$

For the wave equation this implies (assuming $f \in H^2(\mathbb{R}^3)$ such that it is continuous and using Lemma 2.18 to compute the derivative pointwise)

$$\begin{aligned}
u(t,x) &= \frac{1}{4\pi} \frac{\partial}{\partial t} \int_{B_{|t|}(x)} \frac{1}{|x-y|} f(y) d^3 y \\
&= \frac{1}{4\pi} \frac{\partial}{\partial t} \int_0^{|t|} \int_{S^2} \frac{1}{r} f(x+r\omega) r^2 d\sigma^2(\omega) dr \\
&= \frac{t}{4\pi} \int_{S^2} f(x+t\omega) d\sigma^2(\omega).
\end{aligned} \tag{8.113}$$

If merely $f \in H^1(\mathbb{R}^3)$ the integrand has to be understood in the sense of traces (Theorem 7.22). Thus we finally arrive at **Kirchhoff's formula**.

$$\begin{aligned}
u(t,x) &= \frac{\partial}{\partial t} \frac{t}{4\pi} \int_{S^2} g(x+t\omega) d\sigma^2(\omega) + \frac{t}{4\pi} \int_{S^2} f(x+t\omega) d\sigma^2(\omega) \\
&= \frac{1}{4\pi} \int_{S^2} \big(g + t\, \partial g \cdot \omega + t\, f \big)(x+t\omega) d\sigma^2(\omega),
\end{aligned} \tag{8.114}$$

for $g \in H^2(\mathbb{R}^3)$, $f \in H^1(\mathbb{R}^3)$ with the integrand to be understood in the sense of traces.

Finally, to obtain a formula in $n = 2$ dimensions we use the method of descent: That is we use the fact, that our solution in two dimensions is also a solution in three dimensions which happens to be independent of the third coordinate direction. Unfortunately this does not fit within our current framework since such functions are not square integrable (unless they vanish identically). However, we can fix this problem by choosing some large value $R > 0$ and a cutoff function $\varphi \in C_c^\infty(\mathbb{R})$ with $\varphi(r) = 1$ for $|r| \le R$. Then if we consider functions of the type $f(x) = f(x_1, x_2)\varphi(x_3)$, the Kirchhoff formula will not see the difference as long as $|x| + |t| \le R$. Hence we can simplify Kirchhoff's formula in the case f does not depend on x_3: Using spherical coordinates we obtain

$$\begin{aligned}
&\frac{t}{4\pi}\int_{S^2} f(x+t\omega)d\sigma^2(\omega) = \\
&\quad = \frac{t}{2\pi}\int_0^{\pi/2}\int_0^{2\pi} f(x_1 + t\sin(\theta)\cos(\varphi), x_2 + t\sin(\theta)\sin(\varphi))\sin(\theta)d\theta d\varphi \\
&\quad \overset{\rho=\sin(\theta)}{=} \frac{t}{2\pi}\int_0^1\int_0^{2\pi} \frac{f(x_1 + t\rho\cos(\varphi), x_2 + t\rho\sin(\varphi))}{\sqrt{1-\rho^2}}\rho d\rho\, d\varphi \\
&\quad = \frac{t}{2\pi}\int_{B_1(0)} \frac{f(x+ty)}{\sqrt{1-|y|^2}}d^2y
\end{aligned}$$

which gives **Poisson's formula**

$$u(t,x) = \frac{\partial}{\partial t}\frac{t}{2\pi}\int_{B_1(0)} \frac{g(x+ty)}{\sqrt{1-|y|^2}}d^2y + \frac{t}{2\pi}\int_{B_1(0)} \frac{h(x+ty)}{\sqrt{1-|y|^2}}d^2y. \tag{8.115}$$

Problem 8.26. *Assume $g \in L^2(\mathbb{R}^n)$. Show that $u(t)$ defined in (8.79) is in $C^1((0,\infty), L^2(\mathbb{R}^n))$ and solves (8.78). (Hint: $|\mathrm{e}^{-\varepsilon|p|^2} - 1| \le \varepsilon|p|^2$ for $\varepsilon \ge 0$.)*

Problem 8.27. *Suppose $u(t) \in C^1(I, X)$. Show that for $s, t \in I$*

$$\|u(t) - u(s)\| \le M|t-s|, \qquad M = \sup_{\tau\in[s,t]} \|\frac{du}{dt}(\tau)\|.$$

(Hint: Consider $d(\tau) = \|u(\tau) - u(s)\| - \tilde{M}(\tau - s)$ for $\tau \in [s,t]$. Suppose τ_0 is the largest τ for which the claim holds with $\tilde{M} > M$ and find a contradiction if $\tau_0 < t$.)

Problem 8.28. *Solve the **transport equation***

$$u_t + a\partial_x u = 0, \qquad u(0) = g,$$

using the Fourier transform.

Problem 8.29. *Suppose $A \in \mathscr{L}(X)$. Show that*

$$T(t) = \exp(tA) = \sum_{j=0}^\infty \frac{t^j}{j!}A^j$$

defines a C_0 (semi)group with generator A. Show that it is fact uniformly continuous: $T(t) \in C([0,\infty), \mathscr{L}(X))$.

Problem 8.30. *Let $r \ge 0$ and $k \in \mathbb{N}_0$. Show that $T_S(t) : H^r(\mathbb{R}^n) \to H^r(\mathbb{R}^n)$ as well as $T_S(t) : H^{k,k}(\mathbb{R}^n) \to H^{k,k}(\mathbb{R}^n)$ (cf. (8.72)) with*

$$\|T_S(t)f\|_{H^r} = \|T_S(t)f\|_{H^r}, \qquad \|T_S(t)f\|_{H^{k,k}} \le C\langle t\rangle^k \|f\|_{H^{k,k}}.$$

Problem* 8.31. *Show the* ***Dirichlet integral***

$$\lim_{R\to\infty} \int_0^R \frac{\sin(x)}{x} dx = \frac{\pi}{2}.$$

Show also that the sine integral is bounded

$$|\operatorname{Si}(x)| \le \min(x, \pi(1 + \frac{1}{2\mathrm{e}x})), \qquad x > 0.$$

(Hint: Write $\operatorname{Si}(R) = \int_0^R \int_0^\infty \sin(x) \mathrm{e}^{-xt} dt\, dx$ and use Fubini.)

8.6. Tempered distributions

In many situation, in particular when dealing with partial differential equations, it turns out convenient to look at generalized functions, also known as distributions.

To begin with we take a closer look at the Schwartz space $\mathcal{S}(\mathbb{R}^m)$, defined in (8.11), which already turned out to be a convenient class for the Fourier transform. For our purpose it will be crucial to have a notion of convergence in $\mathcal{S}(\mathbb{R}^m)$ and the natural choice is the topology generated by the seminorms

$$q_n(f) = \sum_{|\alpha|,|\beta| \le n} \|x^\alpha (\partial_\beta f)(x)\|_\infty, \tag{8.116}$$

where the sum runs over all multi indices $\alpha, \beta \in \mathbb{N}_0^m$ of order less than n. Unfortunately these norms cannot be replaced by a single norm (and hence we do not have a Banach space) but there is at least a metric

$$d(f,g) = \sum_{n=1}^\infty \frac{1}{2^n} \frac{q_n(f-g)}{1 + q_n(f-g)} \tag{8.117}$$

and $\mathcal{S}(\mathbb{R}^m)$ is complete with respect to this metric and hence a Fréchet space:

Lemma 8.40. *The Schwartz space $\mathcal{S}(\mathbb{R}^m)$ together with the family of seminorms $\{q_n\}_{n\in\mathbb{N}_0}$ is a Fréchet space.*

Proof. It suffices to show completeness. Since a Cauchy sequence f_k is in particular a Cauchy sequence with respect to q_n for arbitrary n, there is a limit $f \in C^\infty(\mathbb{R}^m)$ such that all derivatives converge uniformly. Moreover, since Cauchy sequences are bounded, $q_n(f_k) \le C_n$, we conclude $f \in \mathcal{S}(\mathbb{R}^m)$. Finally, $\lim_{k\to\infty} q_n(f_k, f) = 0$ for every n implies $\lim_{k\to\infty} d(f_k, f) = 0$. □

We refer to Section 6.4 from **[22]** for further background on Fréchet spaces. However, for our present purpose it is sufficient to observe that $f_n \to f$ if and only if $q_k(f_n - f) \to 0$ for every $k \in \mathbb{N}_0$. Moreover, (cf. Corollary 6.16 from **[22]**) a linear map $A : \mathcal{S}(\mathbb{R}^m) \to \mathcal{S}(\mathbb{R}^m)$ is continuous if and only if for every $j \in \mathbb{N}_0$ there is some $k \in \mathbb{N}_0$ and a corresponding constant C_k such that $q_j(Af) \le C_k q_k(f)$ and a linear functional $\ell : \mathcal{S}(\mathbb{R}^m) \to \mathbb{C}$ is continuous if and only if there is some $k \in \mathbb{N}_0$ and a corresponding constant C_k such that $|\ell(f)| \le C_k q_k(f)$.

Now the set of of all continuous linear functionals, that is the dual space $\mathcal{S}^*(\mathbb{R}^m)$, is known as the space of **tempered distributions**. As a natural topology on $\mathcal{S}^*(\mathbb{R}^m)$ one takes the weak-$*$ topology defined to be the weakest topology generated by the family of all point evaluations $q_f(T) = |T(f)|$ for all $f \in \mathcal{S}(\mathbb{R}^m)$. Since different tempered distributions must differ at least at one point the weak-$*$ topology is Hausdorff. In particular, a sequence of tempered distributions T_n converges to T precisely if $T_n(f) \to T(f)$ for all $f \in \mathcal{S}(\mathbb{R}^m)$.

To understand why this generalizes the concept of a function we begin by observing that any locally integrable function which does not grow too fast gives rise to a distribution.

Example 8.12. Let g be a locally integrable function of at most polynomial growth, that is, there is some $k \in \mathbb{N}_0$ such that $C_k := \int_{\mathbb{R}^m} |g(x)|(1+|x|)^{-k} d^m x < \infty$. Then

$$T_g(f) := \int_{\mathbb{R}^m} g(x) f(x) d^m x$$

is a distribution. To see that T_g is continuous observe that $|T_g(f)| \le C_k q_k(f)$. Moreover, note that by Lemma 3.23 the distribution T_g and the function g uniquely determine each other. $\diamond$

The next question is if there are distributions which are not of this form.

Example 8.13. Let $x_0 \in \mathbb{R}^m$ then

$$\delta_{x_0}(f) := f(x_0)$$

is a distribution, the **Dirac delta distribution** centered at x_0. Continuity follows from $|\delta_{x_0}(f)| \le q_0(f)$ Formally δ_{x_0} can be written as $T_{\delta_{x_0}}$ as in the previous example where δ_{x_0} is the Dirac δ-function which satisfies $\delta_{x_0}(x) = 0$ for $x \ne x_0$ and $\delta_{x_0}(x) = \infty$ such that $\int_{\mathbb{R}^m} \delta_{x_0}(x) f(x) d^m x = f(x_0)$. This is of course nonsense as one can easily see that δ_{x_0} cannot be expressed as T_g with a locally integrable function of at most polynomial growth (show this). However, giving a precise mathematical meaning to the Dirac δ-function was one of the main motivations to develop distribution theory. $\diamond$

Example 8.14. Let $\phi \in C_c^\infty(\mathbb{R}^n)$ be a mollifier and let $\phi_\varepsilon(x) = \varepsilon^{-n}\phi(\frac{x}{\varepsilon})$ be the associated approximate identity. Then $T_{\phi_\varepsilon(.-x_0)} \to \delta_{x_0}$ as $\varepsilon \downarrow 0$ since $T_{\phi_\varepsilon(.-x_0)}(f) = (\phi_\varepsilon * f)(x_0) \to f(x_0)$ for all $f \in \mathcal{S}(\mathbb{R}^n)$ by Lemma 3.21 (in the $p = \infty$ case). ⋄

Example 8.15. The example of the delta distribution can be easily generalized: Let μ be a Borel measure on $\mathbb{R}^m$ such that $C_k := \int_{\mathbb{R}^m}(1+|x|)^{-k}d\mu(x) < \infty$ for some k, then

$$T_\mu(f) := \int_{\mathbb{R}^m} f(x)d\mu(x)$$

is a distribution since $|T_\mu(f)| \le C_k q_k(f)$. ⋄

Example 8.16. Another interesting distribution in $\mathcal{S}^*(\mathbb{R})$ is given by

$$\big(p.v.\frac{1}{x}\big)(f) := \lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon}\frac{f(x)}{x}dx.$$

To see that this is a distribution note that by the mean value theorem

$$\begin{aligned}
|\big(p.v.\frac{1}{x}\big)(f)| &= \int_{\varepsilon<|x|<1}\frac{f(x)-f(0)}{x}dx + \int_{1<|x|}\frac{f(x)}{x}dx\\
&\le \int_{\varepsilon<|x|<1}\left|\frac{f(x)-f(0)}{x}\right|dx + \int_{1<|x|}\frac{|x\,f(x)|}{x^2}dx\\
&\le 2\sup_{|x|\le 1}|f'(x)| + 2\sup_{|x|\ge 1}|x\,f(x)|.
\end{aligned}$$

This shows $|\big(p.v.\frac{1}{x}\big)(f)| \le 2q_1(f)$. ⋄

Of course, to fill distribution theory with life, we need to extend the classical operations for functions to distributions. First of all, addition and multiplication by scalars comes for free, but we can easily do more. The general principle is always the same: For any continuous linear operator $A : \mathcal{S}(\mathbb{R}^m) \to \mathcal{S}(\mathbb{R}^m)$ there is a corresponding adjoint operator $A' : \mathcal{S}^*(\mathbb{R}^m) \to \mathcal{S}^*(\mathbb{R}^m)$, defined via $(A'T)(f) = T(Af)$, which extends the effect on functions (regarded as distributions of type T_g) to all distributions. We remark that A' is also continuous, but we will not use this here. Note however that sequential continuity is immediate: $(A'T_n)(f) = T_n(Af) \to T(Af) = (A'T)(f)$. We start with a simple example illustrating this procedure.

Let $h \in \mathcal{S}(\mathbb{R}^m)$, then the map $A : \mathcal{S}(\mathbb{R}^m) \to \mathcal{S}(\mathbb{R}^m)$, $f \mapsto h \cdot f$ is continuous. In fact, continuity follows from the **Leibniz rule**

$$\partial_\alpha(h\cdot f) = \sum_{\beta\le\alpha}\binom{\alpha}{\beta}(\partial_\beta h)(\partial_{\alpha-\beta}f),$$

where $\binom{\alpha}{\beta} = \frac{\alpha!}{\beta!(\alpha-\beta)!}$, $\alpha! = \prod_{j=1}^m(\alpha_j!)$, and $\beta \le \alpha$ means $\beta_j \le \alpha_j$ for $1 \le j \le m$. In particular, $q_j(h\cdot f) \le C_j q_j(h)q_j(f)$ which shows that A is continuous and hence the adjoint is well defined via

$$(A'T)(f) = T(Af). \tag{8.118}$$

Now what is the effect on functions? For a distribution T_g given by an integrable function as above we clearly have

$$\begin{aligned}(A'T_g)(f) = T_g(hf) &= \int_{\mathbb{R}^m} g(x)(h(x)f(x))d^m x\\ &= \int_{\mathbb{R}^m} (g(x)h(x))f(x)d^m x = T_{gh}(f).\end{aligned} \tag{8.119}$$

So the effect of A' on functions is multiplication by h and hence A' generalizes this operation to arbitrary distributions. We will write $A'T = h \cdot T$ for notational simplicity. Note that since f can even compensate a polynomial growth, h could even be a smooth functions all whose derivatives grow at most polynomially (e.g. a polynomial):

$$C^\infty_{pg}(\mathbb{R}^m) := \{h \in C^\infty(\mathbb{R}^m) | \forall \alpha \in \mathbb{N}_0^m \, \exists C, n : \, |\partial_\alpha h(x)| \le C(1+|x|)^n\}. \tag{8.120}$$

In summary we can define

$$(h \cdot T)(f) := T(h \cdot f), \qquad h \in C^\infty_{pg}(\mathbb{R}^m). \tag{8.121}$$

Example 8.17. Let h be as above and $\delta_{x_0}(f) = f(x_0)$. Then

$$h \cdot \delta_{x_0}(f) = \delta_{x_0}(h \cdot f) = h(x_0)f(x_0)$$

and hence $h \cdot \delta_{x_0} = h(x_0)\delta_{x_0}$. ⋄

Moreover, since Schwartz functions have derivatives of all orders, the same is true for tempered distributions! To this end let α be a multi-index and consider $D_\alpha : \mathcal{S}(\mathbb{R}^m) \to \mathcal{S}(\mathbb{R}^m)$, $f \mapsto (-1)^{|\alpha|}\partial_\alpha f$ (the reason for the extra $(-1)^{|\alpha|}$ will become clear in a moment) which is continuous since $q_n(D_\alpha f) \le q_{n+|\alpha|}(f)$. Again we let D'_α be the corresponding adjoint operator and compute its effect on distributions given by functions g:

$$\begin{aligned}(D'_\alpha T_g)(f) = T_g((-1)^{|\alpha|}\partial_\alpha f) &= (-1)^{|\alpha|} \int_{\mathbb{R}^m} g(x)(\partial_\alpha f(x))d^m x\\ &= \int_{\mathbb{R}^m} (\partial_\alpha g(x))f(x)d^m x = T_{\partial_\alpha g}(f),\end{aligned} \tag{8.122}$$

where we have used integration by parts in the last step which is (e.g.) permissible for $g \in C^{|\alpha|}_{pg}(\mathbb{R}^m)$ with $C^k_{pg}(\mathbb{R}^m) = \{h \in C^k(\mathbb{R}^m) | \forall |\alpha| \le k \, \exists C, n : |\partial_\alpha h(x)| \le C(1+|x|)^n\}$.

Hence for every multi-index α we define

$$(\partial_\alpha T)(f) := (-1)^{|\alpha|} T(\partial_\alpha f). \tag{8.123}$$

Example 8.18. Let $\Theta(x) = 0$ for $x < 0$ and $\Theta(x) = 1$ for $x \geq 0$ be the Heaviside step function. Then

$$(\partial T_\Theta)(f) = -\int_{\mathbb{R}} \Theta(x) f'(x) dx = -\int_0^\infty f'(x) dx = f(0) = \delta_0(f).$$

Slightly more general, if μ is a finite Borel measure, then using $f(x) = -\int_x^\infty f'(y)dy$ and Fubini shows

$$\begin{aligned}\int_{\mathbb{R}} f(x) d\mu(x) &= -\int_{\mathbb{R}} \int_{\mathbb{R}} \chi_{[x,\infty)}(y) f'(y) dy\, d\mu(x)\\ &= -\int_{\mathbb{R}} \int_{\mathbb{R}} \chi_{(-\infty,y]}(x) f'(y) d\mu(x) dy = -\int_{\mathbb{R}} \mu(x) f'(y) dy.\end{aligned}$$

Hence μ can be obtained as the derivative of its distribution function. It is straightforward to generalize this to higher dimensions. ⋄

Example 8.19. Let α be a multi-index and $\delta_{x_0}(f) = f(x_0)$. Then

$$\partial_\alpha \delta_{x_0}(f) = (-1)^{|\alpha|} \delta_{x_0}(\partial_\alpha f) = (-1)^{|\alpha|} (\partial_\alpha f)(x_0).$$ ⋄

Finally we use the same approach for the Fourier transform $\mathcal{F} : \mathcal{S}(\mathbb{R}^m) \to \mathcal{S}(\mathbb{R}^m)$, which is also continuous since $q_n(\hat{f}) \leq C_n q_n(f)$ by Lemma 8.4. Since Fubini implies

$$\int_{\mathbb{R}^m} g(x) \hat{f}(x) d^m x = \int_{\mathbb{R}^m} \hat{g}(x) f(x) d^m x \tag{8.124}$$

for $g \in L^1(\mathbb{R}^m)$ (or $g \in L^2(\mathbb{R}^m)$) and $f \in \mathcal{S}(\mathbb{R}^m)$ we define the Fourier transform of a distribution to be

$$(\mathcal{F}T)(f) \equiv \hat{T}(f) := T(\hat{f}) \tag{8.125}$$

such that $\mathcal{F}T_g = T_{\hat{g}}$ for $g \in L^1(\mathbb{R}^m)$ (or $g \in L^2(\mathbb{R}^m)$).

Example 8.20. Let us compute the Fourier transform of $\delta_{x_0}(f) = f(x_0)$:

$$\hat{\delta}_{x_0}(f) = \delta_{x_0}(\hat{f}) = \hat{f}(x_0) = \frac{1}{(2\pi)^{m/2}} \int_{\mathbb{R}^m} \mathrm{e}^{-\mathrm{i}x_0 x} f(x) d^m x = T_g(f),$$

where $g(x) = (2\pi)^{-m/2} \mathrm{e}^{-\mathrm{i}x_0 x}$. ⋄

Example 8.21. A slightly more involved example is the Fourier transform of $p.v.\frac{1}{x}$:

$$\begin{aligned}\big((p.v.\frac{1}{x})\big)^\wedge(f) &= \lim_{\varepsilon\downarrow 0} \int_{\varepsilon<|x|} \frac{\hat{f}(x)}{x} dx = \lim_{\varepsilon\downarrow 0} \frac{1}{\sqrt{2\pi}} \int_{\varepsilon<|x|<1/\varepsilon} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}yx} \frac{f(y)}{x} dy\, dx\\ &= \lim_{\varepsilon\downarrow 0} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \int_{\varepsilon<|x|<1/\varepsilon} \frac{\mathrm{e}^{-\mathrm{i}yx}}{x} dx\, f(y) dy\\ &= -\mathrm{i}\sqrt{\frac{2}{\pi}} \lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}} \operatorname{sign}(y) \left(\int_\varepsilon^{1/\varepsilon} \frac{\sin(t)}{t} dt \right) f(y) dy.\end{aligned}$$

Moreover, Problem 8.31 shows that we can use dominated convergence to get

$$((p.v.\frac{1}{x}))^\wedge(f) = -\mathrm{i}\sqrt{\frac{\pi}{2}}\int_{\mathbb{R}} \operatorname{sign}(y) f(y) dy,$$

that is, $((p.v.\frac{1}{x}))^\wedge = -\mathrm{i}\sqrt{\frac{\pi}{2}}\operatorname{sign}(y)$. ◇

Note that since $\mathcal{F} : \mathcal{S}(\mathbb{R}^m) \to \mathcal{S}(\mathbb{R}^m)$ is a homeomorphism, so is its adjoint $\mathcal{F}' : \mathcal{S}^*(\mathbb{R}^m) \to \mathcal{S}^*(\mathbb{R}^m)$. In particular, its inverse is given by

$$\check{T}(f) := T(\check{f}). \tag{8.126}$$

Moreover, all the operations for $\mathcal{F}$ carry over to $\mathcal{F}'$. For example, from Lemma 8.4 we immediately obtain

$$(\partial_\alpha T)^\wedge = (\mathrm{i}p)^\alpha \hat{T}, \qquad (x^\alpha T)^\wedge = \mathrm{i}^{|\alpha|}\partial_\alpha \hat{T}. \tag{8.127}$$

Similarly one can extend Lemma 8.2 to distributions.

Next we turn to convolutions. Since (Fubini)

$$\int_{\mathbb{R}^m} (h*g)(x) f(x) d^m x = \int_{\mathbb{R}^m} g(x)(\tilde{h}*f)(x) d^m x, \qquad \tilde{h}(x) = h(-x), \tag{8.128}$$

for integrable functions f, g, h we define

$$(h * T)(f) := T(\tilde{h} * f), \qquad h \in \mathcal{S}(\mathbb{R}^m), \tag{8.129}$$

which is well defined by Corollary 8.14. Moreover, Corollary 8.14 immediately implies

$$(h*T)^\wedge = (2\pi)^{n/2}\hat{h}\hat{T}, \qquad (hT)^\wedge = (2\pi)^{-n/2}\hat{h}*\hat{T}, \qquad h \in \mathcal{S}(\mathbb{R}^m). \tag{8.130}$$

Example 8.22. Note that the Dirac delta distribution acts like an identity for convolutions since

$$(h * \delta_0)(f) = \delta_0(\tilde{h} * f) = (\tilde{h} * f)(0) = \int_{\mathbb{R}^m} h(y) f(y) d^m y = T_h(f). \qquad \diamond$$

In the last example the convolution is associated with a function. This turns out always to be the case.

Theorem 8.41. *For every $T \in \mathcal{S}^*(\mathbb{R}^m)$ and $h \in \mathcal{S}(\mathbb{R}^m)$ we have that $h * T$ is associated with the function*

$$h * T = T_g, \qquad g(x) := T(h(x - .)) \in C^\infty_{pg}(\mathbb{R}^m). \tag{8.131}$$

Proof. By definition $(h * T)(f) = T(\tilde{h} * f)$ and since $(\tilde{h} * f)(x) = \int h(y - x) f(y) d^m y$ the distribution T acts on $h(y - .)$ and we should be able to pull out the integral by linearity. To make this idea work let us replace the integral by a Riemann sum

$$(\tilde{h} * f)(x) = \lim_{n\to\infty} \sum_{j=1}^{n^{2m}} h(y_j^n - x) f(y_j^n) |Q_j^n|,$$

where Q_j^n is a partition of $[-\frac{n}{2}, \frac{n}{2}]^m$ into n^{2m} cubes of side length $\frac{1}{n}$ and y_j^n is the midpoint of Q_j^n. Then, if this Riemann sum converges to $h * f$ in $\mathcal{S}(\mathbb{R}^m)$, we have

$$(h * T)(f) = \lim_{n\to\infty} \sum_{j=1}^{n^{2m}} g(y_j^n) f(y_j^n) |Q_j^n|$$

and of course we expect this last limit to converge to the corresponding integral. To be able to see this we need some properties of g. Since

$$|h(z-x) - h(z-y)| \le q_1(h)|x-y|$$

by the mean value theorem and similarly

$$q_n(h(z-x) - h(z-y)) \le C_n q_{n+1}(h)|x-y|$$

we see that $x \mapsto h(.-x)$ is continuous in $\mathcal{S}(\mathbb{R}^m)$. Consequently g is continuous. Similarly, if $x = x_0 + \varepsilon e_j$ with e_j the unit vector in the j'th coordinate direction,

$$q_n\left(\frac{1}{\varepsilon}(h(.-x) - h(.-x_0)) - \partial_j h(.-x_0)\right) \le C_n q_{n+2}(h)\varepsilon$$

which shows $\partial_j g(x) = T((\partial_j h)(x-.))$. Applying this formula iteratively gives

$$\partial_\alpha g(x) = T((\partial_\alpha h)(x-.)) \tag{8.132}$$

and hence $g \in C^\infty(\mathbb{R}^m)$. Furthermore, g has at most polynomial growth since $|T(f)| \le C q_n(f)$ implies

$$|g(x)| = |T(h(.-x))| \le C q_n(h(.-x)) \le \tilde{C}(1+|x|^n) q(h).$$

Combining this estimate with (8.132) even gives $g \in C^\infty_{pg}(\mathbb{R}^m)$.

In particular, since $g \cdot f \in \mathcal{S}(\mathbb{R}^m)$ the corresponding Riemann sum converges and we have $h * T = T_g$.

It remains to show that our first Riemann sum for the convolution converges in $\mathcal{S}(\mathbb{R}^m)$. It suffices to show

$$\sup_x |x|^N \left| \sum_{j=1}^{n^{2m}} h(y_j^n - x) f(y_j^n)|Q_j^n| - \int_{\mathbb{R}^m} h(y-x) f(y) d^m y \right| \to 0$$

since derivatives are automatically covered by replacing h with the corresponding derivative. The above expressions splits into two terms. The first one is

$$\sup_x |x|^N \left| \int_{|y|>n/2} h(y-x) f(y) d^m y \right| \le C q_N(h) \int_{|y|>n/2} (1+|y|^N)|f(y)| d^m y \to 0.$$

The second one is

$$\sup_x |x|^N \left| \sum_{j=1}^{n^{2m}} \int_{Q_j^n} \big(h(y_j^n - x)f(y_j^n) - h(y-x)f(y)\big) d^m y \right|$$

and the integrand can be estimated by

$$\begin{aligned}
&|x|^N \big|h(y_j^n - x)f(y_j^n) - h(y-x)f(y)\big| \\
&\quad \le |x|^N \big|h(y_j^n - x) - h(y-x)\big| |f(y_j^n)| + |x|^N |h(y-x)| \big|f(y_j^n) - f(y)\big| \\
&\quad \le \big(q_{N+1}(h)(1 + |y_j^n|^N)|f(y_j^n)| + q_N(h)(1+|y|^N)\big|\partial f(\tilde{y}_j^n)\big|\big)\, |y - y_j^n|
\end{aligned}$$

and the claim follows since $|f(y)| + |\partial f(y)| \le C(1+|y|)^{-N-m-1}$. □

Example 8.23. If we take $T = \frac{1}{\pi}\big(p.v.\frac{1}{x}\big)$, then

$$(T * h)(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_{|y|>\varepsilon} \frac{h(x-y)}{y} dy = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_{|x-y|>\varepsilon} \frac{h(y)}{x-y} dy$$

which is known as the **Hilbert transform** of h. Moreover,

$$(T * h)^\wedge(p) = \frac{-\mathrm{i}}{\sqrt{2\pi}} \operatorname{sign}(p) \hat{h}(p)$$

as distributions and hence the Hilbert transform extends to a bounded operator on $L^2(\mathbb{R})$. ⋄

As a consequence we get that distributions can be approximated by functions.

Theorem 8.42. *Let ϕ_ε be the standard mollifier. For every $T \in \mathcal{S}^*(\mathbb{R}^m)$ we have $\phi_\varepsilon * T \to T$ in $\mathcal{S}^*(\mathbb{R}^m)$.*

Proof. We need to show $\phi_\varepsilon * T(f) = T(\phi_\varepsilon * f)$ for any $f \in \mathcal{S}(\mathbb{R}^m)$. This follows from continuity since $\phi_\varepsilon * f \to f$ in $\mathcal{S}(\mathbb{R}^m)$ as can be easily seen (the derivatives follow from Lemma 3.20 (ii)). □

Note that Lemma 3.20 (i) and (ii) implies

$$\partial_\alpha (h * T) = (\partial_\alpha h) * T = h * (\partial_\alpha T). \tag{8.133}$$

When working with distributions it is also important to observe that, in contradistinction to smooth functions, they can be supported at a single point. Here the **support** $\operatorname{supp}(T)$ of a distribution is the smallest closed set V (namely the intersection of all closed sets with this property) for which

$$\operatorname{supp}(f) \subseteq \mathbb{R}^m \setminus V \Longrightarrow T(f) = 0.$$

An example of a distribution supported at 0 is the Dirac delta distribution δ_0 as well as all of its derivatives. It turns out that these are in fact the only examples.

Lemma 8.43. *Suppose T is a distribution supported at x_0. Then*

$$T = \sum_{|\alpha|\le n} c_\alpha \partial_\alpha \delta_{x_0}. \tag{8.134}$$

Proof. For simplicity of notation we suppose $x_0 = 0$. First of all there is some n such that $|T(f)| \le C q_n(f)$. Write

$$T = \sum_{|\alpha|\le n} c_\alpha \partial_\alpha \delta_0 + \tilde{T},$$

where $c_\alpha = \frac{T(x^\alpha)}{\alpha!}$. Then $\tilde{T}$ vanishes on every polynomial of degree at most n, has support at 0, and still satisfies $|\tilde{T}(f)| \le \tilde{C} q_n(f)$. Now let $\phi_\varepsilon(x) = \phi(\frac{x}{\varepsilon})$, where ϕ has support in $B_1(0)$ and equals 1 in a neighborhood of 0. Then $\tilde{T}(f) = \tilde{T}(g) = \tilde{T}(\phi_\varepsilon g)$, where $g(x) = f(x) - \sum_{|\alpha|\le n} \frac{f^{(\alpha)}(0)}{\alpha!} x^\alpha$. Since $|\partial_\beta g(x)| \le C_\beta \varepsilon^{n+1-|\beta|}$ for $x \in B_\varepsilon(0)$ Leibniz' rule implies $q_n(\phi_\varepsilon g) \le C\varepsilon$. Hence $|\tilde{T}(f)| = |\tilde{T}(\phi_\varepsilon g)| \le \tilde{C} q_n(\phi_\varepsilon g) \le \hat{C}\varepsilon$ and since $\varepsilon > 0$ is arbitrary we have $|\tilde{T}(f)| = 0$, that is, $\tilde{T} = 0$. □

Example 8.24. Let us try to solve the Poisson equation in the sense of distributions. We begin with solving

$$-\Delta T = \delta_0.$$

Taking the Fourier transform we obtain

$$|p|^2 \hat{T} = (2\pi)^{-m/2}$$

and since $|p|^{-2}$ is a locally integrable function in $\mathbb{R}^m$ for $m \ge 3$ the above equation will be solved by

$$\Phi := (2\pi)^{-m/2} \left(\frac{1}{|p|^2}\right)^\vee.$$

Explicitly, Φ must be determined from

$$\Phi(f) = (2\pi)^{-m/2} \int_{\mathbb{R}^m} \frac{\check{f}(p)}{|p|^2} d^m p = (2\pi)^{-m/2} \int_{\mathbb{R}^m} \frac{\hat{f}(p)}{|p|^2} d^m p$$

and evaluating Lemma 8.27 at $x = 0$ we obtain for $m \ge 3$ that $\Phi = (2\pi)^{-m/2} T_{I_2}$ where I_2 is the Riesz potential. (Alternatively one can also compute the weak derivative of the Riesz potential directly, see Problems 7.11 and 7.12.) Note, that Φ is not unique since we could add a harmonic polynomial corresponding to a solution of the homogenous equation — Problem 8.35.

Given $h \in \mathcal{S}(\mathbb{R}^m)$ we can now consider $h * \Phi$ which solves

$$-\Delta(h * \Phi) = h * (-\Delta\Phi) = h * \delta_0 = h,$$

where in the last equality we have identified h with T_h. Note that since $h * \Phi$ is associated with a function in $C^\infty_{pg}(\mathbb{R}^m)$ our distributional solution is also a classical solution. This gives the formal calculations with the Dirac delta function found in many physics textbooks a solid mathematical meaning. $\diamond$

Note that while we have been quite successful in generalizing many basic operations to distributions, our approach is limited to linear operations! In particular, it is not possible to define nonlinear operations, for example the product of two distributions, within this framework. In fact, there is no associative product of two distributions extending the product of a distribution by a function from above.

Example 8.25. Consider the distributions δ_0, x, and $p.v.\frac{1}{x}$ in $\mathcal{S}^*(\mathbb{R})$. Then

$$x \cdot \delta_0 = 0, \qquad x \cdot \big(p.v.\frac{1}{x}\big) = 1.$$

Hence if there would be an associative product of distributions we would get $0 = (x \cdot \delta_0) \cdot p.v.\frac{1}{x} = \delta_0 \cdot (x \cdot p.v.\frac{1}{x}) = \delta_0$. $\diamond$

This is known as Schwartz' impossibility result. However, if one is content with preserving the product of functions, Colombeau algebras will do the trick.

Problem 8.32. *Compute the derivative of $g(x) = \operatorname{sign}(x)$ in $\mathcal{S}^*(\mathbb{R})$.*

Problem 8.33. *Let $h \in C^\infty_{pg}(\mathbb{R}^m)$ and $T \in \mathcal{S}^*(\mathbb{R}^m)$. Show*

$$\partial_\alpha(h \cdot T) = \sum_{\beta \le \alpha} \binom{\alpha}{\beta} (\partial_\beta h)(\partial_{\alpha-\beta} T).$$

Problem 8.34. *Show that $\operatorname{supp}(T_g) = \operatorname{supp}(g)$ for locally integrable functions.*

Problem 8.35. *Find all harmonic distributions $T \in \mathcal{S}^*(\mathbb{R})$, that is, which satisfy $\Delta T = 0$. (Hint: Take the Fourier transform and use Lemma 8.43.)*

Chapter 9

Interpolation and some applications

9.1. Interpolation and the Fourier transform on L^p

We will fix some measure space (X, μ) and abbreviate $L^p := L^p(X, d\mu)$ for notational simplicity. 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 the **Lyapunov inequality**

$$\|f\|_p \le \|f\|_{p_0}^{1-\theta} \|f\|_{p_1}^{\theta}, \tag{9.1}$$

where $\frac{1}{p} := \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$, $\theta \in (0,1)$ (Problem 3.13). Note that $L^{p_0} \cap L^{p_1}$ contains all integrable simple functions which are a convenient dense set of functions in L^p for $1 \le p < \infty$ (for $p = \infty$ this is only true if the measure is finite — cf. Problem 3.23).

This is a first occurrence of an interpolation technique. Next we want to turn to operators. For example, 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 the question is, if this can be used to extend the Fourier transform to the spaces in between.

Since it will be convenient to have a space which contains both L^{p_0} and L^{p_1} as subspaces, we denote by $L^{p_0} + L^{p_1}$ the space of (equivalence classes) of measurable functions f which can be written as a sum $f = f_0 + f_1$ with $f_0 \in L^{p_0}$ and $f_1 \in L^{p_1}$ (clearly such a decomposition is not unique and different decompositions will differ by elements from $L^{p_0} \cap L^{p_1}$). Then we have

$$L^p \subseteq L^{p_0} + L^{p_1}, \qquad p_0 \le p \le p_1, \tag{9.2}$$

since we can always decompose a function $f \in L^p$, $1 \le p < \infty$, as $f = f\chi_{\{x|\,|f(x)|\le 1\}} + f\chi_{\{x|\,|f(x)|>1\}}$ with $f\chi_{\{x|\,|f(x)|\le 1\}} \in L^p \cap L^\infty$ and $f\chi_{\{x|\,|f(x)|>1\}} \in$

$L^1 \cap L^p$. Hence, if we have two operators $A_0 : L^{p_0} \to L^{q_0}$ and $A_1 : L^{p_1} \to L^{q_1}$ which coincide on the intersection, $A_0|_{L^{p_0}\cap L^{p_1}} = A_1|_{L^{p_0}\cap L^{p_1}}$, we can extend them by virtue of

$$A : L^{p_0} + L^{p_1} \to L^{q_0} + L^{q_1}, \qquad f_0 + f_1 \mapsto A_0 f_0 + A_1 f_1 \tag{9.3}$$

(check that A is indeed well-defined, i.e., independent of the decomposition of f into $f_0 + f_1$). In particular, this defines A on L^p for every $p \in (p_0, p_1)$ and the question is if A restricted to L^p will be a bounded operator into some L^q provided A_0 and A_1 are bounded.

To answer this question we begin with a result from complex analysis.

Theorem 9.1 (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*

$$|F(z)| \le \begin{cases} M_0, & \mathrm{Re}(z) = 0, \\ M_1, & \mathrm{Re}(z) = 1, \end{cases} \tag{9.4}$$

then

$$|F(z)| \le M_0^{1-\mathrm{Re}(z)} M_1^{\mathrm{Re}(z)} \tag{9.5}$$

for every $z \in \overline{S}$.

Proof. Without loss of generality we can assume $M_0, M_1 > 0$ (otherwise the estimate holds for any positive constants and we can take limits) 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 $|\mathrm{Im}(z)| \ge \sqrt{\log(M) n}$ (assuming $M > 1$ w.l.o.g.). Moreover, applying the maximum modulus principle on the rectangle $\{z \in S | |\mathrm{Im}(z)| < \sqrt{\log(M) n}$ we see $|F_n(z)| \le 1$ on this rectangle and hence for all $z \in \overline{S}$. Finally, letting $n \to \infty$ the claim follows. □

Now we are able to show the **Riesz–Thorin interpolation theorem**[1]:

Theorem 9.2 (Riesz–Thorin). *Let $(X, d\mu)$ and $(Y, d\nu)$ be σ-finite measure spaces and $1 \le p_0, p_1, q_0, q_1 \le \infty$. If A is a linear operator on*

$$A : L^{p_0}(X, d\mu) + L^{p_1}(X, d\mu) \to L^{q_0}(Y, d\nu) + L^{q_1}(Y, d\nu) \tag{9.6}$$

satisfying

$$\|Af\|_{q_0} \le M_0 \|f\|_{p_0}, \qquad \|Af\|_{q_1} \le M_1 \|f\|_{p_1}, \tag{9.7}$$

[1]Marcel Riesz (1886 –1969), Hungarian mathematician

[1]Olof Thorin (1912–2004), Swedish mathematician

then A has continuous restrictions

$$A_\theta : L^{p_\theta}(X, d\mu) \to L^{q_\theta}(Y, d\nu), \quad \frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \ \frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1} \quad (9.8)$$

satisfying $\|A_\theta\| \le M_0^{1-\theta} M_1^\theta$ for every $\theta \in [0,1]$.

Proof. By Lemma 3.6 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 $\frac{1}{q_\theta} + \frac{1}{q'_\theta} = 1$.

To this end let $f(x) = \sum_j \alpha_j \chi_{A_j}(x)$, $g(x) = \sum_k \beta_k \chi_{B_k}(x)$ be simple functions with $\|f\|_{p_\theta} = \|g\|_{q'_\theta} = 1$ and set $f_z(x) := \sum_j |\alpha_j|^{p_\theta/p_z} \operatorname{sign}(\alpha_j) \chi_{A_j}(x)$, $g_z(y) := \sum_k |\beta_k|^{q'_\theta/q'_z} \operatorname{sign}(\beta_k) \chi_{B_k}(y)$ such that $\|f_z\|_{p_x} = \|g_z\|_{q'_x} = 1$ for $x := \mathrm{Re}(z) \in [0,1]$. In the case $p_\theta = \infty$ the quotient p_θ/p_z has to replaced by 1 and similarly if $q'_\theta = \infty$.

Next, note that

$$F(z) := \int (Af_z) g_z d\nu$$

is entire (being a linear combination of exponential functions) with

$$|F(z)| \le \sum_{jk} \int |A\chi_{A_j}| \chi_{B_k} d\nu |\alpha_j|^{p_\theta/p_x} |\beta_k|^{q'_\theta/q'_x}.$$

In particular F is bounded for $0 \le x := \mathrm{Re}(z) \le 1$. Moreover, for $x = 0$ we have by Hölders inequality $|F(z)| \le M_0 \|f_z\|_{p_0} \|g_z\|_{q'_0} = M_0$ and for $x = 1$ we have similarly $|F(z)| \le M_1$. Hence, as required, $|F(z)| \le M_0^{1-x} M_1^x$ by the tree lines theorem. □

Note that the proof shows even a bit more

Corollary 9.3. *Let A be an operator defined on the space of integrable simple functions satisfying* (9.7)*. Then A has continuous extensions A_θ as in the Riesz–Thorin theorem which will agree on $L^{p_0}(X, d\mu) \cap L^{p_1}(X, d\mu)$. If $p_1 = \infty$, then A_1 is only defined on the closure of the space of integrable simple functions in L^∞.*

Also observe that it is important to work with complex spaces. If A is an operator on real Lebesgue spaces, we can of course extend it to the complex spaces (by setting $A(f + \mathrm{i}g) = Af + \mathrm{i}Ag$) but this will in general increase the norm by a factor of 2 and this factor hence also needs to be added to (9.5).

As an application we get two important inequalities:

Corollary 9.4 (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*

$$(2\pi)^{-n/(2q)} \|\hat{f}\|_q \le (2\pi)^{-n/(2p)} \|f\|_p. \tag{9.9}$$

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 9.3). Moreover, its range is dense for $1 < p \le 2$ but not all of $L^q(\mathbb{R}^n)$ unless $p = q = 2$.

Corollary 9.5 (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*

$$\|f * g\|_r \le \|f\|_p \|g\|_q, \tag{9.10}$$

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 Lemma 3.20. 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 $A_{g,q} : L^p \to L^r$, 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. □

Combining the last two corollaries we obtain:

Corollary 9.6. *Let $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$ with $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1 \ge 0$ and $1 \le r, p, q \le 2$. Then*

$$(f * g)^\wedge = (2\pi)^{n/2} \hat{f} \hat{g}.$$

Proof. By Corollary 8.14 the claim holds for $f, g \in \mathcal{S}(\mathbb{R}^n)$. Now take a sequence of Schwartz functions $f_m \to f$ in L^p and a sequence of Schwartz functions $g_m \to g$ in L^q. Then the left-hand side converges in $L^{r'}$, where $\frac{1}{r'} = 2 - \frac{1}{p} - \frac{1}{q}$, by the Young and Hausdorff–Young inequalities. Similarly, the right-hand side converges in $L^{r'}$ by the generalized Hölder (Problem 3.10) and Hausdorff–Young inequalities. □

Problem 9.1. *Show that*

$$\|f\| := \inf\{\|f_0\|_{p_0} + \|f_1\|_{p_1}\} | f = f_0 + f_1,\ f_0 \in L^{p_0},\ f_1 \in L^{p_1}\}$$

turns $L^{p_0} + L^{p_1}$ into a Banach space.

Problem 9.2. *Use dilations $f(x) \mapsto f(\lambda x)$, $\lambda > 0$, to show that an inequality $\|\hat{f}\|_q \le C_{p,q}\|f\|_p$ can only hold if $\frac{1}{p} + \frac{1}{q} = 1$.*

Problem* 9.3. *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 $1 \le p < 2$. (Hint for the case $n = 1$: Consider $\phi_z(x) = \exp(-zx^2/2)$ for $z = \lambda + \mathrm{i}\omega$ with $\lambda > 0$.)*

Problem 9.4 (Young inequality). *Let $K(x, y)$ be measurable and suppose*

$$\sup_x \|K(x,.)\|_{L^r(Y,d\nu)} \le C, \qquad \sup_y \|K(.,y)\|_{L^r(X,d\mu)} \le C.$$

where $\frac{1}{r} = \frac{1}{q} - \frac{1}{p} \ge 1$ for some $1 \le p \le q \le \infty$. Then the operator $K : L^p(Y, d\nu) \to L^q(X, d\mu)$, defined by

$$(Kf)(x) = \int_Y K(x,y)f(y)d\nu(y),$$

for μ-almost every x, is bounded with $\|K\| \le C$. (Hint: Show $\|Kf\|_\infty \le C\|f\|_{r'}$, $\|Kf\|_r \le C\|f\|_1$ and use interpolation.)

9.2. The Marcinkiewicz interpolation theorem

In this section we are going to look at another interpolation theorem which might be helpful in situations where the Riesz–Thorin interpolation theorem does not apply. In this respect recall, that $f(x) := \frac{1}{x}$ just fails to be integrable over $\mathbb{R}$. To include such functions we begin by slightly weakening the L^p norms. To this end we fix some measure space (X, μ) and consider the **distribution function**

$$E_f(r) := \mu\big(\{x \in X \,|\, |f(x)| > r\}\big) \tag{9.11}$$

of a measurable function $f : X \to \mathbb{C}$ with respect to μ. Note that E_f is decreasing and right continuous. Moreover, $E_f(r) = 0$ for $r \ge \|f\|_\infty$.

Example 9.1. Considering the characteristic function χ_A of a measurable set A. Then $E_{\chi_A}(r) = |A|$ for $0 \le r < 1$ and $E_{\chi_A}(r) = 0$ for $r \ge 1$.

In general you can think of the area under the graph of f as some mass points which are only allowed to move horizontally like beads on an abacus. Then you obtain E_f by rotating the abacus by 90 degrees and let gravity do the rest. Note that if you rotate the abacus back, the result is known as decreasing rearrangement of f (Problem 9.10). ⋄

Given, the distribution function we can compute the L^p norm via (Problem 2.20)

$$\|f\|_p^p = p\int_0^\infty r^{p-1}E_f(r)dr, \qquad 1 \le p < \infty. \tag{9.12}$$

In the case $p = \infty$ we have

$$\|f\|_\infty = \inf\{r \geq 0 | E_f(r) = 0\}. \tag{9.13}$$

Another relationship follows from the observation

$$\|f\|_p^p = \int_X |f|^p d\mu \geq \int_{|f|>r} r^p d\mu = r^p E_f(r) \tag{9.14}$$

which yields **Markov's inequality**

$$E_f(r) \leq r^{-p} \|f\|_p^p. \tag{9.15}$$

Motivated by this we define the **weak L_p norm**

$$\begin{aligned} \|f\|_{p,w} &:= \inf\{C > 0 | E_f(r) \leq r^{-p} C^p,\ r > 0\} \\ &= \sup_{r>0} r E_f(r)^{1/p}, \qquad 1 \leq p < \infty, \end{aligned} \tag{9.16}$$

and the corresponding spaces $L^{p,w}(X, d\mu)$ consist of all equivalence classes of functions which are equal a.e. for which $\|.\|_{p,w}$ is finite. Clearly the distribution function and hence the weak L^p norm depend only on the equivalence class. Despite its name, the weak L_p *norm* turns out to be only a quasinorm (Problem 9.5). By construction we have

$$\|f\|_{p,w} \leq \|f\|_p \tag{9.17}$$

and thus $L^p(X, d\mu) \subseteq L^{p,w}(X, d\mu)$. In the case $p = \infty$ we set $\|.\|_{\infty,w} := \|.\|_\infty$.

Example 9.2. Consider $f(x) = \frac{1}{x}$ in $\mathbb{R}$. Then clearly $f \notin L^1(\mathbb{R})$ but

$$E_f(r) = |\{x|\, |\frac{1}{x}| > r\}| = |\{x|\, |x| < r^{-1}\}| = \frac{2}{r}$$

shows that $f \in L^{1,w}(\mathbb{R})$ with $\|f\|_{1,w} = 2$. Slightly more general, the function $f(x) = |x|^{-n/p} \notin L^p(\mathbb{R}^n)$ but $f \in L^{p,w}(\mathbb{R}^n)$. Hence $L^{p,w}(\mathbb{R}^n)$ is strictly larger than $L^p(\mathbb{R}^n)$. ◇

Now we are ready for our interpolation result. We call an operator $T : L^p(X, d\mu) \to L^q(X, d\nu)$ **subadditive** if it satisfies

$$|T(f + g)| \leq |T(f)| + |T(g)|. \tag{9.18}$$

It is said to be of **strong type** (p, q) if

$$\|T(f)\|_q \leq C_{p,q} \|f\|_p \tag{9.19}$$

and of **weak type** (p, q) if

$$\|T(f)\|_{q,w} \leq C_{p,q,w} \|f\|_p. \tag{9.20}$$

By (9.17) strong type (p, q) is indeed stronger than weak type (p, q) and we have $C_{p,q,w} \leq C_{p,q}$.

Theorem 9.7 (Marcinkiewicz[2]). *Let $(X, d\mu)$ be a measure space and $1 \le p_0 < p_1 \le \infty$. Let T be a subadditive operator defined for all $f \in L^p(X, d\mu)$, $p \in [p_0, p_1]$. If T is of weak type (p_0, p_0) and (p_1, p_1) then it is also of strong type (p, p) for every $p_0 < p < p_1$ with*

$$\|T(f)\|_p \le 2\left(\frac{p}{p-p_0} + \frac{p}{p_1-p}\right)^{\frac{1}{p}} C_{p_0}^{\frac{p_0}{p}\frac{p_1-p}{p_1-p_0}} C_{p_1}^{\frac{p_1}{p}\frac{p-p_0}{p_1-p_0}} \|f\|_p. \tag{9.21}$$

Proof. We begin by assuming $p_1 < \infty$. Fix $f \in L^p$ as well as some number $s > 0$ and decompose $f = f_0 + f_1$ according to

$$f_0 := f\chi_{\{x|\,|f|>s\}} \in L^{p_0} \cap L^p, \qquad f_1 := f\chi_{\{x|\,|f|\le s\}} \in L^p \cap L^{p_1}.$$

Next we use (9.12),

$$\|T(f)\|_p^p = p\int_0^\infty r^{p-1} E_{T(f)}(r)dr = p2^p \int_0^\infty r^{p-1} E_{T(f)}(2r)dr$$

and observe

$$E_{T(f)}(2r) \le E_{T(f_0)}(r) + E_{T(f_1)}(r)$$

since $|T(f)| \le |T(f_0)| + |T(f_1)|$ implies $|T(f)| > 2r$ only if $|T(f_0)| > r$ or $|T(f_1)| > r$. Now using (9.15) our assumption implies

$$E_{T(f_j)}(r) \le \left(\frac{\|T(f_j)\|_{p_j,w}}{r}\right)^{p_j} \le \left(\frac{C_j\|f_j\|_{p_j}}{r}\right)^{p_j}, \quad j = 0, 1,$$

and choosing $s = \rho r$ (with $\rho > 0$ to be chosen later) we obtain

$$E_{T(f)}(2r) \le \frac{C_0^{p_0}}{r^{p_0}} \int_{\{x|\,|f|>\rho r\}} |f|^{p_0} d\mu + \frac{C_1^{p_1}}{r^{p_1}} \int_{\{x|\,|f|\le\rho r\}} |f|^{p_1} d\mu.$$

In summary we have $\|T(f)\|_p^p \le p2^p(C_0^{p_0} I_0 + C_1^{p_1} I_1)$ with

$$\begin{aligned} I_0 &= \int_0^\infty \int_X r^{p-p_0-1}\chi_{\{(x,r)|\,|f(x)|>\rho r\}} |f(x)|^{p_0} d\mu(x)\, dr \\ &= \int_X |f(x)|^{p_0} \int_0^{|f(x)|/\rho} r^{p-p_0-1} dr\, d\mu(x) = \frac{\rho^{p_0-p}}{p-p_0}\|f\|_p^p \end{aligned}$$

and

$$\begin{aligned} I_1 &= \int_0^\infty \int_X r^{p-p_1-1}\chi_{\{(x,r)|\,|f(x)|\le\rho r\}} |f(x)|^{p_1} d\mu(x)\, dr \\ &= \int_X |f(x)|^{p_1} \int_{|f(x)|/\rho}^\infty r^{p-p_1-1} dr\, d\mu(x) = \frac{\rho^{p_1-p}}{p_1-p}\|f\|_p^p. \end{aligned}$$

In summary we obtain

$$\|T(f)\|_p \le 2\big(p/(p-p_0)C_0^{p_0}\rho^{p_0-p} + p/(p_1-p)C_1^{p_1}\rho^{p_1-p}\big)^{1/p}\|f\|_p$$

---

[2] Józef Marcinkiewicz (1910–1940), Polish mathematician

and choosing ρ such that

$$C_0^{p_0}\rho^{p_0-p} = C_1^{p_1}\rho^{p_1-p}$$

produces the desired estimate.

The case $p_1 = \infty$ is similar: Split $f \in L^p$ according to

$$f_0 := f\chi_{\{x|\,|f|>s/C_1\}} \in L^{p_0} \cap L^p, \qquad f_1 := f\chi_{\{x|\,|f|\le s/C_1\}} \in L^p \cap L^\infty$$

(if $C_1 = 0$ there is nothing to prove). Then $\|T(f_1)\|_\infty \le s/C_1$ and hence $E_{T(f_1)}(s/C_1) = 0$. Thus

$$E_{T(f)}(2r) \le \frac{C_0^{p_0}}{r^{p_0}} \int_{\{x|\,|f|>r/C_1\}} |f|^{p_0} d\mu$$

and we can proceed as before to obtain

$$\|T(f)\|_p \le 2\big(p/(p-p_0)\big)^{1/p} C_0^{p_0/p} C_1^{1-p_0/p} \|f\|_p,$$

which is again the desired estimate. □

As with the Riesz–Thorin theorem there is also a version for operators which are of weak type (p_0, q_0) and (p_1, q_1) but the proof is slightly more involved and the above diagonal version is frequently sufficient.

As a first application we will use it to investigate the **Hardy–Littlewood maximal function** defined for any locally integrable function in $\mathbb{R}^n$ via

$$\mathcal{M}(f)(x) = \sup_{r>0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |f(y)| d^n y. \tag{9.22}$$

By the dominated convergence theorem, the integral is continuous with respect to x and consequently (Problem 1.23) $\mathcal{M}(f)$ is lower semicontinuous (and hence measurable). Moreover, its value is unchanged if we change f on sets of measure zero, so $\mathcal{M}$ is well defined for functions in $L^p(\mathbb{R}^n)$. However, it is unclear if $\mathcal{M}(f)(x)$ is finite a.e. at this point. If f is bounded we of course have the trivial estimate

$$\|\mathcal{M}(f)\|_\infty \le \|f\|_\infty. \tag{9.23}$$

Example 9.3. The maximal function of $f(x) := \chi_{(-1,1)}(x)$ is given by (Problem 9.12)

$$\mathcal{M}(f)(x) = \begin{cases} 1, & |x| < 1, \\ \frac{1}{1+|x|}, & |x| \ge 1. \end{cases}$$

Note that even though f has compact support, $\mathcal{M}(f)$ only decays like $O(1/x)$ and hence ist not integrable. So $\mathcal{M}$ is not of strong type $(1,1)$. The slow decay could be fixed by taking the sup not for $r > 0$ but (e.g.) for $0 < r < 1$. However, this will still not render $\mathcal{M}$ of strong type $(1,1)$ (Problem 9.13). ⋄

Theorem 9.8 (Hardy–Littlewood maximal inequality)**.** *The maximal function is of weak type* $(1,1)$*,*

$$E_{\mathcal{M}(f)}(r) \le \frac{3^n}{r}\|f\|_1, \tag{9.24}$$

and of strong type (p,p)*,*

$$\|\mathcal{M}(f)\|_p \le 2\left(\frac{3^n p}{p-1}\right)^{1/p}\|f\|_p, \tag{9.25}$$

for every $1 < p \le \infty$*.*

Proof. The first estimate follows literally as in the proof of Lemma 4.5 and combining this estimate with the trivial one (9.23) the Marcinkiewicz interpolation theorem yields the second. □

Using this fact, our next aim is to prove the Hardy–Littlewood–Sobolev inequality. As a preparation we show

Lemma 9.9. *Let* $\phi \in L^1(\mathbb{R}^n)$ *be a radial,* $\phi(x) = \phi_0(|x|)$ *with* ϕ_0 *nonnegative and nonincreasing. Then we have the following estimate for convolutions with integrable functions:*

$$|(\phi * f)(x)| \le \|\phi\|_1 \mathcal{M}(f)(x). \tag{9.26}$$

Proof. We start with the case where $\phi_0 = \sum_{j=1}^p \alpha_j \chi_{[0,r_j]}$ with $\alpha_j > 0$ is a simple function. Then

$$(\phi * f)(x) = \sum_j \alpha_j |B_{r_j}(0)| \frac{1}{|B_{r_j}(x)|}\int_{B_{r_j}(x)} f(y) d^n y$$

and the estimate follows upon taking absolute values and observing $\|\phi\|_1 = \sum_j \alpha_j |B_{r_j}(0)|$.

To see the general case choose a sequence of simple functions $\phi_0^n \nearrow \phi_0$ and observe

$$\begin{aligned}|(\phi * f)(x)| &\le (\phi * |f|)(x) = \lim_n (\phi^n * |f|)(x) \le \lim_n \|\phi^n\|_1 \mathcal{M}(f)(x)\\ &= \|\phi\|_1 \mathcal{M}(f)(x)\end{aligned}$$

by monotone convergence. □

Now we will apply this to the Riesz potential (8.45) of order α:

$$\mathcal{I}_\alpha f = I_\alpha * f. \tag{9.27}$$

Theorem 9.10 (Hardy–Littlewood–Sobolev inequality)**.** *Let* $0 < \alpha < n$*,* $p \in (1, \frac{n}{\alpha})$*, and* $q = \frac{pn}{n - p\alpha} \in (\frac{n}{n-\alpha}, \infty)$ *(i.e.,* $\frac{\alpha}{n} = \frac{1}{p} - \frac{1}{q}$*). Then* $\mathcal{I}_\alpha$ *is of strong type* (p,q)*,*

$$\|\mathcal{I}_\alpha f\|_q \le C_{p,\alpha,n}\|f\|_p. \tag{9.28}$$

Proof. We split the Riesz potential into two parts

$$I_\alpha = I_\alpha^0 + I_\alpha^\infty, \qquad I_\alpha^0 = I_\alpha \chi_{(0,\varepsilon)},\ I_\alpha^\infty = I_\alpha \chi_{[\varepsilon,\infty)},$$

where $\varepsilon > 0$ will be determined later. Note that $I_\alpha^0(|.|) \in L^1(\mathbb{R}^n)$ and $I_\alpha^\infty(|.|) \in L^r(\mathbb{R}^n)$ for every $r \in (\frac{n}{n-\alpha}, \infty)$. In particular, since $p' = \frac{p}{p-1} \in (\frac{n}{n-\alpha}, \infty)$, both integrals converge absolutely by the Young inequality (9.10). Next we will estimate both parts individually. Using Lemma 9.9 we obtain

$$|\mathcal{I}_\alpha^0 f(x)| \le \int_{|y|<\varepsilon} \frac{d^n y}{|y|^{n-\alpha}} \mathcal{M}(f)(x) = \frac{(n-1)V_n}{\alpha - 1} \varepsilon^n \mathcal{M}(f)(x).$$

On the other hand, using Hölder's inequality we infer

$$|\mathcal{I}_\alpha^\infty f(x)| \le \left(\int_{|y|\ge\varepsilon} \frac{d^n y}{|y|^{(n-\alpha)p'}} \right)^{1/p'} \|f\|_p = \left(\frac{(n-1)V_n}{p'(n-\alpha)-n} \right)^{1/p'} \varepsilon^{\alpha - n/p} \|f\|_p.$$

Now we choose $\varepsilon = \left(\frac{\|f\|_p}{\mathcal{M}(f)(x)}\right)^{p/n}$ such that

$$|\mathcal{I}_\alpha f(x)| \le \tilde{C} \|f\|_p^\theta \mathcal{M}(f)(x)^{1-\theta}, \qquad \theta = \frac{\alpha p}{n} \in (\frac{\alpha}{n}, 1),$$

where $\tilde{C}/2$ is the larger of the two constants in the estimates for $\mathcal{I}_\alpha^0 f$ and $\mathcal{I}_\alpha^\infty f$. Taking the L^q norm in the above expression gives

$$\|\mathcal{I}_\alpha f\|_q \le \tilde{C} \|f\|_p^\theta \|\mathcal{M}(f)^{1-\theta}\|_q = \tilde{C} \|f\|_p^\theta \|\mathcal{M}(f)\|_{q(1-\theta)}^{1-\theta} = \tilde{C} \|f\|_p^\theta \|\mathcal{M}(f)\|_p^{1-\theta}$$

and the claim follows from the Hardy–Littlewood maximal inequality. □

Problem* 9.5. *Show that $E_f = 0$ if and only if $f = 0$. Moreover, show $E_{f+g}(r+s) \le E_f(r) + E_g(s)$ and $E_{\alpha f}(r) = E_f(r/|\alpha|)$ for $\alpha \neq 0$. Conclude that $L^{p,w}(X, d\mu)$ is a quasinormed space with*

$$\|f+g\|_{p,w} \le 2\big(\|f\|_{p,w} + \|g\|_{p,w}\big), \qquad \|\alpha f\|_{p,w} = |\alpha| \|f\|_{p,w}.$$

Problem 9.6. *Show $f(x) = |x|^{-n/p} \in L^{p,w}(\mathbb{R}^n)$. Compute $\|f\|_{p,w}$.*

Problem 9.7. *Let A be a set with finite measure and $0 < q < p < \infty$. Show that*

$$\int_A |f|^q d\mu \le \frac{p}{p-q} \mu(A)^{1-\frac{q}{p}} \|f\|_{p,w}^q$$

and conclude that

$$L^p(X, d\mu) \subseteq L^{p,w}(X, d\mu) \subseteq L^q(X, d\mu).$$

provided μ is finite. (Hint: $\mu(\{x \in A |\, |f(x)| > r\}) \le \min(\mu(A), r^{-p}\|f\|_{p,w}^p)$.)

Problem 9.8. *Let $0 < q < p < \infty$ and consider*

$$|||f|||_{p,w} := \sup_{0<\mu(A)<\infty} \mu(A)^{\frac{1}{p}-\frac{1}{q}} \left(\int_A |f|^q d\mu \right)^{1/q}.$$

Show that

$$\|f\|_{p,w} \le |||f|||_{p,w} \le \left(\frac{p}{q-p}\right)^{1/q} \|f\|_{p,w}.$$

Show that for $p > 1$ the choice $q = 1$ gives a proper norm. (Hint: For one direction use the previous problem. For the other direction make a special choice for A.)

Problem 9.9. *Show that if $f_n \to f$ in weak-L^p, then $f_n \to f$ in measure.*

Problem 9.10. *The right continuous generalized inverse of the distribution function is known as the* ***decreasing rearrangement*** *of f:*

$$f^\star(t) := \inf\{r \ge 0 | E_f(r) \le t\}$$

(see Section 2.5 for basic properties of the generalized inverse). Note that $f^\star$ is decreasing with $f^\star(0) = \|f\|_\infty$. Show that

$$E_{f^\star} = E_f$$

and in particular $\|f^\star\|_p = \|f\|_p$.

Problem 9.11. *Show that $f = 0$ if and only if $\mathcal{M}(f)(x_0) = 0$ for one x_0.*

Problem 9.12. *Compute the maximal function of $f(x) := \chi_{(-1,1)}(x)$.*

Problem 9.13. *Show that the maximal function of an integrable function might not be (locally) integrable. Specifically, show that*

$$f(x) := \frac{\chi_{[-1/2,1/2]}(x)}{|x|\log(|x|)^2}$$

is in $L^1(\mathbb{R})$ and satisfies

$$\mathcal{M}(f)(x) \ge \frac{1}{2|x|\log(1/|2x|)}, \qquad x \in [-1/4, 1/4],$$

which is not locally integrable.

Problem 9.14. *Show that the maximal function of an integrable function satisfies*

$$|f(x)| \le \mathcal{M}(f)(x) < \infty$$

at every Lebesgue point.

Problem* 9.15. *Let ϕ be a nonnegative nonincreasing radial function with $\|\phi\|_1 = 1$. Set $\phi_\varepsilon(x) = \varepsilon^{-n}\phi(\frac{x}{\varepsilon})$. Show that for integrable f we have $(\phi_\varepsilon * f)(x) \to f(x)$ at every Lebesgue point. (Hint: Split $\phi = \phi^\delta + \tilde{\phi}^\delta$ into a part with compact support ϕ^δ and a rest by setting $\tilde{\phi}^\delta(x) = \min(\delta, \phi(x))$. To handle the compact part use Problem 3.31. To control the contribution of the rest use Lemma 9.9.)*

Problem 9.16. *For $f \in L^1(0,1)$ define*

$$T(f)(x) = \mathrm{e}^{\mathrm{i}\arg\left(\int_0^1 f(y)dy\right)} f(x).$$

Show that T is subadditive and norm preserving. Show that T is not continuous in L^1.

9.3. Calderón–Zygmund operators

In this section we want to look at convolution-type integral operators

$$(Kf)(x) := \int_{\mathbb{R}^n} K(x-y)f(y)d^n y. \tag{9.29}$$

If $K \in L^1(\mathbb{R}^n)$, then K will be a bounded map on $L^p(\mathbb{R}^n)$ whose norm can be estimated using Young's inequality (3.25). Here we want to look at the case where this assumption is not satisfied. The prototypical example being the Hilbert transform with $K(x) := \frac{1}{x}$ in $\mathbb{R}$.

To this end we will call a measurable function $K : \mathbb{R}^n \to \mathbb{C}$ a **Calderón–Zygmund kernel**[3] provided

(i) $|K(x)| \le \frac{B}{|x|^n}$, for all $x \in \mathbb{R}^n$,

(ii) $\int_{|x|>2|y|} |K(x) - K(x-y)| d^n x \le B$ for all $y \in \mathbb{R}^n$ (**Hörmander condition**[4]),

(iii) $\int_{r<|x|<R} K(x) d^n x = 0$ for all $0 < r < R$ (**cancellation condition**).

Such operators are also known as **singular integral operators** since K is not necessarily integrable.

As with the Hilbert transform, special care has to be taken when defining an associated operator and we will set

$$(Kf)(x) := \lim_{\varepsilon \downarrow 0} \int_{\varepsilon<|x-y|} K(x-y)f(y)d^n y \tag{9.30}$$

for $f \in \mathcal{S}(\mathbb{R}^n)$. To see that this limit exists we first split the integral and then use the cancellation condition to obtain

$$\begin{aligned}(Kf)(x) &= \lim_{\varepsilon \downarrow 0} \int_{\varepsilon<|x-y|<1} K(x-y)f(y)d^n y + \int_{1\le|x-y|} K(x-y)f(y)d^n y \\ &= \int_{|x-y|<1} K(x-y)(f(y)-f(x))d^n y + \int_{1\le|x-y|} K(x-y)f(y)d^n y.\end{aligned} \tag{9.31}$$

[3] Alberto Calderón (1920–1998), Argentinian mathematician

[3] Antoni Zygmund (1900–1992), Polish mathematician

[4] Lars Hörmander (1931–2012), Swedish mathematician

Here the first integrand can be estimated according to $|K(x-y)(f(y)-f(x))| \le \frac{B}{|x-y|^n}C|x-y|$ and hence is integrable and the limit follows from dominated convergence. Of course a weaker condition like $f \in C^{0,\alpha}(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ for some $1 \le p < \infty$ would also be sufficient for this argument.

Example 9.4. The **Hilbert transform** $K(x) := \frac{1}{x}$ clearly satisfies these conditions. $\diamond$

Note that the Hörmander condition is a smoothness condition and for differentiable kernels there is the following variant which is easier to check.

Lemma 9.11. *Suppose*

$$|\nabla K(x)| \le \frac{B}{|x|^{n+1}}, \tag{9.32}$$

then the Hörmander condition holds with B replaced by $2^n S_n B$.

Proof. We begin by observing

$$\begin{aligned}|K(x)-K(x-y)| &= \left|\int_0^1 \nabla K(x-ty)\cdot y\, dt\right| \le \int_0^1 \frac{B}{|x-ty|^{n+1}}|y|dt\\ &\le \int_0^1 \frac{B}{|x-ty|^{n+1}}|y|dt \le \frac{2^{n+1}B|y|}{|x|^{n+1}},\end{aligned}$$

where that last estimate holds for $|x| > 2|y|$ since in this case $|x-ty| \ge |x|-|y| \ge \frac{1}{2}|x|$. Consequently

$$\int_{|x|>2|y|} |K(x)-K(x-y)|d^n x \le 2^{n+1}B|y| \int_{|x|>2|y|} \frac{1}{|x|^{n+1}} d^n x = 2^n B S_n$$

as required. $\square$

Example 9.5. Let $n \ge 2$. The **Riesz transform** $K_j(x) := \frac{x_j}{|x|^{n+1}}$ as well as the double Riesz transform $K_{jk}(x) := \frac{x_j x_k}{|x|^{n+2}}$ for $j \ne k$ and $K_{jj}(x) := \frac{x_j^2 - n^{-1}|x|^2}{|x|^{n+2}}$ satisfy these conditions. Note that the double Riesz kernel is (up to a constant) the second derivative of the fundamental solution of the Laplace equation (8.47).

Indeed (i) and (ii) are straightforward (via Lemma 9.11). Finally, (iii) is evident for K_j and K_{jk} if $k \ne j$ since we integrate an odd function over a symmetric domain. In the case $j = k$ note that (by symmetry) the integral over $x_j^2/|x|^{n+2}$ is independent of j. Hence we also have (iii) in this case. $\diamond$

Applying the Fourier transform, a Calderón–Zygmund operator is given by a multiplication operator which provides insight into it properties on $L^2(\mathbb{R}^n)$.

Example 9.6. By Example 8.23 the multiplier associated with the Cauchy kernel is $m(p) := -\mathrm{i}(2\pi)^{-1/2}\,\mathrm{sign}(p)$. Hence the Hilbert transform is unitary when multiplied by $(2\pi)^{1/2}$. ◇

Example 9.7. Consider the fundamental solution of the Laplace equation (8.47). Then we know that its Fourier multiplier is $|p|^{-2}$. Consequently the Fourier multiplier of the double Riesz transform is (up to constants) $m_{jk}(p) := n^{-1}\delta_{jk} - p_j p_k |p|^{-2}$. Since m_{jk} is bounded, the double Riesz transform extends to a bounded operator on $L^2(\mathbb{R}^n)$. ◇

In general, it will not be possible to compute the associated multiplier explicitly, but we can still show that it is always bounded:

Lemma 9.12. *Let K be a Calderón–Zygmund kernel. Then K gives rise to a bounded operator on $L^2(\mathbb{R}^n)$.*

Proof. The idea is that K is given by multiplication with $(2\pi)^{n/2}\hat{K}$ when taking the Fourier transform and hence the statement is equivalent to the fact that $\hat{K}$ is bounded. To this end consider the (up to constants) Fourier transform of the truncated kernel

$$m_{r,s}(p) := \int_{r<|x|<s} K(x)\mathrm{e}^{-\mathrm{i}px} d^n x.$$

Then we have

$$K(f)(x) = \lim_{r\to 0, s\to\infty} \int_{r<|x|<s} K(x) f(y-x) d^n x$$

pointwise for $f \in \mathcal{S}(\mathbb{R}^n)$ and hence by Fatou

$$\|Kf\|_2 \le \liminf_{r\to 0, s\to\infty} \|K_{r,s} f\|_2 \le \liminf_{r\to 0, s\to\infty} \|m_{r,s}\|_\infty \|f\|_2.$$

Consequently it suffices to bound $m_{r,s}$.

We split the integral into the regions $|x| < 2\pi|p|^{-1}$ and $|x| > 2\pi|p|^{-1}$. For the first part we obtain by virtue of conditions (i) and (iii)

$$\begin{aligned} \left|\int_{r<|x|<2\pi|p|^{-1}} K(x)\mathrm{e}^{-\mathrm{i}px} d^n x\right| &= \left|\int_{r<|x|<2\pi|p|^{-1}} K(x)(\mathrm{e}^{-\mathrm{i}px}-1) d^n x\right| \\ &\le B \int_{r<|x|<2\pi|p|^{-1}} \frac{|x||p|}{|x|^n} d^n x = 2\pi S_n B\Big(1 - \frac{|x||p|}{2\pi}\Big) \le 2\pi S_n B. \end{aligned}$$

For the second part we set $q := \pi\frac{p}{|p|^2}$ (such that $\mathrm{e}^{-\mathrm{i}pq} = -1$) and use

$$\begin{aligned} \int_{2\pi|p|^{-1}<|x|<s} K(x)\mathrm{e}^{-\mathrm{i}px} d^n x &= -\int_{2\pi|p|^{-1}<|x|<s} K(x)\mathrm{e}^{-\mathrm{i}p(x+q)} d^n x \\ &= -\int_{2\pi|p|^{-1}<|x-q|<s} K(x-q)\mathrm{e}^{-\mathrm{i}px} d^n x. \end{aligned}$$

Now we can add these two expressions in order to use (ii). The only nuisance being that the domain of integration slightly differs. Fortunately, this will not affect the required bound. To simplify notation, let $A := \{x \in \mathbb{R}^n | 2|q| < |x| < s\}$ and $A_q := A - q$ (note $\pi|p|^{-1} = |q|$). Then

$$\int_{2\pi|p|^{-1}<|x|<s} K(x)\mathrm{e}^{-\mathrm{i}px} d^n x = \frac{1}{2}\int_{A_q} (K(x) - K(x-q))\mathrm{e}^{-\mathrm{i}px} d^n x \\ + \frac{1}{2}\int_{A\setminus A_q} K(x)\mathrm{e}^{-\mathrm{i}px} d^n x - \frac{1}{2}\int_{A_q\setminus A} K(x)\mathrm{e}^{-\mathrm{i}px} d^n x$$

and we need to estimate these three integrals. The first one follows from (ii)

$$\left|\int_{A_q} (K(x) - K(x-q))\mathrm{e}^{-\mathrm{i}px} d^n x\right| \le \int_A |K(x+q) - K(x)| d^n x \le B.$$

Concerning the second one, note that $A \setminus A_q \subset \{x | 2|q| < |x| < 3|q|\} \cup \{x | s - |q| < |x| < s\}$ and hence

$$\left|\int_{A\setminus A_q} K(x)\mathrm{e}^{-\mathrm{i}px} d^n x\right| \le \int_{2|q|<|x|<2|q|} |K(x)| d^n x + \int_{s-|q|<|x|<s} |K(x)| d^n x \\ = S_n B \left(\int_{2|q|}^{3|q|} \frac{dr}{r} + \int_{s-|q|}^{s} \frac{dr}{r}\right) = S_n B\big(\log(\frac{3}{2}) + \log(\frac{s}{s-|q|})\big).$$

Similarly, $A \setminus A_q \subset \{x || q| < |x| < 2|q|\} \cup \{x | s < |x| < s + |q|\}$ and

$$\left|\int_{A_q\setminus A} K(x)\mathrm{e}^{-\mathrm{i}px} d^n x\right| \le S_n B\big(\log(\frac{2}{1}) + \log(\frac{s+|q|}{s})\big).$$

In summary this shows

$$\left|\int_{r<|x|<s} K(x)\mathrm{e}^{-\mathrm{i}px} d^n x\right| \le B\big(2\pi S_n + 1 + S_n \log(3) + \log(\frac{s|p|+\pi}{s|p|-\pi})\big)$$

for $\frac{2\pi}{s} < |p| < \frac{2\pi}{r}$ and we conclude

$$\liminf_{r\to 0, s\to\infty} \|m_{r,s}\|_\infty \le B(1 + (2\pi + \log(3))S_n). \qquad \square$$

Our next aim is to provided a weak-L^1 bound for Calderón–Zygmund operators such that we can apply the Marcinkiewicz interpolation theorem. To this end we will need a decomposition of an integrable function $f \in L^1(\mathbb{R}^n)$ into a *good* and a *bad* part according to the Hardy–Littlewood maximal function. Recall that the Hardy–Littlewood maximal function is based on the quantity

$$\frac{1}{|Q|}\int_Q |f(x)| d^n x, \tag{9.33}$$

where $Q \subset \mathbb{R}^n$ are balls. However, here it will be more convenient to use cubes instead since they can be chosen to tile the domain nicely. To this end we introduce the **dyadic cubes**

$$Q_{k,m} := [2^k m_1, 2^k(m_1+1)) \times \cdots \times [2^k m_n, 2^k(m_n+1)), \quad k \in \mathbb{Z},\ m \in \mathbb{Z}^n. \tag{9.34}$$

Note that two dyadic cubes are either disjoint or one is contained within the other. Moreover, due to our integrability assumption

$$\frac{1}{|Q|}\int_Q |f(x)| d^n x \le \frac{\|f\|_1}{|Q|} \tag{9.35}$$

and our averages will decay as the side length of the cubes increases. On the other hand, if we decrease the side lengths, then they will either stay bounded (if e.g. f is bounded on the cube) or increase for a sequence of cubes containing singularities of f. Hence we will choose a threshold and divide the domain into good and bad cubes depending on whether our above quantity remains below the chosen threshold no matter how small we make the subdivision.

Lemma 9.13 (Calderón–Zygmund)**.** *Fix $\lambda > 0$. Any function $f \in L^1(\mathbb{R}^n)$ can be decomposed according to $f = g + b$, such that:*

(i) $\|g\|_\infty \le \lambda$ *and* $\|g\|_1 + \|b\|_1 = \|f\|_1$.

(ii) $b = \sum_j \chi_{Q_j} f$, *where the sum runs over disjoint dyadic cubes such that for each cube one has*

$$\lambda < \frac{1}{|Q_j|}\int_{Q_j} |f(x)| d^n x \le 2^n \lambda.$$

(iii)

$$\sum_j |Q_j| < \frac{\|f\|_1}{\lambda}.$$

Proof. We call a dyadic cube *bad* if

$$\lambda < \frac{1}{|Q_{k,m}|}\int_{Q_{k,m}} |f(x)| d^n x$$

and if it is not already contained in a larger bad cube. The collection of bad cubes is obtained by observing that there are no bad cubes $Q_{k_0,m}$ provided $\|f\|_1 \le 2^{k_0 n}\lambda$. Hence one can start at such a k_0 and proceed by increasing k and removing the bad cubes at each level while further subdividing the others. This gives a collection of disjoint bad cubes Q_j and a corresponding decomposition $f = g + b$.

Note that since every bad cube $Q_j = Q_{k,m}$ is contained in some parent cube from the previous step, $Q_{k,m} \subset Q_{k-1,m'}$, we have

$$\frac{1}{|Q_{k,m}|}\int_{Q_{k,m}} |f(x)| d^n x \le \frac{1}{|Q_{k,m}|}\int_{Q_{k-1,m'}} |f(x)| d^n x \le 2^n \lambda$$

and (ii) holds. Moreover,

$$\sum_j |Q_j| < \frac{1}{\lambda}\sum_j \int_{Q_j} |f(x)| d^n x \le \frac{\|f\|_1}{\lambda}$$

and consequently (iii) also holds. Finally, if $x_0 \not\in \bigcup_j Q_j$, then there is a sequence $\tilde{Q}_k := Q_{k,m_k}$ of dyadic cubes containing x_0 such that

$$\frac{1}{|\tilde{Q}_k|}\int_{\tilde{Q}_k} |f(x)| d^n x \le \lambda.$$

Consequently $|f(x_0)| \le \lambda$ by Lebesgue's differentiation theorem (Lemma 4.7). Since $g(x_0) = f(x_0)$ and since g vanishes on $\bigcup_j Q_j$, item (i) also holds. □

This is known as Calderón–Zygmund decomposition of f at height λ. Note that by interpolation (9.1) we have $g \in L^p(\mathbb{R}^n)$ for every $1 \le p \le \infty$.

Subtracting the average over each bad cube from b and adding them to g we can assume that b has average zero over each bad cube.

Corollary 9.14. *Fix $\lambda > 0$. Any function $f \in L^1(\mathbb{R}^n)$ can be decomposed according to $f = g + b$, such that:*

(i) $\|g\|_\infty \le 2^n\lambda$ *and* $\|g\|_1 \le \|f\|_1$, $\|b\|_1 \le 2\|f\|_1$.

(ii) $b = \sum_j \chi_{Q_j}(f - \frac{1}{|Q_j|}\int_{Q_j} f\, d^n x)$, *where the sum runs over disjoint dyadic cubes such that for each cube one has*

$$\lambda < \frac{1}{|Q_j|}\int_{Q_j} |f(x)| d^n x \le 2^n \lambda.$$

(iii)

$$\sum_j |Q_j| < \frac{\|f\|_1}{\lambda}.$$

Now are ready to show the anticipated weak-L^1 estimate. Our formulation emphasises that we only require the Hörmander condition (ii) together with boundedness on L^2. This is convenient in case the latter condition can be obtained by other means than Lemma 9.15. This will come handy in the proof of Theorem 9.18 below.

Lemma 9.15. *Let K be a kernel satisfying the Hörmander condition (ii). Moreover, suppose that there is an bounded operator K on $L^2(\mathbb{R}^n)$ such that*

$$(Kf)(x) = \int_{\mathbb{R}^n} K(x-y)f(y)d^n y \tag{9.36}$$

for all $f \in L^2_c(\mathbb{R}^n)$ and all $x \notin \operatorname{supp}(f)$. Then for every $f \in L^2(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$ there is the weak-L^1 bound

$$|\{x \in \mathbb{R}^n | |(Kf)(x)| \ge \lambda\}| \le \frac{C}{\lambda}\|f\|_1, \qquad \lambda > 0. \tag{9.37}$$

Proof. Let $f \in L^2(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$ and write $f = g + b$ according to Corollary 9.14 (note that $b \in L^2 \cap L^1$). Next (cf. Problem 9.5)

$$\begin{aligned}
&|\{x \in \mathbb{R}^n | |(Kf)(x)| \ge \lambda\}| \\
&\quad \le |\{x \in \mathbb{R}^n | |(Kg)(x)| \ge \frac{\lambda}{2}\}| + |\{x \in \mathbb{R}^n | |(Kb)(x)| \ge \frac{\lambda}{2}\}| \\
&\quad \le \frac{C}{\lambda^2}\|g\|_2^2 + |\{x \in \mathbb{R}^n | |(Kb)(x)| \ge \frac{\lambda}{2}\}|,
\end{aligned}$$

where we have used our assumption that K is bounded on L^2. Moreover, since $\|g\|_2^2 \le \|g\|_1 \|g\|_\infty$ (Lyapunov inequality (9.1)) we obtain

$$|\{x \in \mathbb{R}^n | |(Kf)(x)| \ge \lambda\}| \le \frac{2^n C}{\lambda}\|f\|_1 + |\{x \in \mathbb{R}^n | |(Kb)(x)| \ge \frac{\lambda}{2}\}|$$

and it remains to control the second summand. In order to be able to apply (ii) we will work with slightly expanded cubes. To this end denote by Q^* the cube with the same center but side length increased by $2\sqrt{n}$. Then

$$\begin{aligned}
|\{x \in \mathbb{R}^n | |(Kb)(x)| \ge \frac{\lambda}{2}\}| &\le |\cup_j Q_j^*| + |\{x \in \mathbb{R}^n \setminus \cup_j Q_j^* | |(Kb)(x)| \ge \frac{\lambda}{2}\}| \\
&\le (2\sqrt{n})^n \frac{\|f\|_1}{\lambda} + \frac{2}{\lambda}\int_{\mathbb{R}^n \setminus \cup_j Q_j^*} |(Kb)(x)| d^n x.
\end{aligned}$$

Next we use $b = \sum_k b_k$, where $b_k := b\chi_{Q_k}$, to obtain

$$\begin{aligned}
\frac{2}{\lambda}\int_{\mathbb{R}^n \setminus \cup Q_j^*} |(Kb)(x)| d^n x &= \frac{2}{\lambda}\sum_k \int_{\mathbb{R}^n \setminus \cup_j Q_j^*} |(Kb_k)(x)| d^n x \\
&\le \frac{2}{\lambda}\sum_k \int_{\mathbb{R}^n \setminus Q_k^*} |(Kb_k)(x)| d^n x.
\end{aligned}$$

Since b_k has mean zero, we have

$$(Kb_k)(x) = \int_{Q_k} (K(x-y) - K(x-y_k)) b_k(y) d^n y,$$

where y_k is the center of Q_k, and we can now exploit (ii) (and Fubini) to obtain

$$\int_{\mathbb{R}^n\setminus Q_k^*} |(Kb_k)(x)|d^n x \le \int_{\mathbb{R}^n\setminus Q_k^*}\int_{Q_k} |K(x-y)-K(x-y_k)||b_k(y)|d^n y\, d^n x$$
$$\le B\int_{Q_k} |b_k(y)|d^n y.$$

In summary, we have

$$\frac{2}{\lambda}\int_{\mathbb{R}^n\setminus\cup Q_j^*} |(Kb)(x)|d^n x \le \frac{2B}{\lambda}\|b\|_1 \le \frac{4B}{\lambda}\|f\|_1. \qquad \square$$

Given these preparations we are now able to prove the main result:

Theorem 9.16 (Calderón–Zygmund)**.** *Let K be a Calderón–Zygmund kernel. Then K gives rise to a bounded operator on $L^p(\mathbb{R}^n)$ for all $1 < p < \infty$.*

Proof. Combining Lemma 9.12 and Lemma 9.15 with the Marcinkiewicz interpolation theorem establishes the claim for $1 < p \le 2$. To cover the remaining case we apply a duality argument: To this end we identify the dual space of $L^p(\mathbb{R}^n)$ with $L^{p'}(\mathbb{R}^n)$ as usual (Theorem 6.1). Then the adjoint is defined via

$$\int_{\mathbb{R}^n} (K'g)(x)f(x)d^n x = \int_{\mathbb{R}^n} g(x)(Kf)(x)d^n x$$

for all $g \in L^{p'}$ and $f \in L^p$. Since it suffices to consider a dense set we choose $f, g \in \mathcal{S}(\mathbb{R}^n)$ and abbreviate $K_\varepsilon(x) := K(x)\chi_{\varepsilon<|x|<1/\varepsilon}(x)$. Then using dominated convergence and (iii) we obtain

$$\begin{aligned}\int_{\mathbb{R}^n} g(x)(Kf)(x)d^n x &= \int_{\mathbb{R}^n} g(x)\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^n} K_\varepsilon(x-y)f(y)d^n y\, d^n x = \\ &= \int_{\mathbb{R}^n} g(x)\lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^n} K_\varepsilon(x-y)(f(y)-f(x))d^n y\, d^n x \\ &= \lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^n} g(x)\int_{\mathbb{R}^n} K_\varepsilon(x-y)(f(y)-f(x))d^n y\, d^n x \\ &= \lim_{\varepsilon\downarrow 0}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} g(x)K_\varepsilon(x-y)f(y)d^n y\, d^n x.\end{aligned}$$

Now invoking Fubini and repeating the previous operations in reverse order shows that the adjoint operator is associated with the kernel $K'(x) = K(-x)$. Since $K'(x)$ is a Calderón–Zygmund kernel if and only if $K(x)$ is, we conclude that K is also bounded on $L^{p'}$ for $1 < p \le 2$, that is for $2 \le p' < \infty$. $\square$

Example 9.8. Note that the Hilbert transform is not bounded on $L^\infty(\mathbb{R})$. Indeed, observe that

$$\int_{\mathbb{R}} \frac{1}{x-y}\chi_{[-1,1]}(y)dy = \log\big(\frac{1+x}{1-x}\big)$$

is unbounded. Hence we cannot cover the case $p=\infty$ in general. Moreover, by duality, the claim also fails at $p=1$ in general. ◇

Example 9.9. Let us look once again at the Poisson problem on $\mathbb{R}^n$. Given $f\in\mathcal{S}(\mathbb{R}^n)$ the solutions is

$$u(x) := \int_{\mathbb{R}^n} \Phi(x-y)f(y)d^n y,$$

where Φ is the fundamental solution (8.47). Moreover, we have

$$\begin{aligned}\partial_j u(x) &= \int_{\mathbb{R}^n} (\partial_j\Phi)(x-y)f(y)d^n y\\ \partial_j\partial_k u(x) &= \lim_{\varepsilon\downarrow 0}\int_{\varepsilon<|x-y|} (\partial_j\partial_k\Phi)(x-y)f(y)d^n y - \frac{1}{n}f(x).\end{aligned}$$

Hence the Calderón–Zygmund theorem implies that convolution with the fundamental solution provides a solution $u\in W^{2,p}(\mathbb{R}^n)$ for every $f\in L^p(\mathbb{R}^n)$, $1<p<\infty$. Moreover, the solution satisfies $\|u\|_{2,p}\le C\|f\|_p$ and is unique (since the difference of two solutions is harmonic on $\mathbb{R}^n$ and hence must vanish by the man value property). ◇

So far we took a singular integral operator as our point of departure. However, when using the Fourier transform to solve a constant coefficient partial differential equation (cf. Section 8.3) the solution operator arises naturally as a Fourier multiplier. Hence it is desirable to read off mapping properties of the corresponding operator directly from the multiplier. This is the content of the **Mikhlin multiplier theorem**.[5] Its proof uses a decomposition of a function along a geometric scale.

Lemma 9.17 (dyadic partition of unity)**.** *There exists a function $\psi\in\mathbb{C}_c^\infty(\mathbb{R}^n\setminus 0)$ such that*

$$\sum_{j\in\mathbb{Z}}\psi(2^{-j}x) = 1, \qquad x\neq 0. \tag{9.38}$$

Moreover, ψ can be chosen radial with support in $1/2\le|x|\le 2$ and such that at most two terms in the above sum are nonzero.

Proof. Choose $\chi\in C_c^\infty((0,\infty))$ such that $\chi(r)=1$ for $0<r\le 1$ and $\chi(r)=0$ for $r\ge 2$. Then $\psi(x) := \chi(|x|)-\chi(2|x|)$ is the required function.

[5]Solomon Mikhlin (1908–1990), Soviet mathematician

Indeed
$$\sum_{j=-m}^{m} \psi(2^{-j}x) = \chi(2^{-m}|x|) - \chi(2^{m+1}|x|)$$
and for $2^{-m} \le |x| \le 2^m$ this equals one as required. The additional properties are straightforward. □

Recall that m is called an L^p multiplier if the associated operator
$$A_m f := (m\hat{f})^\vee, \qquad f \in \mathcal{S}(\mathbb{R}^n)$$
extends to a bounded map in $L^p(\mathbb{R}^n)$.

Theorem 9.18 (Mikhlin). *Suppose $m : \mathbb{R}^n \to \mathbb{C}$ satisfies*
$$|\partial^\alpha m(p)| \le \frac{B}{|p|^{|\alpha|}} \tag{9.39}$$
for any multi-index of order $|\alpha| \le n+2$. Then m is an L^p multiplier for any $1 < p < \infty$.

Proof. We use a dyadic partition of unity ψ to decompose m into a sum of $m_j(p) := \psi(2^{-j}p)m(p)$. Then the corresponding operator is associated with the kernel $K_j(x) := (2\pi)^{-n/2}(m)^\vee(x)_j$. By Lemma 8.3 we conclude that $K_j \in L^1(\mathbb{R}^n)$ and hence satisfies the claim for any $1 \le p \le \infty$ by Young's inequality. To establish the claim for m we will show that
$$K^N(x) := \sum_{j=-N}^{N} K_j(x)$$
satisfies
$$|\nabla K^N(x)| \le C_n |x|^{-n-1}$$
uniformly in N. Here, and below, C_n stands for a constant which is an n-dependent multiple of B. Then Lemma 9.13 implies
$$\|K^N * f\|_p \le C_{p,n}\|f\|_p$$
(since both the constant in (ii) as well as the L^2 bound, which is just $\|m\|_\infty \le B$, are linear in B and independent of N) and since for $f \in \mathcal{S}(\mathbb{R}^n)$ we have $K^N * f \to (m\hat{f})^\vee$ uniformly, Fatou's lemma implies
$$\int |(m\hat{f})^\vee|^p d^n x = \int \lim_N |K^N * f|^p d^n x \le \liminf_N \int |K^N * f|^p d^n x \le C_{p,n}\|f\|_p^p$$
as required. Hence it remains to establish the estimate for K^N. Since $\psi(2^{-j}p)$ has support in $2^{j-1} \le |x| \le 2^j$, we conclude $|\partial^\alpha m_j(p)| \le C_n 2^{-j|\alpha|}$ as well as $\|\partial^\alpha m_j\|_1 \le C_n 2^{jn} 2^{-j|\alpha|}$ for all $|\alpha| \le n+2$. Similarly we obtain $\|\partial^\alpha p_k m_j(p)\|_1 \le C_n 2^{jn} 2^{-j(|\alpha|-1)}$ for all $|\alpha| \le n+2$. Taking the Fourier transform we get
$$\|x^\alpha \nabla K_j(x)\|_\infty \le C_n 2^{j(n-|\alpha|+1)}$$

for all $|\alpha| \le n+2$. Finally, summing over the corresponding monomials we get

$$\|\nabla K_j(x)\|_\infty \le \frac{C_n 2^{j(n-r+1)}}{|x|^r}$$

for all $r \le n+2$.

Now using these estimates with $r = 0, n+2$ we have

$$\begin{aligned}
|\nabla K^N(x)| &\le \sum_j |\nabla K_j(x)| \le \sum_{2^j \le |x|^{-1}} |\nabla K_j(x)| + \sum_{2^j > |x|^{-1}} |\nabla K_j(x)| \\
&\le C_n \sum_{2^j \le |x|^{-1}} 2^{j(n+1)} + \frac{C_n}{|x|^{n+2}} \sum_{2^j > |x|^{-1}} 2^{-j} \\
&\le C_n \frac{1}{|x|^{n+1}} \frac{1}{1-2^{-n-1}} + \frac{C_n}{|x|^{n+2}} |x| \frac{1}{1-2^{-1}} \le \frac{4C_n}{|x|^{n+1}}
\end{aligned}$$

as required. □

Example 9.10. The prototypical examples are the multipliers corresponding to the Hilbert, Riesz, and double Riesz transform. These examples also show that the estimate fails for $p=1$ and $p=\infty$ in general. Somewhat more general, any multiplier of the form

$$m(p) = \frac{P(p)}{|p|^r},$$

where P is a homogenous polynomial of degree r, satisfies the assumption (9.39) for any multi-index α. ⋄

Problem 9.17. *Show that the condition*

$$\sup_{2|y| \le |x|} \frac{|K(x) - K(x-y)|}{|y|^\gamma} \le \frac{B}{|x|^{n+\gamma}}$$

for any fixed $\gamma \in (0,1]$ implies the Hörmander condition (ii) (with a different constant). Moreover, show that if this condition holds for some γ_0, then it holds for all $\gamma \le \gamma_0$. Finally, show that (9.32) *implies this condition for $\gamma = 1$ (and hence for all $\gamma \le 1$).*

Problem 9.18. *Show that the norm of $K_r(x) := \frac{1}{x}\chi_{|x| \le r}(x)$ in $L^p(\mathbb{R})$ is independent of r.*

Problem 9.19. *Compute the Fourier multiplier of the Riesz transform $K_j(x) := x_j |x|^{-n-1}$ (for $n \ge 2$). (Hint: Lemma 8.27.)*

Bibliography

[1] H. W. Alt, *Lineare Funktionalanalysis*, 4th ed., Springer, Berlin, 2002.

[2] H. Bauer, *Measure and Integration Theory*, de Gruyter, Berlin, 2001.

[3] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, New York, 2011.

[4] E. DiBenedetto, *Real Analysis*, Birkhäuser, Boston, 2002.

[5] L. C. Evans, *Weak Convergence Methods for nonlinear Partial Differential Equations*, CBMS 74, American Mathematical Society, Providence, 1990.

[6] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd. ed., Wiley, Hoboken NJ, 1999.

[7] L. Grafakos, *Classical Fourier Analysis*, 2nd ed., Springer, New York, 2008.

[8] L. Grafakos, *Modern Fourier Analysis*, 2nd ed., Springer, New York, 2009.

[9] G. Grubb, *Distributions and Operators*, Springer, New York, 2009.

[10] E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer, Berlin, 1965.

[11] E. Lieb and M. Loss, *Analysis*, 2nd ed., Amer. Math. Soc., Providence, 2000.

[12] G. Leoni, *A First Course in Sobolev Spaces*, Amer. Math. Soc., Providence, 2009.

[13] C. Muscalu and W. Schlag, *Classical and Multilinear Harmonic Analysis*, Cambridge University Press, Cambridge, 2013.

[14] F. W. J. Olver et al., *NIST Handbook of Mathematical Functions*, Cambridge University Press, Cambridge, 2010.

[15] I. K. Rana, *An Introduction to Measure and Integration*, 2nd ed., Amer. Math. Soc., Providence, 2002.

[16] H. Royden, *Real Analysis*, Prentice Hall, New Jersey, 1988.

[17] W. Rudin, *Real and Complex Analysis*, 3rd edition, McGraw-Hill, New York, 1987.

[18] B. Simon, *A Comprehensive Course in Analysis*, Amer. Math. Soc., Providence, 2015.

[19] E. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton UP, Princeton, 1970.

[20] M. E. Taylor, *Measure Theory and Integration*, Amer. Math. Soc., Providence, 2006.

[21] G. Teschl, *Mathematical Methods in Quantum Mechanics; With Applications to Schrödinger Operators*, Amer. Math. Soc., Providence, 2009.

[22] G. Teschl, *Topics in Linear and Nonlinear Functional Analysis*, Amer. Math. Soc., Providence, to appear.

[23] G. Teschl, *Partial Differential Equations*, Lecture Notes.

Glossary of notation

$AC[a,b]$	... absolutely continuous functions, 127
$\arg(z)$	... argument of $z \in \mathbb{C}$; $\arg(z) \in (-\pi, \pi]$, $\arg(0) = 0$
$B_r(x)$	... open ball of radius r around x
$B(X)$	... Banach space of bounded measurable functions
$BV[a,b]$	... functions of bounded variation, 124
$\mathfrak{B}$	$:= \mathfrak{B}^1$
$\mathfrak{B}^n$	... Borel σ-algebra of $\mathbb{R}^n$, 6
$\mathbb{C}$	... the set of complex numbers
$C(U)$	... set of continuous functions from U to $\mathbb{C}$
$C_0(U)$	... set of continuous functions vanishing on the boundary ∂U, 175
$C_c(U)$	... set of compactly supported continuous functions
$C_{per}[a,b]$	... set of periodic continuous functions (i.e. $f(a) = f(b)$)
$C^k(U)$	... set of k times continuously differentiable functions
$C_b^k(U)$	... functions in C^k with derivatives bounded, 176
$C_0^k(U)$	... functions in C^k with derivatives vanishing on the boundary ∂U, 176
$C_{pg}^\infty(U)$	... set of smooth functions with at most polynomial growth, 254
$C_c^\infty(U)$	... set of compactly supported smooth functions
$C(U,Y)$	... set of continuous functions from U to Y
$\mathscr{C}(X,Y)$	... set of compact linear operators from X to Y

$\chi_\Omega(.)$	... characteristic function of the set Ω
$\mathfrak{D}(.)$	... domain of an operator
$\delta_{n,m}$	... Kronecker delta
det	... determinant
dim	... dimension of a linear space
div	... divergence of a vector filed, 60
$\mathrm{diam}(U)$	$:= \sup_{(x,y)\in U^2} d(x,y)$ diameter of a set
$\mathrm{dist}(U,V)$	$:= \inf_{(x,y)\in U\times V} d(x,y)$ distance of two sets
e	... Napier's constant, $\mathrm{e}^z = \exp(z)$
$\mathrm{GL}(n)$	... general linear group in n dimensions
$\Gamma(z)$	... gamma function, 55
$\Gamma(f_1,\dots,f_n)$	... Gram determinant, 57
$\mathfrak{H}$	... a Hilbert space
conv(.)	... convex hull
$\mathcal{H}(U)$	... set of holomorphic functions on a domain $U \subseteq \mathbb{C}$
$H^k(U)$	... Sobolev space, 180, 239
$H^k_0(U)$	... Sobolev space, 184
$H^{r,s}(\mathbb{R}^n)$	... weighted Sobolev space, 243
i	... complex unity, $\mathrm{i}^2 = -1$
Im(.)	... imaginary part of a complex number
inf	... infimum
$J_f(x)$	$:= \det df(x)$ Jacobi determinant of f at x, 50
$\mathrm{Ker}(A)$	... kernel of an operator A
λ^n	... Lebesgue measure in $\mathbb{R}^n$, 48
$\mathscr{L}(X,Y)$	... set of all bounded linear operators from X to Y
$\mathscr{L}(X)$	$:= \mathscr{L}(X,X)$
$L^p(X,d\mu)$	... Lebesgue space of p integrable functions, 73
$L^\infty(X,d\mu)$	... Lebesgue space of bounded functions, 74
$L^p_{loc}(X,d\mu)$	... locally p integrable functions, 75
$\mathcal{L}^1(X.d\mu)$	... space of integrable functions, 38
max	... maximum
$\mathcal{M}(X)$	... finite complex measures on X, 119
$\mathcal{M}_{reg}(X)$	... finite regular complex measures on X, 122
$\mathcal{M}(f)$	... Hardy–Littlewood maximal function, 268

$\mathbb{N}$	... the set of positive integers
$\mathbb{N}_0$	$:= \mathbb{N} \cup \{0\}$
$n(\gamma, z_0)$	... winding number
∇f	$= (\partial_1 f, \dots, \partial_m f)$ gradient in $\mathbb{R}^n$
$\nu(x_0)$	... outward pointing unit normal vector, 59
$O(.)$	... Landau symbol, $f = O(g)$ iff $\limsup_{x\to x_0} \lvert f(x)/g(x)\rvert < \infty$
$o(.)$	... Landau symbol, $f = o(g)$ iff $\lim_{x\to x_0} \lvert f(x)/g(x)\rvert = 0$
$\mathbb{Q}$	... the set of rational numbers
$\mathbb{R}$	... the set of real numbers
$\rho(A)$	... resolvent set of an operator A
$\mathrm{Ran}(A)$	... range of an operator A
$\mathrm{Re}(.)$	... real part of a complex number
$\sigma(A)$	... spectrum of an operator A
σ^{n-1}	... surface measure on S^{n-1}, 53
S^{n-1}	$:= \{x \in \mathbb{R}^n \mid \lvert x\rvert = 1\}$ unit sphere in $\mathbb{R}^n$
S_n	$:= n\pi^{n/2}/\Gamma(\frac{n}{2}+1)$, surface area of the unit sphere in $\mathbb{R}^n$, 53
$\mathrm{sign}(z)$	$:= z/\lvert z\rvert$ for $z \neq 0$ and 1 for $z = 0$; complex sign function
$\mathcal{S}^n$	... semialgebra of rectangles in $\mathbb{R}^n$, 2
$\bar{\mathcal{S}}^n$	... algebra of finite unions of rectangles in $\mathbb{R}^n$, 2
$\mathcal{S}(\mathbb{R}^n)$	... Schwartz space, 219
sup	... supremum
$\mathrm{supp}(f)$	... support of a function f
$\mathrm{supp}(\mu)$	... support of a measure μ, 18
$\mathrm{span}(M)$	... set of finite linear combinations from M
$W^{k,p}(U)$	... Sobolev space, 180
$W_0^{k,p}(U)$	... Sobolev space, 184
V_n	$:= \pi^{n/2}/\Gamma(\frac{n}{2}+1)$, volume of the unit ball in $\mathbb{R}^n$, 54
$\mathbb{Z}$	... the set of integers
$\mathbb{I}$	... identity operator
$\sqrt{z}$	... square root of z with branch cut along $(-\infty, 0)$
z^*	... complex conjugation
$f^\star$	... symmetric rearrangement of a function, 104
A^*	... adjoint of an operator A
$\overline{A}$	... closure of an operator A
$\hat{f}$	$:= \mathcal{F}f$, Fourier transform of f, 217
$\check{f}$	$:= \mathcal{F}^{-1}f$, inverse Fourier transform of f, 221
$\lvert x\rvert$	$:= \sqrt{\sum_{j=1}^n \lvert x_j\rvert^2}$ Euclidean norm in $\mathbb{R}^n$ or $\mathbb{C}^n$
$\lvert\Omega\rvert$	... Lebesgue measure of a Borel set Ω
$\Vert.\Vert$	... norm in a Banach space
$\Vert.\Vert_p$	... norm in the Banach space L^p, 73
$\langle ., .. \rangle$	... scalar product in a Hilbert space $\mathfrak{H}$

$\oplus$	... direct/orthogonal sum of vector spaces
$\otimes$	... product of σ-algebras and measures
$\dot\cup$	... union of disjoint sets, 6
$\lfloor x \rfloor$	$:= \max\{n \in \mathbb{Z} \| n \leq x\}$, floor function
$\lceil x \rceil$	$:= \min\{n \in \mathbb{Z} \| n \geq x\}$, ceiling function
$\langle x \rangle$	$:= \sqrt{1+\|x\|^2}$, $x \in \mathbb{C}^n$ (Japanese bracket)
∂	$:= (\partial_1 f, \dots, \partial_m f)$ gradient in $\mathbb{R}^m$
∂_α	... partial derivative in multi-index notation, 176
$\partial_x F(x,y)$	... partial derivative with respect to x
∂U	$:= \overline{U} \setminus U^\circ$ boundary of the set U
$\overline{U}$	... closure of the set U
U°	... interior of the set U
$V \subset\subset U$	... V is relatively compact with $\overline{V} \subset U$
$M^\perp$	... orthogonal complement in a Hilbert space
(λ_1, λ_2)	$:= \{\lambda \in \mathbb{R} \,\|\, \lambda_1 < \lambda < \lambda_2\}$, open interval
$[\lambda_1, \lambda_2]$	$:= \{\lambda \in \mathbb{R} \,\|\, \lambda_1 \leq \lambda \leq \lambda_2\}$, closed interval
$x_n \to x$	... norm convergence
$x_n \rightharpoonup x$	... weak convergence
$x_n \overset{*}{\rightharpoonup} x$	... weak-$*$ convergence

Index