Distributions and Function Spaces\textsuperscript{1}

\textbf{Jose-Luis Menaldi}\textsuperscript{2}

Current Version: 11 November 2016\textsuperscript{3}

First Version: 2014\textsuperscript{4}

\textsuperscript{1}©Copyright 2014. No part of this book may be reproduced by any process without prior written permission from the author.

\textsuperscript{2}Wayne State University, Department of Mathematics, Detroit, MI 48202, USA (e-mail: menaldi@wayne.edu).

\textsuperscript{3}Long Title. \textit{Distributions and Function Spaces: Schwartz Theory of Distributions, Sobolev and Besov Spaces}

\textsuperscript{4}This book is being progressively updated and expanded. If you discover any errors or you have any improvements to suggest, please e-mail the author.
## Contents

Preface vii

Introduction ix

1 Abstract Integration 1
   1.1 Daniell Integrals .............................................. 1
       1.1.1 Null or Negligible Sets .................................. 6
       1.1.2 Integrable Functions ...................................... 8
       1.1.3 Measurable Functions ...................................... 18
   1.2 Uniform Integrability ....................................... 23
       1.2.1 Main Properties ............................................ 23
       1.2.2 Mean Convergence .......................................... 29
       1.2.3 Convergence in Norm ...................................... 34
   1.3 Vector-valued Integrals .................................... 41
       1.3.1 Metric Space of Measurable Functions ..................... 42
       1.3.2 With Values in a Banach Space ........................... 43

2 Basic Functional Analysis 49
   2.1 Background and Introduction ............................... 49
       2.1.1 Simple Spectral Analysis ................................. 51
       2.1.2 Three Basic Results ...................................... 54
       2.1.3 Examples and Comments ................................... 57
   2.2 Compactness and Separability .............................. 60
       2.2.1 Linear Functionals ........................................ 60
       2.2.2 Nonlinear Functional ..................................... 62
       2.2.3 Baire Category Arguments ................................ 65
   2.3 Three Essential Principles ................................. 66
       2.3.1 Uniformly Boundedness Principle ......................... 68
       2.3.2 Open Mappings Theorem .................................. 70
       2.3.3 Closed Graph Theorem .................................... 71
       2.3.4 Hahn-Banach Theorem .................................... 72
   2.4 More on Lebesgue Spaces ................................... 75
       2.4.1 Weak Convergence ......................................... 77
       2.4.2 Totally Bounded Sets ..................................... 79
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 Basic Interpolation Questions</td>
<td>83</td>
</tr>
<tr>
<td>2.5.1 Preliminary Interpolation</td>
<td>86</td>
</tr>
<tr>
<td>2.5.2 Marcinkiewicz Interpolation Theorem</td>
<td>87</td>
</tr>
<tr>
<td>2.5.3 Riesz-Thorin Interpolation Theorem</td>
<td>90</td>
</tr>
<tr>
<td>2.5.4 Intermediate Spaces</td>
<td>93</td>
</tr>
<tr>
<td>3 Elements of Distributions Theory</td>
<td>97</td>
</tr>
<tr>
<td>3.1 Locally Convex Spaces</td>
<td>97</td>
</tr>
<tr>
<td>3.1.1 Dual Spaces</td>
<td>103</td>
</tr>
<tr>
<td>3.1.2 Inductive Limits</td>
<td>105</td>
</tr>
<tr>
<td>3.1.3 Test Function Spaces</td>
<td>108</td>
</tr>
<tr>
<td>3.2 Calculus with Distributions</td>
<td>114</td>
</tr>
<tr>
<td>3.2.1 Positivity, Differentiability and Integrability</td>
<td>118</td>
</tr>
<tr>
<td>3.2.2 Support and Finite Order</td>
<td>126</td>
</tr>
<tr>
<td>3.2.3 Distribution on Manifolds</td>
<td>131</td>
</tr>
<tr>
<td>3.2.4 Avoiding Inductive Limits</td>
<td>132</td>
</tr>
<tr>
<td>3.3 More Operations and Localization</td>
<td>134</td>
</tr>
<tr>
<td>3.3.1 Product of Distributions</td>
<td>134</td>
</tr>
<tr>
<td>3.3.2 Convolution of Distributions</td>
<td>136</td>
</tr>
<tr>
<td>3.3.3 Local Structure</td>
<td>144</td>
</tr>
<tr>
<td>3.3.4 Recap on Inductive Limits</td>
<td>146</td>
</tr>
<tr>
<td>4 Introduction to Sobolev Spaces</td>
<td>151</td>
</tr>
<tr>
<td>4.1 Density and Extension</td>
<td>152</td>
</tr>
<tr>
<td>4.1.1 Regularity on the Domain</td>
<td>153</td>
</tr>
<tr>
<td>4.1.2 Lipschitz Transformation</td>
<td>157</td>
</tr>
<tr>
<td>4.2 Imbedding and Compactness</td>
<td>158</td>
</tr>
<tr>
<td>4.2.1 Some Typical Estimates</td>
<td>161</td>
</tr>
<tr>
<td>4.2.2 General Imbedding</td>
<td>165</td>
</tr>
<tr>
<td>4.3 Traces on the Boundary</td>
<td>166</td>
</tr>
<tr>
<td>4.3.1 In Half-space</td>
<td>166</td>
</tr>
<tr>
<td>4.3.2 In a Smooth Domain</td>
<td>169</td>
</tr>
<tr>
<td>4.3.3 Spaces on the Boundary</td>
<td>172</td>
</tr>
<tr>
<td>4.4 Fractional Order Spaces</td>
<td>174</td>
</tr>
<tr>
<td>4.4.1 Discussion and Definition</td>
<td>174</td>
</tr>
<tr>
<td>4.4.2 Basic Properties</td>
<td>175</td>
</tr>
<tr>
<td>5 Basic Fourier Transform</td>
<td>183</td>
</tr>
<tr>
<td>5.1 Smooth Functions</td>
<td>184</td>
</tr>
<tr>
<td>5.2 Tempered Distributions</td>
<td>187</td>
</tr>
<tr>
<td>5.3 Integrable Functions</td>
<td>190</td>
</tr>
<tr>
<td>5.4 Periodic Functions</td>
<td>192</td>
</tr>
<tr>
<td>5.5 Fourier Multiplier</td>
<td>198</td>
</tr>
</tbody>
</table>
# Contents

## 6 Besov and Sobolev Spaces

6.1 Hilbertian Sobolev Spaces .................................................. 208
   6.1.1 In the Whole Space .................................................... 208
   6.1.2 In Continuous Domains .................................................. 216
   6.1.3 Trace Operator ......................................................... 223

6.2 Riesz and Bessel Potentials .............................................. 229
   6.2.1 Initial Discussion ..................................................... 230
   6.2.2 Bessel Kernel and Potentials ....................................... 235
   6.2.3 Fundamental Solutions .............................................. 245

6.3 Besov and Sobolev Relations ............................................. 247
   6.3.1 Besov spaces ......................................................... 248
   6.3.2 Bessel Potential Spaces ........................................... 251
   6.3.3 Traces on Besov Spaces ............................................ 255

## Appendix A Solutions to Exercises

A.1 Abstract Integration ....................................................... 263

A.2 Basic Functional Analysis ............................................... 273

A.3 Elements of Distributions Theory ...................................... 281

A.4 Introduction to Sobolev Spaces ........................................ 311

A.5 Basic Fourier Transform ................................................. 313

A.6 Besov and Sobolev Spaces ............................................... 317

## Appendix B Measure and Integration

B.1 Classes of Sets .............................................................. 319
   B.1.1 First Properties ...................................................... 320
   B.1.2 Topology Included .................................................... 322
   B.1.3 Measurable Functions ............................................... 324
   B.1.4 Some Tools ............................................................. 327

B.2 Caratheodory’s Arguments ................................................. 329
   B.2.1 Caratheodory’s construction ....................................... 329
   B.2.2 From a Semi-Ring ..................................................... 333

B.3 Inner Measure Approach .................................................. 335
   B.3.1 Inner Measures ....................................................... 335
   B.3.2 Inner Construction ................................................... 336

B.4 Examples and Convergence ............................................... 340
   B.4.1 Lebesgue Measures ................................................... 340
   B.4.2 Convergence in Measure ............................................ 344
   B.4.3 Almost Measurable Functions ..................................... 348

B.5 Integration Theory ......................................................... 351
   B.5.1 Definition and Properties .......................................... 352
   B.5.2 Cartesian Products .................................................. 358
   B.5.3 Some Inequalities .................................................... 361
   B.5.4 Orthogonal Projection .............................................. 365
   B.5.5 Lebesgue Spaces ..................................................... 367
   B.5.6 Radon-Nikodym Derivative ....................................... 373

B.6 Essential Complements ................................................... 375

---

Preface

This project has several parts, of which this book is the second one. The first part deals with measure and integration theory, while part three is dedicated to elementary probability (after measure theory). In part four, stochastic integrals are studied in some details, and in part five, stochastic ordinary differential equations are discussed, with a clear emphasis on estimates. Each part was designed independent (as much as possible) of the others, but it makes a lot of sense to consider all five parts as a sequence.

This part two begins with a second iteration on measure and integration theory. First we reset the theory following Daniell integral, we make a deep recall the uniform integrability concept and we complete the discussion with vector-valued integrable functions. For reader convenience, Appendix B is essentially an enlarged review of part one, to which Chapter 1 should be added to motivate our interest in basic functional analysis of the Chapter 2. In Chapters 3 and 5, we introduce the theory of distributions with its delicate inductive limit topology and ending with the definition of the Fourier transform and including a proof of Bochner Theorem. Chapters 4 and 6 gives a comfortable beginning of Sobolev and Besov spaces, with the idea of before and after the Fourier transform. Thus, this book could be used as a second-semester course in Real Analysis. Depending on the instructor’s interest and the amount of exercises added, Chapter 1 may be (partially, e.g., Daniell and Vector integration) skipped (even if this is not recommended for an advanced level), and Chapters 2, 3 and 5 could be used as a short introduction to Schwartz theory of distribution.

Most of the style is formal (propositions, theorems, remarks), but there are instances where a more narrative presentation is used, the purpose being to force the student to pause and fill-in the details. Practically, there are no specific section of exercises, giving to the instructor the freedom of choosing problems from various sources (and according to a particular interest of subjects) and reinforcing the desired orientation. There is no intention to diminish the difficulty of the material to put students at ease, on the contrary, all points presented as blunt as possible, even some times shorten some proofs, but with appropriate references. For instance, we assume that most of the material in our previous book Menaldi [89] (on Measure and Integration) has been reviewed.

This book is written for the instructor rather than for the student in a sense that the instructor (familiar with the material) has to fill-in some (small) details and selects exercises to give a personal direction to the course. Therefore, this
material should be taken more as Lecture Notes, addressed indirectly (via an instructor) to the student. In a way, the student seeing this material for the first time may be overwhelmed, but with time and dedication the reader can check most of the points indicated in the references to complete some hard details, perhaps the expression of a guided tour could be used here. Essentially, it is known that a Proposition in one textbook may be an exercise in another, so that most of the exercises at this level are hard or simple, depending on the experience of the student.

In Appendix A, all exercises are re-listed by section, but now, most of them have a (possible) solution. Certainly, this appendix is not for the first reading, i.e., this part is meant to be read after having struggled (a little) with the exercises. Sometimes, there are many ways of solving problems, and depending of what was developed “in the theory”, solving the exercises could have alternative ways. The instructor will find that some exercises are trivial while other are not simple. It is clear that what we may call “Exercises” in one textbook could be called “Propositions” in others. This part two does not have a large number of exercises as in part one does, but the instructor may find a lot of exercises in some of the references quoted in the text.

The combination of parts I, II, and III is neither a comprehensive course in measure and integration (but a certain number of generalizations suitable for probability are included), nor a basic course in probability (but most of language used in probability is discussed), nor a functional analysis course (but function spaces and the three essential principles are addressed), nor a course in theory of distribution (but most of the key component are there), and certainly nor a course in Sobolev or Besov spaces (but a quick introduction is there). One of the objectives of these first three books is to show the reader a large open door to probability (and partial differential equations), without committing oneself to probability (or partial differential equations) and without ignoring hard parts in measure and integration theory.

Michigan (USA),  

Jose-Luis Menaldi, October 2016
Introduction

Even if this is a continuation of our first part-book Menaldi [89], only certain material is essential to the understanding of what follow. However, it may be convenient for the reader to check certainly points as needed. Actually, in Appendix B, the reader will find a quick summary of some of the key ‘background’ material somehow useful (from the notation and conceptual viewpoint). Certainly, this is not necessary, it should be taken as a ‘service chapter’ for the ‘convenience’ of the reader.

After covering a so-called more basic measure theory, we are interested in more advanced topic. The points discussed in the previous first part were basic and abstract, without committing ourself to any particular direction, e.g., including subjects such that Borel measures and approximation by smooth functions.

In this second part, we begin by reinforcing some points regarding the theory of integrals. We consider in great details the concept of uniform integrability, which is very important various aspect of analysis. We reset the integral theory as a Daniell functional and we extend the integral to functions with values in a Banach space.

Therefore, we assume that the reader is familiar with the rudiments of metric, Banach and Hilbert spaces, and we declare our interest in functional analysis, without making a course on it. However, the order in which the material is developed can be modified.

Next, most of the key results in functional analysis are only stated, with some comments on the proofs, the point is to reach the so-called Schwartz’ theory of distributions and discuss elements of the Fourier analysis. By now, the reader is able to remark the various aspects of integrable, continuous and differentiable functions. For instance, in the distribution sense, we are able to differentiate locally integrable functions and to make a rigorous sense out of several formal calculations.
Chapter 1

Abstract Integration

Recall that given a measure space \((\Omega, \mathcal{F}, \mu)\), we denote by \(L^1 = L^1(\Omega, \mathcal{F}, \mu)\) the vector space of integrable functions. Besides defining integrable functions, we call a function \(f\) integrable on \(F\) in \(\mathcal{F}\) if \(1_F f\) is integrable and we write

\[
\int_{F} f \, d\mu = \int_{\Omega} 1_F f \, d\mu.
\]

Moreover, \(f\) is called \(\sigma\)-integrable if there exists a sequence of measurable sets \(F_n \in \mathcal{F}\) with \(\Omega = \bigcup_n F_n\) such that \(1_{F_n} f\) is integrable. Sometimes, a measurable function \(f\) is called quasi-integrable if either \(f^+\) or \(f^-\) is integrable. Furthermore, in some books, the terms summable and integrable are used instead of integrable and quasi-integrable, respectively.

Now, we reset the theory of measure and integral by considering first the integral, and a posteriori the measure. This is more a functional analysis approach, where we build up the theory from an elementary class of functions, e.g., from continuous functions and their Riemann integrals. Even if measure theory is not a priori necessary, it is preferred to have some minimum exposure to it before reading these sections. Next, we discuss (independently) the so-called uniform integrability property and the vector-valued integrals.

In any case, this chapter could replace (briefly) our previous volume [89], but the intention is to complement (rather than to substitute) a traditional approach to measure and integration theory. Also, some readers may want to check some points (as needed) in the comprehensive guide (to infinity dimensional analysis) Aliprantis and Border [6], specially for following first three chapters.

1.1 Daniell Integrals

Three independent ways for constructing measures has been seen previously, namely, the outer approach (or Caratheodory’s arguments, in Section B.2), the inner approach (or compact technique, in Section B.3), and the geometric approach (or Hausdorff construction). In this section the inner approach is, in a
way, reconsidered. Moreover, the theory of integration is first developed and as a consequence, measure theory is deduced.

As mentioned early, the Lebesgue integral is based on the measure theory (i.e., first we learn how to measure sets and then we develop the integral); while the Riemann integral was intended as a means of defining area or volumes (i.e., we introduce the integral to be able to measure sets). Actually, measure and integral theory are tied together, and we may begins with either of them.

Certainly, besides the theorems of passage-to-the-limit inside the integral, the construction of the Lebesgue spaces $L^p$, $p \geq 1$, is of great importance, $L^p$ is a Banach space and $L^2$ is a Hilbert space. Let us take a closed look at the completeness of $L^1$, the space of integrable functions. Without developing the measure theory first, we may consider a class of elementary functions $E$ for which the integral $I(\cdot)$ is defined. For instance, if $(X, T)$ is a (local or $\sigma$-compact) topological space and we know how to integrate continuous functions which the integral theory are tied together, and we may begin with either of them.

Hence, without knowing the Lebesgue theory, we may define a pre-Lebesgue normed space $(E^1, || \cdot ||_1)$, which (in general) is not a Banach space, and then we need to complete the space. For instance, we may consider Cauchy sequences in $(E^1, || \cdot ||_1)$, similar to the argument used to pass from the rational numbers to the real numbers, but we need to workout the details on what those limiting functions are. For instance, the reader may check the books Ash [11, Section 4.2, pp. 170–177] or Phillips [100, Chapter 12, pp. 363–394], among others.

The defining properties to complete this approach are the following: Firstly, the vector space $E$ is a lattice of functions from $X$ into $\mathbb{R}$, namely,

$$
\begin{align*}
(a) & \text{ If } \varphi, \psi \in E \text{ and } a, b \in \mathbb{R} \text{ then } a\varphi + b\psi \in E, \\
(b) & \text{ If } \varphi, \psi \in E \text{ then } \max\{\varphi, \psi\} \in E, \\
(c) & \text{ If } \varphi \in E \text{ and } \varphi \geq 0 \text{ then } \min\{\varphi, 1\} \in E.
\end{align*}
$$

The last property is irrelevant if the function 1 (identically to the number 1) belongs to $E$. Usually, only the first two property are used to define a (vector) lattice, and with (1.1-c) it is referred to a Stone (vector) lattice. Clearly, on a lattice the max and the min operators are defined pointwise, we have $\min\{\varphi, \psi\} = -\max\{-\varphi, -\psi\}$ and we may define $|\varphi| = \varphi^+ + \varphi^-$, where $\varphi^+ = \max\{\varphi, 0\}$ and $\varphi^- = -\min\{\varphi, 0\}$. Usually, the symbols $\wedge$ and $\vee$ denote the min and the max, respectively. From an abstract point of view, the name vector lattice include a compatibility condition between the vector and the lattice structures, namely, the positive cone $P = \{\varphi \in E : \varphi \geq 0\}$ defines a partial order $\leq$ with the properties: if $\varphi \leq \psi$ then $\varphi + \phi \leq \psi + \phi$ for every $\phi$ and $\alpha\varphi \leq \alpha\psi$ (or $\alpha\varphi \geq \alpha\psi$) for every scalar $\alpha \geq 0$ (or $\alpha \leq 0$). For instance, the reader can check the following equalities: $(\varphi \vee \psi) + \phi = (\varphi + \phi) \vee (\psi + \phi)$, $\alpha(\varphi \vee \psi) = (\alpha\varphi) \vee (\alpha\psi)$, for $\alpha \geq 0$, $\alpha \leq 0$, $\alpha \neq 0$, $\alpha = 0$.
among others, e.g., see Yosida [135, Section XII.2, pp. 364–370].

Secondly, a pre-integral real-valued maps \( I \) is defined on \( E \), i.e.,

\[
\begin{align*}
\text{(a) If } & \varphi, \psi \in E \text{ and } a, b \in \mathbb{R} \text{ then } I(a\varphi + b\psi) = aI(\varphi) + bI(\psi), \\
\text{(b) If } & \varphi, \psi \in E \text{ and } \varphi \geq \psi \text{ (pointwise) then } I(\varphi) \geq I(\psi), \\
\text{(c) If } & \{\varphi_k\} \subset E \text{ and } \varphi_k \downarrow 0 \text{ (pointwise decreasing) then } I(\varphi_k) \to 0.
\end{align*}
\]

The first property states the linear character of the integral \( I \), and instead of monotonicity (1.2-b) it suffices the positivity, i.e., \( \varphi \geq 0 \) implies \( I(\varphi) \geq 0 \). Also, we have \( |I(\varphi)| \leq I(|\varphi|) \), for every \( \varphi \in E \). Next, based on the first two properties, the condition (1.2-c) is equivalent to the monotone continuity, namely, if \( \{\varphi_k\} \subset E \) is a monotone sequence and \( \varphi_k \to \varphi \) with \( \varphi \in E \) then \( I(\varphi_k) \to I(\varphi) \).

If every elementary function is bounded then we may endow \( E \) with the sup-norm \( \|\varphi\| = \sup\{|\varphi(x)| : x \in X\} \) and consider \( I \) as a linear functional on \((E, \| \cdot \|)\). Thus, if \( 1_X \) belongs to \( E \) then a map \( I \) satisfying (1.2-a) and (1.2-b) will also satisfy

\[ |I(\varphi)| \leq \|\varphi\| I(1_X), \quad \forall \varphi \in E \]

i.e., \( I \) is a bounded linear functional on \((E, \| \cdot \|)\), still not exactly the monotone continuity (1.2-c). However, if \( X \) is a compact space, \( E = C(X) \) is the space of continuous real functions on \( X \) and \( \varphi_k \to \varphi \) (pointwise decreasing) with \( \varphi_k, \varphi \in C(X) \) then Dini’s Theorem implies that \( \|\varphi_k - \varphi\| \to 0 \) and so the monotone continuity (1.2-c) is satisfies, i.e., only (1.2-a) and (1.2-b) are relevant in this case.

**Remark 1.1.** Dini’s Theorem affirms that if a sequence of continuous functions with compact support \( \{\varphi_n : n \geq 1\} \) is pointwise decreasing to 0 then \( \|\varphi_n\| \to 0 \). Indeed, because the sequence is pointwise decreasing to 0, we have \( \text{supp}(\varphi_n) \subset \text{supp}(\varphi_1) = K \), and therefore, given \( \varepsilon > 0 \) and \( x \in K \) there exists \( \eta = \eta(\varepsilon, x) \) such that \( 0 \leq \varphi_n(x) \leq \varepsilon/2 \), for every \( n \geq \eta \). The continuity of \( \varphi_\eta \) ensures that there exists an open set \( U(x, \varepsilon) = U(x, \varepsilon, \eta) \) containing \( x \) such that \( |\varphi_\eta(y) - \varphi_\eta(x)| \leq \varepsilon/2 \), for every \( y \in U(x, \varepsilon) \), i.e.,

\[ 0 \leq \varphi_\eta(y) \leq [\varphi_\eta(y) - \varphi_\eta(x)] + \varphi_\eta(x) \leq \varepsilon, \quad \forall y \in U(x, \varepsilon). \]

Since the family \( \{U(x, \varepsilon) : x \in K\} \) is an open cover of \( K \), there exists finite subcover, i.e., \( x_1, \ldots, x_k \) such that \( K \subset \bigcup_{i=1}^k U(x_i, \varepsilon) \). Define \( N(\varepsilon) = \max\{\eta(x_1), \ldots, \eta(x_k)\} \), take \( n \geq N(\varepsilon) \) and \( y \in K \) to obtain an \( i \) such that \( y \) belongs to \( U(x_i, \varepsilon) \) and then

\[ 0 \leq \varphi_n(y) \leq \varphi_\eta(y) \leq \varepsilon, \quad \text{where } n \geq \eta = \eta(\varepsilon, x_i). \]

Hence, \( \varphi_n(x) \to 0 \) uniformly in \( x \) belonging to \( K \). Actually, the above argument proves that if a decreasing sequence of functions \( \{\psi_n\} \) satisfies \( \psi_n(x) \to 0 \) as \( n \to 0 \) for every \( x \) then \( \sup_{x \in K} |\psi_n(x)| \to 0 \), for any compact subset \( K \) of the continuity set \( \bigcap_n C(\psi_n) \), i.e., where \( C(\psi_n) \) denotes the set of points where \( \psi_n \) is continuous. \( \square \)
We have

**Lemma 1.2.** Let \( E = C_0(X) \) be the space of real-valued continuous functions with compact support on a locally compact Hausdorff topological space \( X \). Thus \( E \) is a lattice satisfying (1.1) and if \( I \) is a linear monotone functional, i.e., (1.2-a) and (1.2-b) hold, then \( I \) is a pre-integral, i.e., (1.2-c) is also valid.

**Proof.** First let \( K \) be a compact subset of \( X \) and denote by \( C_K(X) \) the subspace of \( C_0(X) \) of all functions with support inside \( K \). Because \( X \) is locally compact Hausdorff space, there exists and open set \( U \supseteq K \) with compact closure \( \overline{U} \). Now by Urysohn’s Theorem there exists a continuous function \( \rho : X \to [0, 1] \) such that \( \rho = 1 \) on \( K \) and \( \rho = 0 \) on \( X \setminus \overline{U} \). For every \( \varphi \) in \( C_K(X) \), belonging to \( C_0(X) \) with \( \text{supp}(\varphi) \subseteq K \), we have \( \varphi = \varphi \rho \) and \( -\|\varphi\| \leq \varphi \rho \leq \|\varphi\| \). Hence

\[
I(\varphi) = I(\varphi \rho), \quad |I(\varphi)| \leq \|\varphi\| I(\rho), \quad \forall \varphi \in C_K(X).
\]

Next, if \( \{ \varphi_n \} \) is a sequence in \( C_0(X) \) pointwise decreasing to 0 then \( 0 \leq \varphi_n \leq \varphi_1 \) and \( \varphi_n \) belongs to \( C_K(X) \), for every \( n \geq 1 \), with \( K = \text{supp}(\varphi_1) \). By Dini’s Theorem, \( \|\varphi_n\| \to 0 \) and therefore \( I(\varphi_n) \to 0 \), as \( n \to \infty \). \( \square \)

A typical application of the above result is the Riemann integral on \( \mathbb{R}^d \) with \( d \geq 1 \), i.e., Lemma 1.2 proves that the class \( E = C_0(\mathbb{R}^d) \) of real-valued continuous functions is a vector lattice and \( I \) defined as the Riemann integral satisfies (1.2), i.e., \( I \) is a pre-integral.

In contract with Lemma 1.2, if \( E \) is a semi-ring of \( 2^X \) and \( E \) is the class of linear combinations of characteristic functions \( 1_E \) with \( E \) in \( E \), i.e., \( \varphi = \sum_{i=1}^{n} a_i 1_{E_i} \) and \( E_i \cap E_j = \emptyset \), when \( i \neq j \), then it suffices to have \( I(\cdot) \) defined (and finite) on characteristic functions \( 1_E \) with \( E \) in \( E \). Therefore, if \( I(\cdot) \) is a \( \sigma \)-additive set function defined on a semi-ring \( E \) then \( I(\cdot) \) can be extended to the lattice \( E \) satisfying (1.2-a) and (1.2-b). However, condition (1.2-c) is the continuity from above, i.e., the \( \sigma \)-additivity assumption. Note that for any function \( \varphi \) in \( E \), the pre-image \( \varphi^{-1}(a) \cap \{ \varphi \neq 0 \} \) belongs to the ring generated by \( E \), i.e., the class of disjoint unions of set in \( E \). For convenience, we restate these assertions

**Lemma 1.3.** Let \( E \) be a semi-ring of \( 2^X \) which is identify to the class of characteristic functions \( 1_E \) with \( E \) in \( E \). If \( I : E \to [0, \infty) \) is a \( \sigma \)-additive functional then \( I \) can be extended to the vector lattice space \( E \) of finite linear combination of characteristic functions in \( E \) and \( I \) satisfies (1.2), i.e., \( I \) becomes a pre-integral.

**Proof.** First remark that the \( \sigma \)-additivity of \( I \) considered as either a (finite!) measure on the semi-ring \( E \), i.e., (a) \( I(1_0) = 0 \) (b) \( I(1_{A+B}) = I(1_A)+I(1_B) \) and (b) \( \lim_n I(1_{A_n}) = 0 \) for any decreasing sequence \( \{ A_n \} \) with \( \bigcap_n A_n = \emptyset \), is equivalent to the shorter condition \( I\left( \sum_k 1_{E_k} \right) = \sum_k I(1_{E_k}) \), for any sequence \( \{ E_n \} \) of disjoint sets in \( E \) such that \( E = \sum_k 1_{E_k} \) belongs also to \( E \).

The extension of \( I \) to \( E \) (the class of linear combinations of characteristic functions \( 1_E \) with \( E \) in \( E \)) is clearly accomplished by linearity, i.e., if \( \varphi = \sum_{i=1}^{n} a_i 1_{A_i} \) with \( A_i \cap A_j = \emptyset \), when \( i \neq j \), and \( A_i \) in \( E \) then we have \( I(\varphi) = [\text{Preliminary}] \) Menaldi November 11, 2016
\[ \sum_{i=1}^{n} a_i I(1_{A_i}) \], which is a good definition in view of the linearity of \( I \) on the initial semi-ring.

To check property (1.2-c), we consider a decreasing sequence \( \{ \varphi_n \} \) of functions in \( E \) satisfying \( \lim_n \varphi_n(x) = 0 \) for every \( x \). For every \( \varepsilon > 0 \), the set \( E_{\varepsilon,n} = \{ \varphi_n(x) > \varepsilon \} \) is a finite disjoint union of set in the semi-ring \( \mathcal{E} \) satisfying \( E_{\varepsilon,n} \supset E_{\varepsilon,n+1} \) and \( \cap_n E_{\varepsilon,n} = \emptyset \). Therefore, the \( \sigma \)-additivity of \( I \) implies \( \lim_n I(1_{E_{\varepsilon,n}}) = 0 \). Since
\[
I(\varphi_n) \leq ||\varphi|| I(1_{E_{\varepsilon,n}}) + \varepsilon I(1_A), \quad A = \{ x : \varphi_1(x) \neq 0 \},
\]
we deduce \( \lim_n I(\varphi_n) = 0 \).

It is clear that a typical application of the above result is the case of the (Lebesgue) measure defined on the semi-ring of all \( d \)-dimensional intervals \( (a,b] = (a_1,b_1] \times \cdots \times (a_d,b_d] \), \( I(1_{(a,b]}) = (b_1 - a_1) \cdots (b_d - a_d) \). This also agree with the Riemann integral.

\textbf{Remark 1.4.} Let us consider the case of the Lebesgue measure in \( \mathbb{R} \), i.e., \( E \) is the vector lattice space of step functions \( \varphi = \sum_{i=1}^{n} \alpha_i 1_{(a_i,b_i]} \) and the functional \( I \) is defined by linearity \( I(\varphi) = \sum_{i=1}^{n} \alpha_i I(1_{(a_i,b_i]}) \) with \( I(1_{(a_i,b_i]}) = (b_i - a_i) \) being the length as above. It is simple to show that \( I \) is well defined, monotone and linear on \( E \). To check the continuity condition (1.2-c), let \( \{ \varphi_n \} \) be a decreasing sequence of functions in \( E \) with \( \lim \varphi_n(x) = 0 \) for every \( x \). Since each step function \( \varphi_n \) is discontinuous (actually has a jump) only at a finite number points, for any \( \varepsilon > 0 \) we can find a sequence of open intervals \( \{(c_k,d_k)\} \) covering all points of discontinuity for any \( \varphi_n \) and such that \( \sum_k (d_k - c_k) < \varepsilon \). Now, the argument in Remark 1.1 proves that there exists another sequence of open intervals \( \{(e_k,f_k)\} \) covering all points of continuity, i.e., the union of the two sequences of open intervals cover the compact interval where \( \varphi_1(x) \neq 0 \). Thus, we can extract a finite subcover, and we repeat the argument in Lemma 1.3, to see that \( I(\varphi_n) \downarrow 0 \).

\textbf{Exercise 1.1.} Fill in the details of the previous Remark 1.4. Moreover, consider in great detail the case of the Lebesgue-Stieltjes measures in \( \mathbb{R} \), i.e., \( I(1_{(a,b]}) = F(b) - F(a) \) for a given right-continuous increasing real-valued function. Furthermore, discuss the changes necessary to extend the arguments used for the length of a interval to the case of the hyper-volume of a \( d \)-dimensional interval (i.e., Lebesgue measure).

\textbf{Exercise 1.2.} Let \( E_i \) be a vector lattice of functions on a (Hausdorff) space \( X_i \), for \( i = 1, 2 \). Denote by \( E_1 \otimes E_2 \) the vector lattice generated by functions of the form \( \varphi_1(x_1)\varphi_2(x_2) \) with \( \varphi_i \) in \( E_i \), i.e., the smallest vector lattice containing the above class of functions. Verify that any element \( \varphi(x_1,x_2) \) in \( E_1 \otimes E_2 \) satisfies: for every fixed \( x_1 \), the function \( \varphi_{x_1} : x_2 \mapsto \varphi(x_1,x_2) \) belongs to \( E_2 \), and for every fixed \( x_2 \), the function \( \varphi_{x_2} : x_1 \mapsto \varphi(x_1,x_2) \) belongs to \( E_1 \).

Daniell integral is the procedure to extend the definition of the pre-integral \( I \) to a larger class of functions containing \( E \), i.e., based on the hypotheses (1.1) and (1.2), we need to redefine (1) null or negligible sets, (2) integrable functions and (3) measurable functions.
1.1.1 Null or Negligible Sets

A subset $N$ of $X$ is called a null or negligible set (with respect to the pre-integral $I$) if there exists an increasing (i.e., non-decreasing) sequence $\{\varphi_k\} \subset E$ and a constant $C$ such that (a) $\varphi_k(x) \uparrow +\infty$ for every $x$ in $N$ and (b) $I(\varphi_k) \leq C$, for every $k \geq 1$. Certainly, by considering the new sequence $\{a(\varphi_k - \varphi_1)\}$ for a fix $a > 0$, we deduce that a set $N$ is null if and only if for every $\varepsilon > 0$ there exists a nonnegative and increasing sequence $\{\varphi_k\} \subset E$ such that (a) $\varphi_k(x) \uparrow +\infty$ for every $x$ in $N$ and (b) $I(\varphi_k) \leq \varepsilon$, for every $k \geq 1$. As in previous sections, we say that a pointwise property is satisfied almost everywhere if it is true for every point except on a null set.

It is then clear that null sets are countable completes, this means that (i) every subset of a null set is a null set, and (ii) any countable union of null sets is a null set. Indeed, if $N = \bigcup_i N_i$ and $\{\varphi_{i,k}\}_k$ is a nonnegative and increasing sequence of elementary functions corresponding to $N_k$ and $\varepsilon = 2^{-i}$, then $\varphi_k = \sum_{i=1}^k \varphi_{i,k}$ yields a nonnegative and increasing sequence such that $\varphi(x) \to +\infty$, for every $x \in N$ and $I(\varphi_k) \leq 1$, for every $k \geq 1$. The linearity of $I$ yields $I(\varphi) = 0$ for the identically zero function $\varphi = 0$, for the converse we have

**Proposition 1.5.** Let $\{\psi_k\}$ be an increasing sequence in $E$. If $\lim_k \psi_k(x) \geq 0$, for every $x$ in $X \setminus N$, with $N$ a null set, then $\lim_k I(\psi_k) \geq 0$. Conversely, if $\lim_k I(\psi_k) = 0$ and $\lim_k \psi_k(x) \geq 0$, for every $x$ in $X$, then $\lim_k \psi_k(x) = 0$, for every $x$ in $X \setminus N$, with $N$ a null set.

**Proof.** Indeed, let $N$ be the null set where $\lim_k \psi_k(x) < 0$ and $\{\varphi_k\}$ be an increasing sequence of (nonnegative) elementary functions satisfying $\varphi_k(x) \uparrow +\infty$, for every $x$ in $N$ and $I(\varphi_k) \leq 1$, for every $k \geq 1$. Given an $\varepsilon > 0$, the sequence $\{\phi_k = (\psi_k^- - \varepsilon \varphi_k)^+\}$ satisfies $\phi_k(x) \downarrow 0$, for every $x$ and therefore $I(\phi_k) \to 0$. Since $\phi_k \geq \psi_k^- - \varepsilon \varphi_k$ we have $\psi_k \geq \psi_k^+ - \varepsilon \varphi_k - \phi_k$, which yields $I(\psi_k) \geq -\varepsilon - I(\phi_k)$, and the desired inequality follows.

The converse is easier, for an increasing sequence $\{\epsilon_k\}$ of positive numbers such that $\epsilon_k I(\psi_k)$ remains bounded (in $k$) and $\epsilon_k \to \infty$, we deduce that the increasing sequence $\{\varphi_k = \epsilon_k \psi_k^+\}$ in $E$ satisfies $\lim_k \varphi_k(x) = \infty$ for every $x$ where $\lim_k \psi_k(x) > 0$, while $I(\varphi_k) = \epsilon_k I(\psi_k^+)$ is bounded.

Note that the argument in Proposition 1.5 also shows that if $\lim_k I(\psi_k) = 0$ and $\lim_k \psi_k(x) \geq 0$, for every $x$ in $X \setminus N'$, and some null set $N'$, then $\lim_k \psi_k(x) = 0$, for every $x$ in $X \setminus N$, with $N \supset N'$ a null set.

Essentially, this result says that null sets can be ignored (or are negligible) for the pre-integral operation. For instance,

**Exercise 1.3.** Based on the technique of the previous Proposition 1.5, prove that if $N$ is a null set and $\psi$ and $\varphi$ are two functions in $E$ such that $\psi(x) = \varphi(x)$, for any $x$ in $X \setminus N$, then $I(\psi) = I(\varphi)$.

**Proposition 1.6.** Let $\{\psi_k\}$ be a sequence in $E$ which is pointwise decreasing to 0 outside of a null set $N$, i.e., $\varphi_{k+1}(x) \leq \varphi_k(x)$ and $\varphi_k(x) \downarrow 0$, for every $x$ in $X \setminus N$ and $k \geq 1$. Then we have $I(\varphi_k) \to 0$ as $k \to \infty$. 

Proof. Suppose \( \{ \psi_k \} \subset E \) and for some null set \( N \) we have \( \psi_{k+1}(x) \leq \psi_k(x) \) and \( \psi_k(x) \downarrow 0 \), for every \( x \in X \setminus N \). This implies that \( \psi_k^\leq = 0 \) and \( \min_{i \leq k} \psi_i = \psi_k \) outside of \( N \), for every \( k \geq 1 \).

First, choose a increasing (nonnegative) sequence \( \{ \varphi_n \} \subset E \) such that \( \varphi_n(x) \uparrow +\infty \), for every \( x \in N \), and \( I(\varphi_n) \leq 1 \), for any \( n \geq 1 \). Now, for every \( \varepsilon > 0 \) and any fix \( k \), the sequence \( \{ \phi_n = (\psi_k^- - \varepsilon \varphi_n)^+ \} \) satisfies \( \psi_k^-(x) \leq \phi_n(x) + \varepsilon \varphi_n(x) \) and \( \phi_n(x) \downarrow 0 \), for every \( x \). Hence \( I(\phi_n) \downarrow 0 \) and \( I(\psi_k^-) \leq \varepsilon \), which implies that \( I(\psi_k^-) = 0 \), i.e., \( I(\psi_k) \geq 0 \). This proves that \( \lim_k I(\psi_k) \geq 0 \).

To show the converse inequality, note that the sequence \( \{ \min_{i \leq k} \psi_i \} \) is decreasing everywhere and \( \min_{i \leq k} \psi_i(x) \uparrow 0 \), for any \( x \in X \setminus N \). Hence, apply Proposition 1.5 with the increasing sequence \( \{ -\min_{i \leq k} \psi_i \} \) to deduce \( \lim_k I(\min_{i \leq k} \psi_i) \leq 0 \).

Thus, our argument will be completed by proving that \( I(\min_{i \leq k} \psi_i) = I(\psi_k) \). To this purpose, for \( \varepsilon > 0 \) and any fix \( k \), note that the sequence \( \{ \phi_n = (\psi_k^- - \varepsilon \varphi_n)^+ \} \) satisfies \( \psi_k^- \leq \phi_n + \varepsilon \varphi_n \) and \( \phi_n(x) \downarrow 0 \), for every \( x \). Hence \( I(\phi_n) \downarrow 0 \) and \( I(\psi_k) - I(\min_{i \leq k} \psi_i) \leq \varepsilon \), i.e., \( I(\psi_k) \leq I(\min_{i \leq k} \psi_i) \) as desired.

To compare with the definition of null sets based on a measure, let us assume that there exits a semi-ring \( S \) of subsets of \( X \) generating \( E \), i.e., \( \varphi \in E \) if and only if there exists \( E_1, \ldots, E_n \in S \) and real numbers \( a_1, \ldots, a_n \) such that \( \varphi = \sum_{i=1}^n a_i \mathbb{1}_{E_i} \) and \( E_i \cap E_j = \emptyset \), when \( i \neq j \). Then, null sets (with respect to the pre-integral \( I \)) can be defined as follows: a set \( N \) is called a set of measure zero if for every \( \varepsilon > 0 \) there exists a sequence \( \{ E_k \} \) of sets in \( S \) such that (i) \( \bigcup_{k=1}^\infty E_k \supset N \) and (ii) \( \sum_{k=1}^\infty I(\mathbb{1}_{E_k}) < \varepsilon \).

Lemma 1.7. With the above notation, sets of measure zero and null sets are equivalent.

Proof. To check the equivalence, let \( N \) be a set of measure zero according to (i) and (ii), then for every \( \varepsilon = 2^{-n} \) we find a sequence \( \{ E_{n,k} \} \) such that

\[
N \subset \bigcup_{k=1}^\infty E_{n,k} \quad \text{and} \quad \sum_{k=1}^\infty I(\mathbb{1}_{E_{n,k}}) \leq 2^{-n}.
\]

Thus the functions \( \psi_n = n \sum_{i=1}^n \mathbb{1}_{E_{n,i}} \) belong to \( E \) and \( \psi_n(x) \geq n \), for every \( x \) in \( \bigcup_{i=1}^n E_{n,i} \). Hence, \( \{ \varphi_k = \sum_{n=1}^\infty \psi_n \} \) is an increasing sequence of functions in \( E \) such that \( \varphi_k(x) \rightarrow +\infty \) for every \( x \) in \( \bigcap_{n=1}^\infty \bigcup_{k=1}^\infty E_{n,k} \supset N \) and \( I(\varphi_k) \leq \sum_{n=1}^\infty n2^{-n} \).

For the converse, let \( N \) be a null set, i.e., a set such that there exists an increasing sequence \( \{ \varphi_k \} \) of functions in \( E \) and a constant \( C > 0 \) such that (a) \( \varphi_k(x) \uparrow +\infty \) for every \( x \) in \( N \) and (b) \( I(\varphi_k) \leq C \), for every \( k \geq 1 \). Given \( \varepsilon > 0 \), consider the sets \( A_{\varepsilon,k} = \{ x : \varphi_k(x) > \varepsilon^{-1}C \} \), for \( k \geq 1 \). Thus, \( A_{\varepsilon,k} \subset A_{\varepsilon,k+1} \) \( I(\mathbb{1}_{A_{\varepsilon,k}}) \leq \varepsilon \), for every \( k \geq 1 \), and \( N \subset \bigcup_{k=1}^\infty A_{\varepsilon,k} \). We conclude by writing each \( A_{\varepsilon,k} \) as a disjoint finite union of elements belonging to the semi-ring \( S \).
class $\mathcal{E}$ results a ring, but it is more complicate (without develop the tools of the following section) to prove that $I(1_{\mathcal{E}}) = \lim_n I(\psi_n)$, with $\psi_n = (n(\varphi - a)^+) \land 1$ is a $\sigma$-additive finite measure on the ring $\mathcal{E}$. Essentially, this would reduce a general vector lattice to the case where $\mathcal{E}$ are step functions as in Lemma 1.7. However, in view of what follows, this would be an almost nonsense exercise.

### 1.1.2 Integrable Functions

Every function $\varphi$ in $\mathcal{E}$ can be written in a unique form $\varphi = \varphi^+ - \varphi^-$, where $\varphi^+ = \max\{\varphi, 0\}$ and $\varphi^- = -\min\{\varphi, 0\}$ belong to the class $\mathcal{E}^+$ of nonnegative elementary functions. Sometimes it is convenient to use the notation $f \lor g = \max\{f, g\}$, $f \land g = \min\{f, g\}$, and $|f| = f^+ + f^-$. Denote by $\bar{\mathcal{E}}$ the semi-space of functions $f$ which are pointwise limit of an increasing sequence in $\mathcal{E}$. Note that $f : X \to (-\infty, +\infty]$ We define the integral of a function $f$ in $\bar{\mathcal{E}}$ as the increasing limit of the numerical sequence $\{I(\varphi_k)\}$, where $\{\varphi_k\}$ is an increasing sequence of elementary functions in $\mathcal{E}$ such that $\varphi_k(x) \uparrow f(x)$, for every $x$ in $X$. Certainly, the value of the limit $I(f) = \lim_k I(\varphi_k)$ may be infinite ($+\infty$), and we say that $f$ in $\bar{\mathcal{E}}$ is integrable if $I(f) < \infty$. To make this definition valid and compatible with the notion of null sets, we need to show

**Lemma 1.8.** If $\{\varphi_k\}$ and $\{\psi_k\}$ are two increasing sequences of functions belonging to $\mathcal{E}$ such that $\lim_k \varphi_k(x) \geq \lim_k \psi_k(x)$, for every $x$ in $X \setminus N$, for some null set $N$, then $\lim_k I(\varphi_k) \geq \lim_k I(\psi_k)$.

**Proof.** First, assume $N = \emptyset$ and consider the double sequence $\rho_{i,j} = \min\{\varphi_i, \psi_j\}$ of functions in $\mathcal{E}$. Then, keeping $j$ fix as $i \to \infty$, we have $\rho_{i,j} \uparrow \psi_j$ and the monotone continuity of the pre-integral implies $I(\rho_{i,j}) \uparrow I(\psi_j)$. Since $I(\varphi_i) \geq I(\rho_{i,j})$, we deduce $\lim_i I(\varphi_i) \geq I(\psi_j)$, for every $j$, which yields $\lim_k I(\varphi_k) \geq \lim_k I(\psi_k)$.

Next, let $N$ be the null set where $\lim_k \varphi_k(x) < \lim_k \psi_k(x)$ and $\{\alpha_k\}$ be an increasing sequence of (nonnegative) elementary functions satisfying $\alpha_k(x) \uparrow +\infty$, for every $x$ in $N$ and $I(\alpha_k) \leq 1$, for every $k \geq 1$. Given an $\varepsilon > 0$, we denote by $\varphi_k = \varphi_k + \varepsilon \alpha_k$ to deduce $\lim_k \varphi_k(x) \geq \lim_k \psi_k(x)$, for every $x$ in $X$, with $\lim_k |I(\varphi_k) - I(\psi_k)| \leq \varepsilon$. Thus, from the previous arguments when was $N = \emptyset$, we deduce $\lim_k I(\varphi_k) \geq \lim_k I(\psi_k)$, which implies $\lim_k I(\varphi_k) + \varepsilon \geq \lim_k I(\psi_k)$, and we conclude by sending $\varepsilon$ to 0. □

Essentially by definition, we deduce that if $f$ in $\bar{\mathcal{E}}$ is integrable then the $N = \{x : f(x) = \infty\}$ is a null set, and conversely, if $N$ is a null set then for every $\varepsilon > 0$ there exists a function $f = f_\varepsilon$ in $\bar{\mathcal{E}}$ such that $|I(f)| < \varepsilon$ and $f(x) = \infty$ for every $x$ in $N$. Now, as in measure theory, we say that a pointwise property holds *almost everywhere* if it holds for every point outside of a null (or negligible) set with respect to $I$.

**Proposition 1.9** (Daniell). A pre-integral $I$ on the lattice $\mathcal{E}$ satisfying (1.1) and (1.2) can be uniquely extended to $\bar{\mathcal{E}}$ as mentioned above, i.e.,

$$\text{if } \varphi_k \uparrow f \text{ with } \varphi_k \in \mathcal{E} \text{ then } I(\varphi_k) \uparrow I(f),$$
and the following properties hold:

(a) $-\infty < I(f) \leq +\infty$, for every $f \in \mathcal{E}$;

(b) if $f, g \in \mathcal{E}$ and $c \geq 0$ then $f + cg, f \wedge g, f \vee g \in \mathcal{E}$, $I(f + cg) = I(f) + cI(g)$ and $I(f) + I(g) = I(f \vee g) + I(f \wedge g)$;

(c) if $f, g \in \mathcal{E}$ and $f(x) \leq g(x)$ almost every $x$ then $I(f) \leq I(g)$;

(d) if $\{f_n\}$ is an increasing sequence of functions in $\mathcal{E}$ then $f = \lim_n f_n$ is in $\mathcal{E}$ and $\lim_n I(f_n) = I(f)$.

Moreover, if $f$ belongs to $\mathcal{E}$, $f \geq 0$ and $I(f) = 0$ then $f = 0$ almost everywhere.

**Proof.** It is clear that assertions (a) and (c) follow from Lemma 1.8. Similarly, by means of the linearity of $I$ over $\mathcal{E}$ and the equality $f + g = f \vee g + f \wedge g$, we establish (b).

Thus to check (d), suppose $f_n \uparrow f$ so that for any $n$ fixed, there exists an increasing sequence in $\mathcal{E}$ such that $\varphi_{n,k} \uparrow f_n$ as $k \to \infty$. Define $\psi_k = \max_{n \leq k} \varphi_{n,k}$ to have $\psi_k \leq \psi_{k+1}$ and $f_k \geq \psi_k \geq \varphi_{n,k}$, for every $k \geq n$. Hence, as $k \to \infty$ we have

$$f = \lim_k f_k \geq \lim_k \psi_k \geq \lim_k \varphi_{n,k} = f_n,$$

$$\lim_k I(f_k) \geq \lim_k I(\psi_k) \geq \lim_k I(\varphi_{n,k}) = I(f_n),$$

and later, as $n \to \infty$, we deduce

$$f = \lim_k \psi_k \quad \text{and} \quad I(f) = \lim_k I(\psi_k) = \lim_k I(f_k),$$

which completes the proof of part (d).

The last statement, follows from Proposition 1.5, i.e., if $f$ belongs to $\mathcal{E}$, $f \geq 0$ and $I(f) = 0$ then $f(x) = 0$ for every $x$ in $X \setminus N$ for some null set $N$. 

If a function $f$ in $\mathcal{E}$ satisfies $I(f) < \infty$ then $f$ is finite almost everywhere. Thus, our main interest is in the class $\mathcal{E}_1$ of $(-\infty, +\infty]$-valued functions $f$ in $\mathcal{E}$ with $I(f)$ finite, i.e., $f$ belongs to $\mathcal{E}_1$ if and only if there exists an increasing sequence of (elementary) functions $\{\varphi_k\}$ in $\mathcal{E}$ such that $f(x) = \lim_k \varphi_k(x)$, for every $x$ in $X$, and $I(f) = \lim_k I(\varphi_k) < \infty$. In other words, using telescoping series, $f$ belongs to $\mathcal{E}_1$ if and only if $f = \varphi + \sum_n \varphi_n$ pointwise with $\varphi, \varphi_n$ in $\mathcal{E}$ and $\varphi_n \geq 0$, and the numerical series $I(\varphi) + \sum_n I(\varphi_n) < \infty$ converges to a finite value denotes by $I(f)$.

Similarly, we consider the class $\mathcal{E}_1^+$ of nonnegative functions in $\mathcal{E}_1$, i.e., $f$ belongs to $\mathcal{E}_1^+$ if and only if there exists an increasing sequence of (elementary) functions $\{\varphi_k\}$ in $\mathcal{E}$ such that $f(x) = \lim_k \varphi_k(x) \geq 0$, for every $x$ in $X$, and $I(f) = \lim_k I(\varphi_k) < \infty$. Actually, using telescoping series, $f$ belongs to $\mathcal{E}_1^+$ if and only if $f = \sum_n \varphi_n$ pointwise with $\varphi_n$ in $\mathcal{E}$ and $\varphi_n \geq 0$, and the numerical series $\sum_n I(\varphi_n) < \infty$ converges to a finite value denotes by $I(f)$. 


Since $E$ is a vector lattice we deduce that $\bar{E}_1$ and $\bar{E}_1^+$ are semi-vector lattices. Moreover, $f$ belongs to $\bar{E}_1$ if and only if there exits $\varphi$ in $E$ and $g$ in $\bar{E}_1^+$ such that $f = \varphi + g$, i.e., $\bar{E}_1 = E + \bar{E}_1^+$.

**Definition 1.10.** The class $L$ of integrable functions is the vector space of (extended) real-valued functions which are almost everywhere equal to a difference of two functions in $\bar{E}_1$, i.e., $f$ belongs to $L$ if and only if $f(x) = g(x) - h(x)$ for every $x$ in $X \setminus N$, for some null set $N$ and some functions $g$ and $h$ in $\bar{E}_1$. Moreover, by linearity, we set (uniquely) $I(f) = I(g) - I(h)$.

A priori, any two functions $g$ and $h$ in $\bar{E}$ have extended value, and then the difference $g - h$ may not be defined. From the definition of null set, we deduce that any function $f$ in $\bar{E}$ with $I(f) < \infty$ has real values outside of a null set, in particular, for any two functions $g$ and $h$ in $\bar{E}_1$, the difference $g - h$ is defined and take real values almost everywhere.

To check that $L$ is vector space, we remark that $cf = (cg) - (ch)$ if $c \geq 0$ and $cf = ((-c)h) - ((-c)g)$ if $c \leq 0$. Next, to verify that $L$ is a lattice we can use the relations $g \lor h = (g + h - |g - h|)/2$ and $g \land h = (g + h + |g - h|)/2$ (which yield $|f| = |g - h| = g \lor h - g \land h$) and (b) of Proposition 1.9. A posteriori, we write $f = f^+ - f^-$ with $f^\pm$ in $L$ and we have $I(f) = I(f^+) - I(f^-)$. It is clear that if $f = 0$ almost everywhere then $I(f) = 0$, but the converse takes more work.

Because $\bar{E}_1 = E + \bar{E}_1^+$, we may say that $f$ belongs to $L$ if and only if $f(x) = \varphi + g^+(x) - h^+(x)$ for every $x$ in $X \setminus N$, for some null set $N$ and some functions $g^+$ and $h^+$ in $\bar{E}_1^+$, with $I(f) = I(\varphi) + I(g^+) - I(h^+)$. Actually, we have

**Proposition 1.11.** Recall that $\bar{E}_1^+$ is the class of $[0, +\infty]$-valued functions $f$ in $\bar{E}$ with $I(f)$ finite. Any integrable function $f$ (i.e., $f$ in $L$) can be written as the different of two functions in $\bar{E}_1^+$.

**Proof.** Recall that if $g \geq 0$ and $\varphi_k \uparrow g$ then $\varphi_k^+ \uparrow g$, therefore we deduce that $g$ belongs to $\bar{E}_1^+$ if and only if there exists an increasing sequence of nonnegative (elementary) functions $\{\varphi_k\}$ in $E$ such that $g(x) = \lim_k \varphi_k(x)$ for every $x$ in $X$ with and $I(g) = \lim_k I(\varphi_k) < \infty$.

If $f$ is integrable, by definition we have $f = g - h$ a.e., with $g$ and $h$ in $\bar{E}_1$, and so $g = \lim_n g_n$ and $h = \lim_n h_n$ for some increasing sequences $\{g_n\}$ and $\{h_n\}$ in $E$ with $I(g) + I(h) < \infty$. Hence $f = (g + g_1^- + h_1^-) - (h + h_1^- + g_1^-)$ a.e, $(g_n - g_1 + g_1^+ + h_1^-) = (g_n + g_1^+ + h_1^-)$ and $(h_n - h_1 + h_1^+ + g_1^-) = (h_n + h_1^- + g_1^-)$ yield two increasing sequences of nonnegative (elementary) functions in $E$, which prove the desired result. In short and formally, we can write either $L = \bar{E}_1^+ - \bar{E}_1^+$ a.e., or $L = \bar{E}_1 - \bar{E}_1$ a.e.

This argument shows that without any changes, we could define the class $L$ of integrable functions as the vector space of (extended) real-valued functions which are almost everywhere equal to a difference of two functions in $\bar{E}_1^+$, instead of two functions in $\bar{E}_1$.

Hence, any integrable function $f$ is an almost everywhere limit of a sequence $\{f_n = g_n - h_n\}$ of elementary functions (i.e., in $E$) where $\{g_n\}$ and $\{h_n\}$ are increasing sequences of (nonnegative) functions in $E$ such that the numerical
sequence \( \{I(g_n) + I(h_n)\} \) is bounded. Note that \( |f_n| \leq g + h \) for every \( n \geq 1 \), and \( g + h \) belongs to \( \tilde{E} \) and \( I(g + h) < \infty \), i.e., \( g + h \) belongs to \( \tilde{E}_1(\tilde{E}_1^+) \).

Integrable functions (or functions in class \( L \)) are defined everywhere, but they are considered as defined almost everywhere. Properly speaking, we form classes of equivalence as we consider equality almost everywhere, essentially, any pointwise property applied to functions in the class \( L \) is considered almost everywhere, e.g., an integrable function \( f \) is nonnegative if indeed \( f(x) \geq 0 \) for every \( x \) in \( X \setminus N \) for some null set \( N \). Moreover, a nonnegative function \( f \) in the class \( L \) is not necessarily equal almost everywhere to a function in \( \tilde{E}_1^+ \), actually we have

**Lemma 1.12.** If \( f \) is a nonnegative integrable then for every \( \varepsilon > 0 \) there exist functions \( g_\varepsilon \) and \( h_\varepsilon \) in \( \tilde{E}_1^+ \) such that \( f = g_\varepsilon - h_\varepsilon \) almost everywhere, \( I(g_\varepsilon) < \infty \) and \( 0 \leq I(h_\varepsilon) < \varepsilon \).

**Proof.** Indeed, we can express \( f = g - h \) a.e., with \( g \) and \( h \) in \( \tilde{E}_1^+ \), i.e., there exist increasing sequences \( \{g_n\} \) and \( \{h_k\} \) in \( E \) such that \( g_n \uparrow g, \ h_k \uparrow h, \ I(g_n) \uparrow I(g) < \infty \) and \( I(h_k) \uparrow I(h) < \infty \). Thus, we write \( f = (g - h_k) - (h - h_k), \) a.e., where \( (h - h_k) \geq 0 \), \( I(h - h_k) \downarrow 0 \), and the function \( (h - h_k) \) belongs to \( \tilde{E}_1^+ \), but a priori, the function \( (g - h_k) \) belongs only to \( \tilde{E}_1 \).

However, \( g(x) - h_k(x) \geq g(x) - h(x) = f(x) \geq 0 \), for every \( x \) in \( X \setminus N \), where \( N = \{x \in X : f(x) < 0\} \) is a null set. Applying the definition of null set, for every \( \varepsilon > 0 \) there exists an increasing sequence (of nonnegative) functions \( \{\varphi_n\} \) in \( E \) with \( 2I(\varphi_n) < \varepsilon \), for every \( n \), and such that \( \varphi_n(x) = \lim_n \varphi_n(x) = +\infty \), for every \( x \) in \( N \). This implies that \( g - h_k + \varphi \geq 0 \), which means that \( g - h_k + \varphi \) belongs to \( \tilde{E}_1^+ \).

Hence, we can choose \( g_\varepsilon = g - h_k + \varphi \) and \( h_\varepsilon = h - h_k + \varphi \) with some \( k \) sufficiently large, to obtain the desired result.

At this point we are able to prove the three basic convergence results, namely,

**Theorem 1.13** (monotone). Let \( \{f_n\} \) be a sequence in \( L \) satisfying \( f_n \geq 0 \) almost everywhere, for any \( n \geq 1 \). The pointwise almost everywhere defined series \( f = \sum_n f_n \) belongs to \( L \) if and only if \( \sum_n I(f_n) < \infty \). In this case \( I(f) = \sum_n I(f_n) \).

**Proof.** Indeed, each \( f_n \) is written as \( g_n - h_n \) with \( g_n \) and \( h_n \) in \( \tilde{E}_1^+ \) and \( 0 \leq I(h_n) < 2^{-n} \). Thus, proving that for \( g = \sum_n g_n \) and \( h = \sum_n h_n \) the property (d) of Proposition 1.9 yields the equality \( I(g) = \sum_n I(g_n) \) and \( I(h) = \sum_n I(h_n) \), with \( \sum_n I(h_n) \leq 1 \) the argument is completed.

To this purpose, because \( h_n \) is in \( \tilde{E}_n \), there is a sequence \( \{\varphi_{n,k}\} \subset E \) such that \( h_n = \sum_k \varphi_{n,k} \) pointwise, \( \varphi_n \geq 0 \) and \( I(h_n) = \sum_k I(\varphi_{n,k}) < 2^{-n} \), for each fixed \( n \). This implies that \( h = \sum_n \varphi_{n,k} \) and \( I(h) = \sum_n I(\varphi_{n,k}) = \sum_n I(h_n) \leq 1 \).

Similarly, since \( g_n \) belongs to \( \tilde{E}_n \), the representation \( g_n = \sum_k \psi_{n,k} \) (with \( \psi_{n,k} \geq 0 \) and \( \psi_n \) in \( E \)) implies that

\[
I(g) = \sum_{n,k} I(\psi_{n,k}) = \sum_n I(g_n) \leq 1 + \sum_n I(f_n) < \infty.
\]
Remark that if $\sum_n I(f_n) = \infty$ then the inequality $g_n \geq f_n$ implies that $I(g) = \sum_n I(g_n) \geq \sum_n I(f_n) = \infty$, and the limiting equality $I(f) = \sum_n I(f_n)$ holds true, but $f$ is almost everywhere equal to the function $g - h$, with $g \geq 0$ in $\mathcal{E}$ and $h$ in $\mathcal{E}_1^+$. □

Sometimes, the almost everywhere difference of two functions, one in $\mathcal{E}$ and another in $\mathcal{E}_1$ is called quasi-integrable, i.e., $f = g - h$ almost everywhere, with either $g$ in $\mathcal{E}$ and $h$ in $\mathcal{E}_1$ or $g$ in $\mathcal{E}_1$ and $h$ in $\mathcal{E}$; and we set $I(f) = I(g) - I(h)$, which may be a positive or negative infinite value. For instance, if $\sum_n I(f_n) = +\infty$ in the above Theorem 1.13 then $f = \sum_n f_n$ is an quasi-integrable function.

• **Remark 1.14.** From the constructions of the classes $\mathcal{E}_1$ and $\mathcal{L}$, it should be clear that for any integrable function $f$ (i.e., any element $f$ in the class $\mathcal{L}$) there exists a sequence $\{\varphi_n\}$ in $\mathcal{E}$ such that $\varphi_n \to f$ pointwise, except in a null set, and $I(|f - \varphi_n|) \to 0$. Now, the converse is ensured by the monotone convergence, i.e., a function $f$ belongs to $\mathcal{L}$ if and only if $f = \sum_n \varphi_n$ pointwise almost everywhere, with $\varphi_n$ in $\mathcal{E}$ and $\sum_n I(|\varphi_n|) < \infty$. □

• **Remark 1.15.** The above convergence can be restated in term of monotone (increasing or decreasing, almost everywhere) sequences, e.g., if $\{f_n\}$ is a sequence of integrable functions satisfying $f_{n+1} \geq f_n$ almost everywhere for every $n \geq 1$ then $f = \lim_n f_n$ is integrable if and only if $\lim_n I(f_n) < \infty$ and in this case $I(f) = \lim_n I(f_n)$. □

• **Remark 1.16.** By means of the monotone convergence we can show that for any nonnegative integrable function $f$ we have $I(f) = 0$ if and only if $f = 0$ almost everywhere. Indeed, if $f \geq 0$ and $I(f) = 0$ then the increasing sequence $\{f_k = kf : k \geq 1\}$ of integrable functions satisfies $I(f_k) = kI(f) = 0$ and therefore the limit $\lim_k f_k$ is an integrable function, which must be finite almost everywhere, i.e., $f = 0$ almost everywhere. □

**Theorem 1.17** (liminf). Let $\{f_n\}$ be a sequence in $\mathcal{L}$ satisfying $f_n \geq 0$ almost everywhere, for any $n \geq 1$. The almost everywhere pointwise inferior limit function $\liminf_n f_n$ belongs to $\mathcal{L}$ if and only if $\liminf_n I(f_n) < \infty$. In this case $I(\liminf_n f_n) = \liminf_n I(f_n)$.

**Proof.** Because $\mathcal{L}$ is a lattice, the functions $\min_{n \leq k \leq m} \{f_k\}$ are integrable. Now, we need to apply the previous Theorem 1.13 (and Remark 1.15) twice. First, for a fixed $n$ and as $k$ goes to $\infty$ we deduce that the decreasing limit functions $g_n = \min_{k \geq n} \{f_k\}$ are also integrable, and that $I(g_n) \leq I(f_n)$. Next, the increasing sequence $\{g_n\}$ with limit $\lim_n g_n = \liminf_n f_n$ yields the desired convergence. □

**Theorem 1.18** (dominate). Let $\{f_n\}$ be a sequence in $\mathcal{L}$ satisfying $|f_n| \leq g$ almost everywhere, for any $n \geq 1$ and some $g$ in $\mathcal{L}$. If the limit function $f = \lim_n f_n$ exists almost everywhere then it belongs to $\mathcal{L}$ and $I(f) = \lim_n I(f_n)$. □

**Proof.** It suffices to use previous Theorem 1.17 with the sequences $g \pm f_n$ to deduce that

$$I(\liminf_n f_n) \leq \liminf_n I(f_n) \leq \limsup_n I(f_n) \leq I(\limsup_n f_n),$$
after simplifying the finite term $I(g)$. Since $\liminf_n f_n = \limsup_n f_n$ almost everywhere and $|I(f_n)| \leq I(g) < \infty$, we conclude.

As mentioned early, an element $f$ of the class of integrable functions $L$ is regarded as a real-valued function defined almost everywhere, so that $f$ can be re-defined to be “anything” outside of a null set. Sometimes, it may be convenient to have $f$ defined everywhere in a proper sense, in this case, $f$ becomes a function with extended real-values, i.e., valued in $[-\infty, +\infty]$. To this purpose, some more spaces are introduced.

**Definition 1.19.** Let $\tilde{E}_1$ be the class of extended real-valued functions which are the limit of a non-increasing sequence in $\tilde{E}$ with bounded integrals, i.e., $f$ belongs to $\tilde{E}_1$ if and only if there exists a double sequence $\{\varphi_{k,n}\}$ in $E$ such that (a) $\varphi_{k,n}(x) \uparrow f_n(x)$ as $k \to \infty$, for every $n \geq 1$, (b) $f_n(x) \downarrow f(x)$ as $n \to \infty$, and (c) $|I(\varphi_{k,n})| \leq C$, for every $k, n$ and some constant $C$.

If the condition (c) on bounded integrals is eliminated then the class of extended real-valued functions satisfying (a) and (b) could be called $\tilde{E}$. The integral $I$ has been extended to $\tilde{E}$ by monotony, and could be extended to $\tilde{E}$ analogously. However, $\tilde{E}$ is not a vector space and because the integral $I$ will take values in $[-\infty, +\infty]$, the convergence theorems are not valid.

Note that elements in $\tilde{E}_1$ are functions defined everywhere (with values in $[-\infty, +\infty]$, but almost everywhere in $\mathbb{R}$) and in view of the monotone convergence Theorem 1.13 (and Remark 1.15) we deduce that $\tilde{E}_1 \subset L$ (in the sense that a function in $\tilde{E}_1$ can also be considered as an element of $L$) and $I(f) = \lim_n f_n$ and $I(f_n) = \lim_k I(\varphi_{k,n})$. Conversely, if $f$ is an element in the class $L$ then there exist increasing sequences $\{\phi_n\}$, $\{\psi_k\}$ and $\{\bar{\phi}_j\}$ in $E$ such that $\phi_n \uparrow g$, $\lim_n I(\phi_n) = I(g) < \infty$, $\psi_n \uparrow h$, $\lim_n I(\psi_n) = I(h) < \infty$, $\lim_j I(\bar{\phi}_j) < \infty$, and the set $N = \{x \in X : \phi_j(x) \uparrow +\infty\}$ is a null set containing every $x$ such that $g(x) = \infty$ or $h(x) = \infty$. Hence, the double sequence $\{\varphi_{n,k}\}$, $\varphi_{n,k} = \phi_n + \bar{\phi}_j - \psi_k - \bar{\phi}_k$ of functions in $E$ is increasing in $n$, decreasing in $k$ and satisfies $\lim_n \varphi_{n,k}(x) = \infty$ only for $x$ in $N$, and the iterate limit $\lim_k \lim_n \varphi_{n,k}(x) = f(x)$ for every $x$ outside of the null set $N$, i.e., for every element in $L$ is equal almost everywhere to some element in $\tilde{E}_1$. Hence $\tilde{E}_1 = L$ in the previous sense. In other words, a function $f$ belongs to $L$ if and only if there exists a sequence $\{\varphi_n\}$ of functions in $E$ satisfying $\sum_n I(|\varphi_n|) < \infty$, and $f(x) = \sum_{n} \varphi_n(x)$ whenever the series $\sum_{n} |\varphi_n(x)| < \infty$ (which implies that $\{x \in X : \sum_n |\varphi_n(x)| = \infty\}$ is a null set).

For any increasing sequence $\{\varphi_k\}$ of nonnegative elementary functions (i.e., in $E$ with $\varphi_k \geq 0$) such that $I(\varphi_k) \leq C < \infty$ and $N = \{x : \lim_k \varphi_k(x) = \varphi_\infty(x) = \infty\}$, and for any strictly decreasing numerical sequence $\{\varepsilon_n\}$ converging to 0, we can define the double sequence $\psi_{k,n} = 1 \wedge (\varepsilon_n \varphi_k)$, which is increasing in $k$ to the function $f_n = 1 \wedge (\varepsilon_n \varphi_\infty)$ in $\tilde{E}$ and $f_n$ decreases to the function $1_N$. Thus, if $A$ is a null set then there exists another null set $N$ such that $A \subset N$ and $1_N$ belongs to $\tilde{E}_1$. However, $1_A$ is not necessarily a function belonging to $\tilde{E}_1$. The completeness property is built into the definition of null sets and passed to the equivalence classes of $L$. 
Chapter 1. Abstract Integration

Therefore, we find again that a real-valued (or extended real-valued finite almost everywhere) function \( f \) is integrable if and only if \( f \) is the limit in almost every point \( x \) of the difference of two increasing sequences \( \{ \varphi_k \} \) and \( \{ \psi_n \} \) of nonnegative elementary functions where the numerical sequences \( \{ I(\varphi_k) \} \) and \( \{ I(\psi_n) \} \) are bounded. However, a priori, an arbitrary nonnegative integrable function \( f \) is not necessarily the limit in almost every point \( x \) of an increasing sequence of nonnegative elementary functions, see Lemma 1.12.

In the construction of the above class \( L \) of integrable functions, we made intensive use of null sets. Actually, we can (temporary) by pass negligible sets with the use of the following alternative construction. Assuming the pre-integral \( I \) has been extended to \( \bar{E} \) (as in Proposition 1.9, replacing almost everywhere statements with everywhere statements), for every real-valued function, we define the upper integral \( I \) and the lower integral \( \bar{I} \) by means of

\[
\bar{I}(f) = \inf \{ I(g) : g \geq f, g \in \bar{E} \} \quad \text{and} \quad I(f) = \sup \{ I(g) : g \leq f, -g \in \bar{E} \},
\]

(1.3)

under the convention that the inf (or sup) is equal to \( +\infty \) (or \( -\infty \)) if there are no function \( g \) satisfying the requested condition. Note that \( I(f) = -\bar{I}(-f) \), and \( f \) could have extended real-values with without any changes in the above setting.

**Exercise 1.4.** Prove that the following properties hold for the upper and the lower integrals: (a) \( \bar{I}(f + g) \leq \bar{I}(f) + \bar{I}(g) \); (b) \( \bar{I}(cf) = c\bar{I}(f) \), for any constant \( c \geq 0 \); (c) if \( f \leq g \) then \( \bar{I}(f) \leq \bar{I}(g) \) and \( I(f) \leq I(g) \); (d) \( I(f) \leq \bar{I}(f) \) for any \( f \), and \( \bar{I}(g) = \bar{I}(g) = I(g) \) for \( g \) in \( \bar{E} \). Moreover, if \( \{ f_k \} \) is a sequence of nonnegative functions and \( f = \sum_k f_k \) then \( \bar{I}(f) \leq \sum_k \bar{I}(f_k) \). \( \square \)

In this context, a null function \( h \) is a function satisfying \( \bar{I}(|h|) = 0 \), which is necessarily integrable. Thus, a proper null set \( N \) is defined as a subset of \( X \) such that \( 1_N \) is a null function, and to have the completeness, a null set is a subset of proper null set. Certainly, this turn out to be equivalent to definition of the previous section.

Therefore, the space \( \tilde{L} \) of integrable functions is defined as functions \( f \) with real values such that \( \bar{I}(f) = I(f) \) is a real number (i.e., finite). It is not hard to show that \( \tilde{L} \) is a vector lattice. Moreover, based on the properties given in Proposition 1.9, particularly (d), we obtain (as in previous sections) the convergence theorems for integrals (namely, the monotone, the dominated and the lim inf convergence theorems) within the spaces \( \tilde{L} \) (as accomplished early on the class \( L \)). Furthermore, we have

**Lemma 1.20.** With the previous notation \( L = \tilde{L} \) and the extension of \( I \) to \( L \) is unique. Moreover, \( f \) belongs to \( L \) if and only if \( f = g \) almost everywhere for some \( g \) in \( \bar{E}_1 \), i.e., \( g \) is a limit of a non-increasing sequence in \( \bar{E} \) with bounded integrals (see Definition 1.19).

**Proof.** Because functions in \( L \) are technically class of equivalence, the equality \( L = \tilde{L} \) actually means that if \( f \) belongs to \( L \) then there is an \( g \) in \( \tilde{L} \) such that...
f = g a.e., and if g belongs to \( \tilde{L} \) then the class of equivalence \( f \) determinate by g (i.e., all functions which are equal to g almost everywhere) belongs to \( L \).

Next we point out that the concept of null set is independent of the extension, but, the property that \( f = g \) almost everywhere implies \( I(f) = I(g) \) needs some consideration within the class \( \tilde{L} \). Indeed, we have seen that if \( N \) is a null set then for every \( \varepsilon > 0 \) there exists a function \( h = h_\varepsilon \geq 0 \) in \( \tilde{E} \) satisfying \( h(x) = \infty \) for every \( x \in N \) and \( I(h) < \varepsilon \). Therefore, if \( f(x) \leq g(x) \) for any \( x \) outside of a null set \( N \) then \( f \leq g + h \) everywhere, with \( g + h \) in \( \tilde{E} \). Hence the upper and lower integrals could be defined as

\[
\overline{I}(f) = \inf \{I(g) : g \geq f, \text{ almost everywhere, } g \in \tilde{E} \}
\]

and \( I(f) = -\overline{I}(-f) \).

Therefore, if \( f \) belongs to \( L \) then \( f = g - h \leq g - h_n \), almost everywhere, with \( g \) and \( h \) in \( \tilde{E} \) and \( \{h_n\} \) an increasing sequence in \( E \), i.e., \( g - h_n \) belongs to \( \tilde{E} \). Hence \( \overline{I}(f) \leq I(f) \). Similarly, we obtain \( I(f) \geq I(f) \), and we deduce that \( f \) belongs to \( \tilde{L} \).

Conversely, if \( f \) is any function with \( |\overline{I}(f)| < \infty \), then there exists a sequence \( \{g_n\} \) of function in \( \tilde{E} \) such that \( f \leq g_n \) and \( I(g_n) \leq \overline{I}(f) + 1/n \). Thus, the decreasing sequence \( \{f_n = g_1 \land \cdots \land g_n\} \) of functions in \( \tilde{E} \) satisfies \( I(f_n) \leq \overline{I}(f) + 1/n \) and so \( f^* = \lim_n f_n \) belongs to \( \tilde{E}_1 \), \( f^* \geq f \) and \( \overline{I}(f) = I(f^*) \). Similarly, if \( |I(f)| < \infty \) then there exists a function \( f_* \) such that \(-f_* \) belongs to \( \tilde{E}_1 \), \( f_* \leq f \) and \( I(f) = I(f_*) \).

Hence, if \( f \) belongs to \( L \) then the function \( f^* - f_* \geq 0 \) and \( I(f^* - f_*) = 0 \), and Remark 1.16 shows that \( f^* - f_* = 0 \) almost everywhere, i.e., \( f \) is equal almost everywhere to a function (i.e., \( f^* \) or \( f_* \)) in \( \tilde{E}_1 \) and thus \( f \) belongs to \( \tilde{L} \).

The statement relative to unique extension of \( I \) means that if \( K \) is a vector lattice containing \( L \), and \( J : K \to \mathbb{R} \) satisfies (1.2) with \( K \) instead of \( E \) and \( J(\varphi) = I(\varphi) \) for every \( \varphi \) in \( E \) then \( J = I \) on \( L \). It clear that this property follows from the convergence theorems for integrals.

Summing up, starting from an integral \( I \) defined on \( E \) and satisfying (1.2) we construct an extension of \( I \) which is defined on \( L \subseteq E \) and enjoys the convergence theorems. Moreover, this extension is unique and repeating this extension argument on \( I \) as defined on \( L \) produces the same lattice vector space \( L \). Furthermore, elements in \( L \) are regarded as functions defined almost everywhere, but as defined everywhere, one may use elements in \( \tilde{E}_1 \) with values in the extended real-numbers, see Definitions 1.10 and 1.19.

Now, denote by \( E^+ = \{\varphi \in E : \varphi \geq 0\} \) and by \( \mathcal{E} \subseteq 2^X \) the ring generated by the class of sets \( E = \{x \in X : \varphi(x) > a\} \) for any \( \varphi \in E^+ \) and \( a > 0 \), i.e., \( A \in \mathcal{E} \) if and only if there exist \( \varphi_i \) and \( 0 < a_i < b_i, i = 1, \ldots, n \) such that \( A = \bigcup_{i=1}^n \{x \in X : a_i < \varphi_i(x) \leq b_i\} \). Remark that the ring \( \mathcal{E} \) of elementary sets may not be, a priori, an algebra. Note the use of Stone's assumption (1.1-c) in the following

**Lemma 1.21.** If \( E \) belongs to \( \mathcal{E} \) then \( 1_E \) is integrable. Moreover, the map \( \mu : \mathcal{E} \to \mathbb{R} \) defined by \( \mu(E) = I(1_E) \) is \( \sigma \)-additive.
Proof. It suffices to consider $E = \{x \in X : \varphi(x) > a\}$ for some $\varphi \in E^+$ and $a > 0$. Then, define $\varphi_n(x) = (n|\varphi(x)/a - 1|^+)^-1$ to see that $\varphi_n(x) \uparrow 1_E(x)$ for every $x$, which proves that $1_E$ belongs to $E$. Next, the inequality $a 1_E \leq \varphi$ implies $a I(1_E) \leq I(\varphi)$, i.e., $I(1_E) < \infty$.

Finally, the $\sigma$-additivity property follows from the linearity and monotone continuity of $I$.

For instance, for further details the reader may consult Neveu [94, Section II.7, pp. 57–65]) regarding the first construction $L$ or Royden [108, Chapter 13] regarding $\tilde{L}$. Also, see Taylor [122, Chapter 6, 281–323]. Anyway, based on the previous Lemma 1.20, we use $L$ as the space of integrable function, which may also be called the space of $I$-integrable functions.

As stated in Lemma 1.21, the arguments used to extend the integral $I$ initially defined on a vector lattice $E$ can be used to extend a $\sigma$-measure initially defined on a semi-ring (or ring) to the generated $\sigma$-ring.

Recall that a class of subsets of a set $X$ is called a lattice if it contains the empty set, and it is stable under the formation of finite intersections and finite unions. A typical example of a lattice is the family of compact sets of a locally compact (Hausdorff topological) space $X$, and certainly, this is related to Radon and inner measures.

If $K_0$ is a $\pi$-class of subsets $X$, i.e., it contains the empty set and it stable under the formation of finite intersections, then denote by $K$ the class of finite disjoint unions of elements in $K_0$ (which may be referred as a semi-lattice) and by $K^+$ the semi-space of all simple functions of the form $\varphi = \sum_{i=1}^n a_i 1_{K_i}$, with nonnegative constants $a_i$ and a finite sequence $\{K_i\}$ of disjoint sets in $K_0$. It is clear that the semi-space $K^+$ is a lattice (i.e., if $\varphi$ and $\psi$ belong to $K^+$ then $\varphi \vee \psi$ and $\varphi \wedge \psi$ also belong to $K^+$), but not necessarily a semi-vector space. Although, $K^+$ contains the zero-function and it has the properties (i) if $\varphi$ and $\psi$ belong to $K^+$ and $\varphi \wedge \psi = 0$ then $\varphi + \psi$ belongs to $K^+$, (ii) if $\varphi$ belongs to $K^+$ and $c$ is a nonnegative constant then $c \varphi$ belongs to $K^+$, (iii) if $\varphi$ and $\psi$ belong to $K^+$ then $\varphi \psi$ belongs to $K^+$, and (iv) even if positive constant functions does not necessarily belong to $K^+$, the Stone’s assumption (1.1-c) is satisfied, i.e., if $\varphi$ belongs to $K^+$ then $1 \wedge \varphi$ belongs to $K^+$.

A mapping $I : K^+ \to [0, \infty)$ is called (a) pre-linear if $\varphi$, $\psi$ and $\varphi + c \psi$ in $K^+$ with a nonnegative constant c imply $I(\varphi + c \psi) = I(\varphi) + c I(\psi)$, and (b) monotone if $\varphi \leq \psi$ both in $K^+$ implies $I(\varphi) \leq I(\psi)$. It is clear that it suffices to define $I(1_K)$ for any $K$ in the semi-lattice $K$ and to use the extension $I(\varphi) = \sum_{i=1}^n a_i I(1_{K_i})$ to complete the definition of $I$ on the lattice $K^+$.

The vector space $E$ spanned by $K^+$ is the space of all simple functions of the form $\varphi = \sum_{i=1}^n a_i 1_{E_i}$ with constants $a_i$ and sets $E_i$ in $E$, where $E$ is the ring generated by the lattice $K$. This vector space is indeed the vector lattice generated by $K^+$. The pre-linearity and monotonicity of $I$ on $K^+$ are certainly necessarily conditions, but in general, not sufficient to extend the definition of $I$ to a linear functional on the vector lattice space $E$.

As seen later, a sufficient assumption to accomplish this objective is given by the so-called $K$-tightness condition, namely, for any $A \subset B$ in $K$ and every
$\varepsilon > 0$ there exists $C$ in $\mathcal{K}$ such that $C \subset B \setminus A$ and $I(1_B) \leq \varepsilon + I(1_A) + I(1_C)$. Due to the monotonicity, the expression

$$I_*(1_A) = \sup \{ I(1_K) : 1_K \leq 1_A, K \in \mathcal{K} \}, \quad \forall A \subset X.$$  \hspace{1cm} (1.4)

yields $I_*(1_K) = I(1_K)$, for every $K$ in $\mathcal{K}$ and the $\mathcal{K}$-tightness condition becomes a linearity-type assumption, i.e., if $A \subset B$ are in $\mathcal{K}$ then $I(1_B) = I(1_A) + I_*(1_{B \setminus A})$, and since the class $\mathcal{K}$ is a semi-lattice, this is equivalent to the condition: if $A$ and $B$ are in $\mathcal{K}$ then $I(1_B) = I(1_{B \cap A}) + I_*(1_{B \setminus A})$. Note that $I_*$ takes values in $[0, +\infty]$, $I_*$ is monotone and super-additive, i.e.,

(a) if $0 \leq 1_A \leq 1_B$ then $I_*(1_A) \leq I_*(1_B)$, and (b) if $A \cap B = \emptyset$ then $I_*(1_A + 1_B) \geq I_*(1_A) + I_*(1_B)$.

**Proposition 1.22.** Let $\{\mathcal{K} \subset \mathcal{I} \}$ be a lattice of subset of $X$ satisfying $I(1_K) \leq I_*(1_{K \cap A}) + I_*(1_{K \setminus A})$ for every $K$ in $\mathcal{K}$. It is clear that $\mathcal{K} \subset \mathcal{A}$ and if $A$ belongs to $\mathcal{A}$ then given any subset $E$ of $X$, for any $K = \bigcup_{i=1}^n K_i \subset E$ with $K_i$ in $\mathcal{K}$, the super-additivity and monotony of $I_*$ imply that

$$\sum_{i=1}^n I(1_{K_i}) \leq \sum_{i=1}^n [I_*(1_{K_i \cap A}) + I_*(1_{K_i \setminus A})] \leq I_*(1_{K \cap A}) + I_*(1_{K \setminus A}) \leq I_*(1_{E \cap A}) + I_*(1_{E \setminus A}),$$

and taking supremum on $K \subset E$, $K$ in $\mathcal{K}$, we deduce that $A$ belongs to $\mathcal{A}$ if and only if $I_*(1_E) \leq I_*(1_{E \cap A}) + I_*(1_{E \setminus A})$ for every $E \subset X$.

Now, we are ready to show that $\mathcal{A}$ is an algebra and that $I_*$ is linear on characteristic functions of $A$. Indeed, first note that the condition on a set for belonging to the class $\mathcal{A}$ is symmetric, i.e., $E \cap A = E \setminus A^c$ and $E \setminus A = E \cap A^c$, which means that $\mathcal{A}$ is stable under the formation of complement. Next, for any $A, B \in \mathcal{A}$ and $E \subset \Omega$, the equality

$$(E \cap A \setminus B) \cup (E \cap A \cap B) \cup (E \cap A \cap B) = E \cap (A \cap B)^c$$

and the super-additivity of $I_*$ imply

$$I_*(1_E) = I_*(1_{E \cap A}) + I_*(1_{E \cap A \setminus B}) + I_*(1_{E \cap A \cap B}) \leq I_*(1_{E \cap (A \cap B)^c}) + I_*(1_{E \cap (A \cap B)^c}).$$

Hence $A \cap B \in \mathcal{A}$, i.e., the class $\mathcal{A}$ is an algebra. Moreover, if $A, B \in \mathcal{A}$ and $A \cap B = \emptyset$ then

$$I_*(1_{A \cup B}) = I_*(1_{(A \cup B) \cap A}) + I_*(1_{(A \cup B) \cap A^c}) = I_*(1_A) + I_*(1_B),$$
i.e., \( I_* \) is linear on characteristic functions of \( \mathcal{A} \).

The \( \mathcal{K} \)-tightness condition implies that \( \mathcal{K} \subset \mathcal{A} \), which yields that the ring \( \mathcal{E} \) is contained in \( \mathcal{A} \). Moreover, the class of subset \( A \subset X \) which are covered by a finite union of sets in \( \mathcal{K} \) is clearly a ring, and therefore, it contains the ring \( \mathcal{E} \).

Summing-up, the extension \( I_* \) takes finite values on characteristic functions of \( \mathcal{E} \) and \( I_* \) is linear on the vector lattice \( \mathcal{E} \).

The uniqueness of the extension is clear, if \( I \) is another linear extension of \( I \) on a vector space \( \overline{\mathcal{E}} \supset \mathcal{E} \) then the family of functions \( \varphi \) in \( \mathcal{E} \) such that \( I(\varphi) = I_*(\varphi) \) is a vector space containing \( \mathcal{K}^+ \), which yields \( I = I_* \) on \( \mathcal{E} \).

Now, if \( I \) is a pre-integral, i.e., it satisfies the monotone continuity (1.2-c), then Daniell Proposition 1.9 can be used to complete the integration arguments. The condition necessary for this extension reduces to the so-called \( \sigma \)-smoothness in \( \mathcal{K} \), namely, if \( \{K_n\} \) is a decreasing sequence of subsets in \( \mathcal{K} \) with \( \bigcap_n K_n = \emptyset \) then \( I(1_{K_n}) \rightarrow 0 \).

**Exercise 1.5.** With the previous notation, for any nonnegative simple function \( f = \sum_{i=1}^n a_i 1_{A_i} \geq 0 \) with \( \{A_i\} \) a finite sequence of disjoint subsets of \( X \), note that the expression

\[
I_*(f) = \sup \{ I(\varphi) : \varphi \leq f, \ \varphi \in \mathcal{K}^+ \},
\]

takes values in \([0, +\infty]\), and verify that (1) \( I_* \) is monotone, i.e., if \( 0 \leq f \leq g \) then \( I_*(f) \leq I_*(g) \), (2) super-additive, i.e., if \( f, g \geq 0 \) and \( f \wedge g = 0 \) then \( I_*(f + g) \geq I_*(f) + I_*(g) \), and (3) homogeneous, i.e., if \( c \geq 0 \) constant then \( I_*(cf) = cI_*(f) \), and that (4) \( I_*(\varphi) = I(\varphi) \), for every \( \varphi \) in \( \mathcal{K} \). Moreover, prove that the \( \mathcal{K} \)-tightness condition may be called \( \mathcal{K}^+ \)-tightness condition when written as (5) if \( \varphi \) and \( \psi \) are in \( \mathcal{K}^+ \) then \( I(\varphi) = I(\varphi \wedge \psi) + I_*(\varphi - \varphi \wedge \psi) \). Furthermore, show that (6) if \( f = \sum_{i=1}^n a_i 1_{A_i} \geq 0 \) with \( \{A_i\} \) a finite sequence disjoint measurable subsets of \( X \) (where measurability of a set \( A \) means that \( I(1_K) \leq I_*(1_{K \cap A}) + I_*(1_{K \setminus A}) \) for every \( K \) in \( \mathcal{K} \)) then \( I_*(f) = \sum_{i=1}^n a_i I_*(1_{A_i}) \). Finally, deduce that the unique linear extension of \( I \) could be given as the mapping \( \varphi \mapsto I_*(\varphi^+) - I_*(\varphi^-) \), with \( I_* \) given as above.

\[ \square \]

### 1.1.3 Measurable Functions

Certainly, integrable functions (or elements in \( L \)) are considered defined almost everywhere and taking real-values, and in view of Stone’s assumption (1.1-c) the function \( g \wedge 1 \) belongs to \( L \) for every nonnegative \( g \) in \( L \). Moreover, a function \( f \) belongs to \( L \) if and only if \( f = g - h \), where both \( g \) and \( h \) are pointwise almost everywhere increasing limit of sequences \( \{g_n\} \) and \( \{h_n\} \) of elementary functions (i.e., in the vector lattice \( \mathcal{E} \)) such that the numerical sequences \( \{I(g_n)\} \) and \( \{I(h_n)\} \) are bounded.

**Definition 1.23.** A function \( f \) from \( X \) into \( \mathbb{R} \) is called measurable if for every nonnegative functions \( \varphi \) and \( \psi \) in \( \mathcal{E} \) we have \((-\psi) \vee (f \wedge \varphi)\) in \( L \). The set of all measurable functions is denoted by \( \mathcal{M} \). Again, classes of equivalence are used implicitly.

\[ \square \]
The definition of integrable functions and the dominate convergence Theorem 1.18 shows that $f$ is measurable if and only if for every nonnegative function $g$ and $h$ in $L$ we have $(-h) \vee (f \wedge g)$ in $L$.

Since $(-h) \vee (f \wedge g) = f^+ \wedge g - f^- \wedge h$, it is clear that $f$ belongs to $M$ if and only if $f^+$ and $f^-$ belongs to $M$. Now, if $f_1, f_2$ and $g$ are nonnegative and $c$ is a positive constant then $(f_1 + f_2) \wedge g = f_1 \wedge (g - g \wedge f_2) + f_2 \wedge g,$ $(c f_1) \wedge g = c (f_1 \wedge g / c)$ and $(f_1 \vee f_2) \wedge g = (f_1 \wedge g) \vee (f_2 \wedge g)$, and because $L$ is a vector lattice, we deduce that $M$ is also a vector lattice containing all constant functions (recall functions in $M$ takes only real-values).

Let $\{f_n\}$ be a sequence such that $f_n(x) \to f(x)$ for almost every $x$ in $X$. First, if each $f_n$ is measurable then the Lebesgue dominate convergence implies that $f$ also belongs to $M$. This show that $M$ a vector lattice stable under the pointwise almost everywhere convergence

Conversely, if there is a sequence $\{g_k\}$ in $L$ such that $\sup_k g_k(x) > 0$, for every $x$ such that $f(x) \neq 0$, then $f$ is a pointwise almost everywhere limit of the sequence $\{f_n\} \subset L$, $f_n = (-h_n) \vee (f \wedge h_n)$ with $h_n(x) = n \max_{k \leq n} g_k(x)$. Note that in general, we cannot express a measurable function as a pointwise (almost everywhere) limit of a suitable sequence in the initial vector lattice $E$.

Yet a further extension, a function $f$ belongs $L^+$ if $f = \lim_k f_k$ (almost everywhere) for some non-decreasing (almost everywhere) sequence $\{f_k\}$ of nonnegative functions in $L$, and $I(f)$ is uniquely definite as the limit $\lim_k I(f_k)$, which may be infinite. Now, let $M^+$ be the semi-space of all $[0, \infty]$-valued functions satisfying the measurability condition, i.e., a $f \geq 0$ belongs to $M^+$ if and only if $f \wedge \varphi$ is in $L$ for every $\varphi$ in $L$. It is clear that (a) $M^+ \subseteq L^+$, (b) any nonnegative function $f$ in $M$ (or $L$) belongs to $M^+$ (or $L^+$), (c) any finite valued function in $M^+$ belongs to $M$, and that (d)

$$I(f) = \sup \{ I(f \wedge \varphi) : \varphi \in L^+ \}, \quad \forall f \in M^+. $$

properly extends the definition of $I$ to $M^+$.

Moreover, the monotony of $I$ on $M^+$ is clear and the inequalities $(f + g) \wedge \varphi \leq f \wedge \varphi + g \wedge \varphi$ and $f \wedge \varphi + g \wedge \psi \leq (f + g) \wedge (f \wedge \varphi + g \wedge \psi)$, valid for any $f, g$ in $M^+$ and $\varphi, \psi$ in $L^+$, show that $I$ is semi-linear on $M^+$, i.e., if $f, g$ belong to $M^+$ (or $L^+$) and $c$ is a nonnegative constant then $f + cg$ belongs to $M^+$ (or $L^+$) and $I(f + cg) = I(f) + cI(g)$. Furthermore, the liminf and dominate convergence theorems are obtained from

**Theorem 1.24** (monotone). If $\{f_n\}$ is a sequence of functions in $M^+$ satisfying $f_{n+1} \geq f_n \geq 0$ almost everywhere for every $n \geq 1$ then $I(f) = \lim_n I(f_n)$.

**Proof.** By definition, for every $r < I(f)$ there exists $\varphi$ in $L^+$ such that $r < I(\varphi) \leq I(f)$. Since $\{f_n \wedge \varphi\}$ is an increasing sequence of functions in $L$ converging to $\varphi$, Theorem 1.13 yields $r < I(\varphi) = \lim_n I(f_n \wedge \varphi) \leq \lim_n I(f_n)$, and the desired convergence follows. 

A priori, due to the $\infty - \infty$ indetermination, we cannot define $I(f)$ for every $f$ in $M$. However, an element $f$ in $M$ such that $I(f^+) < \infty$ or $I(f^-) < \infty$ is
called quasi-integrable and \( I(f) \) makes sense as \( I(f^+) - I(f^-) \). Anyway, \( \mathcal{M}^+ \) is a (semi-vector) lattice and if \( \{f_n\} \) is a sequence in \( \mathcal{M}^+ (\mathcal{M}, \text{respectively}) \) then the functions \( \sup_n f_n(x), \inf_n f_n(x), \limsup_n f_n(x) \) and \( \liminf_n f_n(x) \) (when they are finite almost everywhere, respectively) belongs to \( \mathcal{M}^+ (\mathcal{M}, \text{respectively}) \). There is a subtle difference between \( L^+ \) and \( \mathcal{M}^+ \), but if \( X \) is a Polish space and the elementary functions are Borel function then \( \mathcal{M}^+ \) and \( L^+ \) are quite the same, depending on the initial vector lattice \( \mathcal{E} \). For instance, as mentioned early, if there exists of a strictly positive integrable function \( g \) then every function in \( \mathcal{M} \) is an almost everywhere limit of a sequence of functions in \( L \) (or even in \( \mathcal{E} \) if the strictly positive function \( g \) is an elementary function).

There are two classes of sets “to called measurable” that may be considered. Firstly, let \( \mathcal{M}_0 \subset 2^X \) the class of sets \( A \) such that (a) \( 1_A \) belongs to \( \mathcal{M} \) and (b) there exists a sequence \( \{g_k\} \) in \( L \) such that \( \sup_k g_k(x) > 0 \) for almost every \( x \) in \( A \). Alternatively, we may consider \( \mathcal{M} \subset 2^X \) the class of sets \( A \) such that \( 1_A \) belongs to \( \mathcal{M} \). Since \( \mathcal{M} \) is a vector lattice stable under the pointwise almost every convergence, and in view of Stone’s assumption (1.1-c), we deduce that \( \mathcal{M}_0 \) is a \( \sigma \)-ring (of \( \sigma \)-finite measurable sets) and \( \mathcal{M} \) is a \( \sigma \)-algebra (of measurable sets).

We use the following definition of measurability with respect to a \( \sigma \)-ring: a function \( f : X \to \mathbb{R} \) is \( \mathcal{M}_0 \)-measurable if and only if \( f^{-1}(B) \in \mathcal{M}_0 \) for every Borel set \( B \) in \( \mathbb{R} \) not containing 0. Thus, \( f \) is \( \mathcal{M} \)-measurable if and only if \( f \) is \( \mathcal{M}_0 \)-measurable and the set \( \{x : f(x) = 0\} \) belongs to \( \mathcal{M} \). This really means that a class of equivalence is measurable if there is one member in the class such that the above condition is satisfied.

**Lemma 1.25.** A real-valued function (class of equivalence) \( f \) defined \( X \) is \( \mathcal{M}_0 \)-measurable (or \( \mathcal{M} \)-measurable) if and only if \( f \) belongs to \( \mathcal{M} \).

**Proof.** It suffices to consider the case \( f \geq 0 \). Suppose \( f \) in \( \mathcal{M} \). To show that \( f \) is \( \mathcal{M} \)-measurable we should check that the set \( A = \{x \in X : f(x) > a\} \) belongs to \( \mathcal{M} \) for every \( a > 0 \). Now, by means of Stone’s assumption (1.1-c) we have \( 1_A = \lim_n \left(n[f(\cdot)/a - 1]^+\right) \land 1 \), which proves that \( 1_A \in \mathcal{M} \). Also the set \( \{x : f(x) \neq 0\} = \bigcap_{k=1}^{\infty} \{x : |f(x)| > 1/k\}^c \) belongs to the \( \sigma \)-algebra \( \mathcal{M} \).

For the converse, suppose \( f \) is a (nonnegative) \( \mathcal{M} \)-measurable. Remarking that the function \( 1_B \) belongs to \( \mathcal{M} \), for \( B = f^{-1}(\{a, b\}) \), with \( b > a > 0 \), and approximating \( f \) by a linear combination of characteristic functions, we show that \( f \) is a pointwise limit of a (monotone increasing) sequence in \( \mathcal{M} \), and therefore \( f \) belongs to \( \mathcal{M} \).

Comparing with the measure theory, we may begin with a semi-ring \( \mathcal{S} \) and to define the initial vector lattice \( \mathcal{E} \) as the simple functions \( \varphi = \sum_{i=1}^{n} a_i 1_{A_i} \) with \( A_i \) in \( \mathcal{E} \). In this case, the elementary sets of the form \( \varphi^{-1}(\{1, +\infty]\} \) with \( \varphi \in \mathcal{E} \) constitute the ring generated by \( \mathcal{S} \), i.e., \( \mathcal{E} = \{E = \varphi^{-1}(\{1, +\infty]\) : \varphi \in \mathcal{E}\} \). This definition of elementary sets works even if the vector lattice \( \mathcal{E} \) is initially given.

Now, if \( \mathcal{R} \) is the \( \sigma \)-ring generated by the elementary sets \( \mathcal{E} \) then (1) \( \mathcal{R} \) is a \( \sigma \)-algebra if \( 1_X \) is \( \sigma \)-integrable (i.e., \( X = \bigcup_{i=1}^{\infty} X_i \) with \( 1_{X_i} \) integrable), and (2) \( \mathcal{R} \) is the smallest \( \sigma \)-ring such that every function \( \varphi \in \mathcal{E} \) is measurable.
Only property (2) need some discussion. To this purpose, first recall that a function \( f : X \to \mathbb{R} \) is \( \mathcal{R} \)-measurable if and only if \( f^{-1}(B) \in \mathcal{R} \) for every Borel set \( B \) in \( \mathbb{R} \) not containing 0, and denote by \( \overline{\mathcal{R}} \) the smallest \( \sigma \)-ring such that every function \( \varphi \) in \( \mathcal{E} \) is measurable. Thus, the equalities
\[
(f/a)^{-1}(]1, +\infty[) = (f/a)^{-1}(]1, +\infty[),
\]
and the fact that the intervals \( ]a, +\infty[, \) \( ]-\infty, -a[ \) generated the \( \sigma \)-ring of Borel subsets of \( \mathbb{R} \) non containing 0, show that any function in \( \mathcal{E} \) is \( \mathcal{R} \)-measurable, i.e., \( \mathcal{R} \subseteq \overline{\mathcal{R}} \). Next, if \( E \in \mathcal{E} \), i.e., \( E = \varphi^{-1}(]1, +\infty[) \), then the function \( \varphi_n = (n(\varphi - \varphi \wedge 1)) \wedge 1 \) belongs to \( \mathcal{E} \) and \( \varphi_n \uparrow \mathbb{1}_E \), pointwise increasing, which means that \( \mathbb{1}_E \) is measurable, i.e., \( \mathcal{E} \subseteq \mathcal{R} \), which implies that \( \overline{\mathcal{R}} = \mathcal{R} \).

It is clear that if \( \mathcal{E} \) is the lattice generated by the characteristic functions of some semi-ring \( \mathcal{S} \) then \( \mathcal{R} \) is the \( \sigma \)-ring generated by \( \mathcal{S} \), as expected. However, if \( X \) is a Polish space and \( \mathcal{E} \) is the lattice of continuous functions with compact support then \( \mathcal{R} \) is the Baire \( \sigma \)-algebra, which may be strictly smaller than the Borel \( \sigma \)-algebra.

**Proposition 1.26** (Stone-Daniell). Given a pre-integral \( I \) on the vector lattice \( \mathbb{E} \) satisfying (1.1) and (1.2), there exists a measure space \( (X, \mathcal{M}, \mu) \) such that
\[
I(\varphi) = \int_X \varphi d\mu, \quad \forall \varphi \in \mathcal{E},
\]
and the measure \( \mu \) is uniquely determine on the \( \sigma \)-ring \( \mathcal{R} \), the smallest \( \sigma \)-ring for which all function in \( \mathcal{E} \) are measurable.

**Proof.** Certainly, \( \mathcal{M} \) and \( \mathcal{M}_0 \) are the \( \sigma \)-algebra and \( \sigma \)-ring defined above, and \( \mu(A) = I(\mathbb{1}_A) \) for every \( A \) in \( \mathcal{M}_0 \) and \( \mu(A) = \infty \) if \( A \) belongs to \( \mathcal{M} \setminus \mathcal{M}_0 \). Since any \( \mathcal{M} \)-measurable function can be approximated by an increasing sequence of simple functions, in particular any nonnegative elementary function \( \varphi \) is limit of a linear combination of characteristic functions of sets in \( \mathcal{M} \), we deduce the representation of \( I \) as the integral with respect to \( \mu \).

To show that \( \mu \) is is uniquely determine on the \( \sigma \)-ring \( \mathcal{R} \), first recall that the class
\[
\mathcal{E}_0 = \{ E = \varphi^{-1}(]1, +\infty[) : \varphi \in \mathcal{E} \},
\]
generates the \( \sigma \)-ring \( \mathcal{R} \); and for every \( E \) in \( \mathcal{E}_0 \), there is an increasing sequence \( \{ \varphi_n \} \subseteq \mathcal{E} \) such that \( \varphi_n \to \mathbb{1}_E \) and \( I(\mathbb{1}_E) < \infty \).

Now, let any two measures \( \mu \) and \( \nu \) be such that
\[
I(\varphi) = \int_X \varphi d\mu = \int_X \varphi d\nu, \quad \forall \varphi \in \mathcal{E},
\]
and for a fixed measurable set \( C \in \mathcal{M} \) with \( I(\mathbb{1}_C) < \infty \) (in particular for \( C \) in \( \mathcal{E}_0 \)) define the class \( \mathcal{C} = \mathcal{C}(\mu, \nu, C) \) of all sets \( A \) in \( \mathcal{R} \) such that \( \mu(A \cap C) = \nu(A \cap C) \).

Essentially by assumption, \( \mathcal{E}_0 \subseteq \mathcal{C} \). Since \( \mathcal{E}_0 \) is stable under finite intersections and unions, and all sets in \( \mathcal{E}_0 \) have finite measure (relative to both \( \mu \) and \( \nu \)), the class \( \mathcal{C} \) contains the ring generated by \( \mathcal{E}_0 \). Now, a monotone argument, shows that \( \mathcal{C} \) contains (is equal to) the class \( \mathcal{R} \cap \mathcal{C} = \{ A \cap C : A \in \mathcal{R} \} \).
Finally, another monotone argument proves that the class of sets in \( \mathcal{R} \) which are included in a countable union of sets in \( \mathcal{E}_0 \) is indeed the whole \( \sigma \)-ring \( \mathcal{R} \), we deduce that \( \mu = \nu \) on \( \mathcal{R} \).

- **Remark 1.27.** Let \( E \) be a vector lattice of real-valued functions defined on \( X \) with a pre-integral \( I : E \to \mathbb{R} \) satisfying (1.2) and such that there exists an almost everywhere positive integrable function \( \rho \). If \( \mathcal{F} \) is the smallest \( \sigma \)-algebra for which all function in \( E \) are measurable, then Stone-Daniell Proposition 1.26 proves that there exists a unique measure \( \mu \) on \( (X, \mathcal{F}) \) such that

\[
I(\varphi) = \int_X \varphi \mathrm{d}\mu, \quad \forall \varphi \in E,
\]

and a function is \( \mu \)-integrable if and only if it is \( I \) integrable.

Since integrable functions vanish (almost everywhere) on the complement of \( \sigma \)-finite sets, from the functional \( I \) viewpoint, only \( \sigma \)-finite sets are involved, i.e., the measure \( \mu \) can be unique only on \( \sigma \)-finite sets. Clearly, the class of all measurable \( \sigma \)-finite sets is the \( \sigma \)-ring \( \mathcal{R} \) (which is a \( \sigma \)-algebra if \( \mu \) is \( \sigma \)-finite) of the previous Proposition 1.26. For instance, the reader may consult the books by Dudley [36, Section 4.5, pp. 142–148], Haaser and Sullivan [62, Chapter 6, pp. 107–156], Phillips [100, Chapter 12, pp. 363–394], and Zaanen [136].

**Exercise 1.6.** Let \( E_i \) be a vector lattice of functions on a (Hausdorff) space \( X_i \), for \( i = 1, 2 \), and set \( X = X_1 \times X_2 \) and \( E = E_1 \otimes E_2 \), see Exercise 1.2. Assume that a pre-integral \( I_i \) is given on \( E_i \), \( i = 1, 2 \), and such that for every \( \varphi \) in \( E \), the function \( x_2 \mapsto I_1(\varphi(\cdot, x_2)) \) belongs to \( E_2 \). Prove that the iterate expression \( I(\varphi) = I_2(I_1(\varphi)) \) defines a pre-integral on \( E \). Based on results of this section, try to show that for any \( I \)-integrable function \( f \) there exists a \( I_2 \)-null set \( N_2 \) such that the function \( x_1 \mapsto f(x_1, x_2) \) is \( I_1 \)-integrable for every \( x_2 \) in \( X_2 \setminus N_2 \), e.g., see Taylor [122, Section 7.2, pp. 329–334].

**Exercise 1.7.** Let \( S \) be a (Stone) vector lattice, see (1.1), of bounded (real-valued) functions defined on \( X \). First (1) show that if \( f, g, h \geq 0 \) and \( f + g \geq h \) with \( f, g, h \) in \( S \) then we can write \( h = h_1 + h_2 \) with \( h_i \) in \( S \), \( 0 \leq h_1 \leq f \) and \( 0 \leq h_2 \leq g \). Next, let \( I \) be a linear functional on \( S \) such that there exists a constant \( C \) satisfying \( |I(f)| \leq C \|f\| \) for every function \( f \) in \( S \), where \( \|f\| = \sup \{|f(x)| : x \in X \} \). Define

\[
I^+(f) = \sup_{0 \leq h \leq f} \{I(h)\}, \quad \text{and} \quad I^-(f) = -\inf_{0 \leq h \leq f} \{I(h)\},
\]

for every \( f \geq 0 \) in \( S \), and later \( I^\pm(f) = I^\pm(f^+) - I^\pm(f^-) \). Prove (2) that \( I^+ \) and \( I^- \) are two linear (nonnegative) functionals such that \( I = I^+ - I^- \). Moreover, (3) if \( I \) is a signed pre-integral (i.e., besides being linear it has the monotone convergence property \( I(f_n) \to 0 \) whenever \( f_n \downarrow 0 \) pointwise decreasing to 0) then so are \( I^+ \) and \( I^- \).

With the notation of the above Exercise 1.7, we define the variation functional \( |I| = I^+ + I^- \), and we note that if \( X \) is a locally compact Hausdorff...

[Preli...
topological space then Dini’s Theorem implies that $I$ is indeed a signed pre-
integral, and therefore, $I^+$, $I^-$ and $|I|$ are all (nonnegative) pre-integrals, see
Lemma 1.2, i.e., in this case, (3) carries not additional assumption. The reader
may want to check the book Bogachev [19, Section 7.8, pp. 99–107].

1.2 Uniform Integrability

For the reader convenience, this section is almost a copy of one of the last
sections in our previous book [89]. Let $(\Omega, \mathcal{F}, \mu)$ be a $\sigma$-
finite measure space. On the vector space of integrable functions $L = L^1$ we can define the semi-norm

$$\|f\|_1 = \int_\Omega |f| \, d\mu,$$

and using equivalence classes we obtain a norm and therefore $L^1 = L^1/\sim$ or
$L^1(\Omega, \mathcal{F}, \mu)$ is a normed space. Elements in $L^1$ are classes of equivalence, but we
think of a function defined almost everywhere, and if necessary, we may complete
the definition everywhere as along as the operations involving elements in $L^1$
does not depend on the particular extension used. Special attention is necessary
to this point when dealing with measurable functions (or random variables or
processes) in probability theory. Since

$$\varepsilon \mu(\{x : |f(x)| \geq \varepsilon\}) \leq \|f\|_1,$$

convergence in $L^1$ (also called in mean) implies convergence in measure. Note
that

$$\left| \int_\Omega (f - g) \, d\mu \right| \leq \int_\Omega |f - g| \, d\mu = \|f - g\|_1,$$

if $f_n \to f$ in $L^1$ then the integral of $f_n$ converges to the integral of $f$, i.e, the
integral is a continuous mapping from $L^1$ into $\mathbb{R}$.

In the construction of the integral we allow functions taking valued in the
extended real numbers $\bar{R} = [-\infty, +\infty]$, but integrable functions are finite almost
everywhere, i.e., the set $\{x \in \Omega : |f(x)| = \infty\}$ is negligible, the set $\{x \in \Omega : |f(x)| \geq \varepsilon\}$ is finite and thus, the support $\{x \in \Omega : |f(x)| > 0\}$ is $\sigma$-finite. Therefore, $L^1(\Omega, \mathcal{F}, \mu; \bar{R}) = L^1(\Omega, \mathcal{F}, \mu; \mathbb{R})$. Thus the space $L^0(\Omega, \mathcal{F}, \mu; \bar{R})$ of
equivalence classes of measurable functions with values in $\bar{R}$ almost everywhere
finite is $L^0(\Omega, \mathcal{F}, \mu; \mathbb{R})$, i.e., equivalence classes with the condition $\mu(\{|f| = \infty\}) = 0$. In general, we may use the space $L^1(\Omega, \mathcal{F}, \mu; E)$ of $E$-valued integrable
functions, where $E$ is a Banach space as discussed later in Section 1.3.

1.2.1 Main Properties

**Definition 1.28.** Let $\{f_i : i \in I\}$ be a family of measurable functions almost
everywhere finite (or elements in $L^0$). If for every $\varepsilon > 0$ there exist $\delta > 0$ and a
set \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) such that for every \( F \in \mathcal{F} \) with \( \mu(F) < \delta \) we have

\[
\int_F f_i \, d\mu + \int_{A^c} |f_i| \, d\mu < \varepsilon, \quad \forall i \in I,
\]

then the family \( \{f_i : i \in I\} \) is called \( \mu \)-equicontinuous. While, if for every \( \varepsilon > 0 \) there exist \( \delta > 0 \) and a set \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) such that

\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i| \, d\mu + \int_{A^c} |f_i| \, d\mu < \varepsilon, \quad \forall i \in I,
\]

then the family is called \( \mu \)-uniformly integrable. The words uniform integrability or uniformly integrable may be used when the reference measure \( \mu \) is clear from the context.

It is clear that if \( \mu(\Omega) < \infty \) then we can take \( A = \Omega \) and the above definition is greatly simplified. Both \( \mu \)-equicontinuous and \( \mu \)-uniformly integrable have in common the part relative to the set \( A \), namely, for every \( \varepsilon > 0 \) there exists a set \( A \in \mathcal{F} \) such that

\[
\mu(A) < \infty \quad \text{and} \quad \sup_{i \in I} \int_{A^c} |f_i| \, d\mu < \varepsilon. \tag{1.5}
\]

This condition is useful only when \( \mu(\Omega) = \infty \), it involves the behavior of the set \( \{|f_i| \leq \delta\} \), as \( \delta \to 0 \), and it could be called tightness.

On the other hand, if the family is almost everywhere equibounded, i.e., \( |f_i| \leq M \) almost everywhere, for every index \( i \) in \( I \), then \( \{|f_i| \geq 1/\delta\} \) is the empty set for \( \delta < 1/M \) and

\[
\int_F |f_i| \, d\mu \leq M \mu(F),
\]

proving that a part of \( \mu \)-equicontinuity and \( \mu \)-uniform integrability (except the tightness condition) is satisfied. Moreover, the condition on the set \( F \) could be called uniform or equi absolute continuity of the family of measures obtained from the integrals. The following properties hold:

(1) If \( \Omega = \{1, 2, \ldots\} \) and \( \mu \) is the \( \sigma \)-finite measure \( \mu(F) = \sum_{k=1}^{\infty} 1_{k \in F} \), for every \( F \in 2^\Omega \), then there is no \( F \in 2^\Omega \) such that \( 0 < \mu(F) < 1 \), i.e., \( \mu(F) < \delta < 1 \) implies \( F = \emptyset \) and therefore the condition on the uniform absolute continuity is always satisfied. Now, regarding condition (1.5), for any set \( A \in 2^\Omega \) with \( \mu(A) < n < \infty \) we have \( A^c \supset \{k \geq n\} \). Thus, the sequence of functions \( f_i : \Omega \to \mathbb{R}, f_i(k) = k^{-1/i} - (k + 1)^{-1/i} \) satisfies

\[
\int_{\{k \geq n\}} |f_i| \, d\mu = \sum_{k=n}^{\infty} f_i(k) = \lim_k \left(n^{-1/i} - (k + 1)^{-1/i}\right) = n^{-1/i} \geq n^{-1},
\]

and therefore, \( \{f_i : i \geq 1\} \) fails to be \( \mu \)-equicontinuous (and \( \mu \)-uniformly integrable) because (1.5) is not satisfied.
(2) It is clear that if \( \{f_i : i \in I\} \) is a \( \mu \)-equicontinuous family of functions then the equality
\[
\int_F |f_i| \, d\mu = \int_{F \cap \{f_i > 0\}} f_i \, d\mu - \int_{F \cap \{f_i < 0\}} f_i \, d\mu
\]
shows that the family \( \{|f_i| : i \in I\} \) is also \( \mu \)-equicontinuous. Moreover, if \( \{f_i : i \in I\} \) and \( \{g_j : j \in J\} \) are two families of \( \mu \)-equicontinuous functions then for any constant \( c \) the family \( \{h_{i,j} = f_i + cg_j : i \in I, j \in J\} \) is also \( \mu \)-equicontinuous.

(3) If \( \{f_i : i \in I\} \) is a family of \( \mu \)-uniformly integrable functions then the inequality
\[
\int_F |f_i| \, d\mu \leq \frac{\mu(F)}{\delta} + \int_{\{|f_i| \geq 1/\delta\}} |f_i| \, d\mu, \quad \forall F \in \mathcal{F}, \forall i \in I,
\]
shows that the family is also \( \mu \)-equicontinuous. Moreover, for \( F = A \) with \( \mu(A) < \infty \) as in the Definition 1.28, we deduce that \( \sup_{i \in I} \|f_i\|_1 < \infty \).

(4) For a family \( \{f_i : i \in I\} \) of \( \mu \)-equicontinuous functions, each member \( f_i \) is an integrable function. Indeed if \( F_n = \{|f_i| \geq n\} \) then \( \bigcap_n F_n = \{|f| = \infty\} \), and for any set \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) we have \( \mu(F_n \cap A) \to 0 \) as \( n \to \infty \). Hence, take any \( \varepsilon > 0 \) and find \( \delta > 0 \) and \( A \in F \) as above. Since \( \Omega = F_n \cup (F_n^c \cap A^c) \cup (F_n^c \cap A) \), taking \( n \) such that \( \mu(F_n) < \delta \) and \( F = F_n \) we deduce
\[
\int_{\Omega} |f_i| \, d\mu = \int_F |f_i| \, d\mu + \int_{F_n^c \cap A^c} |f_i| \, d\mu + \int_{F_n^c \cap A} |f_i| \, d\mu \\
\leq \int_F |f_i| \, d\mu + \int_{A^c} |f_i| \, d\mu + n\mu(A_i),
\]
i.e., each \( f_i \) must be integrable.

(5) If \( \{f_i : i \in I\} \) is a \( \mu \)-equicontinuous family of functions then we may have \( \sup_i \|f_i\|_1 = \infty \). Indeed, if the measure \( \mu \) is finite with an atom \( A_1 \) and \( f_i = (i/\mu(A_1)) \mathbb{1}_{A_1} \) then for any \( \varepsilon > 0 \) we can choose \( A = \Omega \) and \( \delta < \mu(A_1) \) to have
\[
0 = \int_F |f_i| \, d\mu + \int_{A^c} |f_i| \, d\mu \leq \delta \mu(\Omega) \leq \varepsilon, \quad \text{but} \quad \int_{\Omega} |f_i| \, d\mu = i,
\]
for every \( F \in \mathcal{F} \) with \( \mu(F) < \delta \). On the other hand, it is clear that if the set \( A \) satisfying (1.5) can be decomposed into a finite number of measurable sets \( A_1, \ldots, A_n \) such that
\[
\int_{A_k} |f_i| \, d\mu < \delta, \quad \forall k = 1, \ldots, n, \quad \forall i \in I,
\]
than \( \sup_i \|f_i\|_1 < \infty \). Therefore, we deduce that if the family of measures induced by the functions \( \{f_i : i \in I\} \) is uniformly absolutely \( \mu \)-continuous, i.e.,
for every $\varepsilon > 0$ there exist $\delta > 0$ such that for every $F \in \mathcal{F}$ with $\mu(F) < \delta$ we have

$$\int_F f_i \, d\mu < \varepsilon, \quad \forall i \in I,$$

and also, for every $\delta > 0$ there exist $A_1, \ldots, A_n$ in $\mathcal{F}$ such that $A = A_1 \cup \cdots \cup A_n$ has finite measure, $\mu(A) < \infty$, and

$$\int_{A_k} |f_i| \, d\mu + \int_{A^c} |f_i| \, d\mu < \delta, \quad \forall k = 1, \ldots, n, \quad \forall i \in I,$$

then $\{f_i : i \in I\}$ is $\mu$-uniformly integrable. In other words, if the measure $\mu$ is diffuse or non-atomic (i.e., for any set $A$ with $\mu(A) < \infty$ and for every $\delta > 0$ there is a decomposition of $A$ into a finite number of measurable sets, $A = A_1 \cup \cdots \cup A_n$ with $\mu(A_i) < \delta$, for every $i = 1, \ldots, n$), then any $\mu$-equicontinuous family $\{f_i : i \in I\}$ is also $\mu$-uniformly integrable.

(6) A family $\{f_i : i \in I\}$ of $\mu$-equicontinuous functions with $\sup_{i \in I} \|f_i\|_1 < \infty$ is $\mu$-uniformly integrable. Indeed, the inequality

$$\mu(\{|f_i| \geq c\}) \leq \frac{1}{c} \int_{\Omega} |f_i| \, d\mu \leq \frac{1}{c} \sup_{i \in I} \|f_i\|_1,$$

shows that for every $\delta > 0$ and any $i$ there exists $c$ sufficiently large so that the set $F_{i,c} = \{|f_i| \geq c\}$ satisfies $\mu(F_{i,c}) < \delta$. Now, for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$F \in \mathcal{F} \text{ with } \mu(F) < \delta \quad \implies \quad \int_F |f_i| \, d\mu \leq \varepsilon,$$

and by taking $F = F_{i,c}$ we conclude. As a consequence we deduce that if $\{f_i : i \in I\}$ and $\{g_j : j \in J\}$ are two families of $\mu$-uniform integrable functions then for any constant $c$ the family $\{h_{i,j} = f_i + cg_j : i \in I, j \in J\}$ is also $\mu$-uniformly integrable.

(7) Any family $\{f_i : i \in I\}$ of measurable functions dominated by an integrable function $g$, i.e., $|f_i| \leq g$ almost everywhere, is $\mu$-uniformly integrable. Indeed, since $g$ is integrable, it is clear that for every $\varepsilon > 0$ there exists a $\delta_1 > 0$ such that $F \in \mathcal{F}$ with $\mu(F) < \delta_1$ implies

$$\int_F |f_i| \, d\mu \leq \int_F |g| \, d\mu \leq \frac{\varepsilon}{3}.$$

Next, if $A_n = \{x \in \Omega : |g(x)| \geq 1/n\}$ then $\mathbb{1}_{A_n} g \to 0$ almost everywhere as $n \to \infty$. Thus, Lebesgue dominate convergence Theorem 1.18, shows that

$$\lim_n \int_{A_n^c} |g| \, d\mu = \lim_n \int_{\Omega} \mathbb{1}_{A_n^c} |g| \, d\mu = 0,$$
i.e., there exists \( A = A_n \) such that
\[
\int_{A^c} |f_i| \, d\mu \leq \int_{A^c} |g| \, d\mu < \frac{\varepsilon}{3}, \quad \text{and} \quad \mu(A) \leq n \int_{\Omega} |g| \, d\mu,
\]
and we conclude by taking \( \delta \in (0, \delta_1] \) such that \( \delta \mu(A) < \varepsilon/3 \).

(8) Similarly to (7), any family \( \{f_i : i \in I\} \) of measurable functions dominated by a \( \mu \)-equicontinuous (or \( \mu \)-uniformly integrable) family \( \{g_j : j \in J\} \) (i.e., for every \( i \) there exists \( j \) such that \( |f_i| \leq g_j \) almost everywhere) results also \( \mu \)-equicontinuous (or \( \mu \)-uniformly integrable).

(9) Let \( r \to p(r), r > 0 \), be a nonnegative Borel measurable function such that \( p(r)/r \to \infty \) as \( r \to \infty \), e.g., \( p(r) = r^\alpha \) with \( \alpha > 1 \) or \( p(r) = r \ln(1 + r) \). If \( \sup_{i \in I} \|p(|f_i|)\|_1 = C < \infty \) then for every \( \varepsilon > 0 \) choose \( \delta > 0 \) such that \( p(r) \geq rC/\varepsilon \), for every \( r \geq 1/\delta \). Thus
\[
\int_{\{|f_i| \geq 1/\delta\}} \|f_i\| \, d\mu \leq \frac{\varepsilon}{C} \|p(|f_i|)\|_1 \leq \varepsilon, \quad \forall i \in I.
\]
Hence, if \( \mu(\Omega) = \infty \) then we need only to add the condition (1.5), to deduce that \( \{f_i : i \in I\} \) is a \( \mu \)-uniformly integrable family.

**Proposition 1.29.** If \( \{f_i : i \in I\} \) is \( \mu \)-equicontinuous then it is uniformly \( \sigma \)-additive, i.e.,
\[
\forall \{B_n\} \subset \mathcal{F}, B_n \supset B_{n+1}, \cap_n B_n = \emptyset,
\]
we have
\[
\lim_{n} \left( \sup_{i \in I} \int_{B_n} |f_i| \, d\mu \right) = 0.
\]

(1.6)
Conversely, if either the index set \( I \) is countable or the measure \( \mu \) is \( \sigma \)-finite then the uniform \( \sigma \)-additive condition (1.6) implies the \( \mu \)-equicontinuous condition.

**Proof.** Indeed, let \( \{B_n\} \) be a decreasing sequence in \( \mathcal{F} \) such that \( \cap_n B_n = \emptyset \). From (1.5), for any \( \varepsilon > 0 \) there exists a measurable set \( A \) with \( \mu(A) < \infty \) such that
\[
\int_{B_n} |f_i| \, d\mu = \int_{B_n \cap A^c} |f_i| \, d\mu + \int_{B_n \cap A} |f_i| \, d\mu \leq \varepsilon + \int_{B_n \cap A} |f_i| \, d\mu.
\]
Since \( \mu(B_n \cap A) < \infty \) we have \( \mu(B_n \cap A) \to 0 \) and the \( \mu \)-equicontinuity (the condition on the set \( F \)) yields (1.6).

Conversely, if the index set \( I \) is countable or the measure \( \mu \) is \( \sigma \)-finite, the set \( \bigcup_{i \in I} \{f_i \neq 0\} \) is contained in a \( \sigma \)-finite measurable set \( E \), and so, there exists an increasing sequence of measurable sets \( \{E_k\} \) with \( \mu(E_k) < \infty \) such that \( E = \bigcup_k E_k \). Thus
\[
\int_{E_k^c} |f_i| \, d\mu = \int_{\Omega \setminus E_k} |f_i| \, d\mu \quad \text{and} \quad \lim_k \int_{E_k \setminus E_k^c} |f_i| \, d\mu = 0
\]
where the limit is uniform in view of (1.6). Hence we deduce (1.5) with \( A = E_k \) and \( k \) sufficiently large. Therefore, if \( \{B_n : n \geq 1\} \) is a sequence in \( \mathcal{F} \) such that \( \mu(B_n) < \infty \) and \( \mu(A) = 0 \) with \( \bigcap_n B_n = A \), then \( C_n = \bigcap_{k=1}^n (B_k \setminus A) \) forms a decreasing sequence satisfying \( \bigcap_n C_n = \emptyset \) and the uniform \( \sigma \)-additivity (1.6) yields a contradiction with the \( \mu \)-equicontinuous condition.

Note that in the above Proposition 1.29, because the set \( \{f_i \neq 0\} \) is \( \sigma \)-finite for every \( i \in I \), the countability of the index set \( I \) can be avoided if we assume that the \( \sigma \)-ring of all \( \sigma \)-finite measurable sets is countable generated. It is also clear that this condition is related to the separability of the Banach space \( L^1(\Omega, \mathcal{F}, \mu) \). Another aspect of the \( \mu \)-uniformly integrability is analyzed later, in Definition 1.37 and Theorem 1.38.

A family of measures \( \{\mu_i : i \in I\} \) is called uniform absolutely continuous on a measure space \( (\Omega, \mathcal{F}, \mu) \) (or \( \mu \)-uniformly absolutely continuous) if for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that for every measurable set \( F \) with \( \mu(F) < \delta \) we have \( \mu_i(F) < \varepsilon \), for every \( i \in I \). To mimic the \( \mu \)-equicontinuity of a family of integrable functions, we may add a tightness condition like: for every \( \varepsilon > 0 \) there exist measurable sets \( A_i \) such that \( \sup_{i \in I} \mu_i(A_i) < \infty \) and \( \mu_i(A_i^c) < \varepsilon \), for every \( i \in I \). Actually, given a family \( (\Omega_i, \mathcal{F}_i, \mu_i) \) of measure spaces and a family \( \{f_i : i \in I\} \) of measurable functions \( f_i : \Omega_i \to \mathbb{R} \) almost everywhere finite, i.e., elements of \( L^0(\Omega_i, \mathcal{F}_i, \mu_i) \), then we could say that they are equi-continuous if for every \( \varepsilon > 0 \) there exist \( \delta > 0 \) and sets \( A_i \) in \( \mathcal{F}_i \) such that \( \sup_{i \in I} \mu_i(A_i) < \infty \) and for every set \( F \) in \( \mathcal{F}_i \) with \( \mu_i(F_i) < \delta \) we have

\[
\int_{F_i} f_i \, d\mu_i + \int_{A_i^c} |f_i| \, d\mu_i < \varepsilon, \quad \forall i \in I,
\]

while the uniform integrability is stated with the condition

\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i| \, d\mu_i + \int_{A_i^c} |f_i| \, d\mu_i < \varepsilon, \quad \forall i \in I.
\]

The reader can verify that most of the previous properties, (1) . . . , (9) above, remain true for this setting, where both the functions \( f_i \) and the measures \( \mu_i \) are indexed by \( i \) in \( I \). For instance, it is clear that property (7) make sense only when \( \Omega_i = \Omega \) the same abstract space. Nevertheless, when comparing with the uniform \( \sigma \)-additivity property as in Proposition 1.29 we get some difficulties. In particular, if we are dealing with probability measures then we could take \( A_i = \Omega \) and virtually, this question does not occur. Similarly, if the abstract spaces \( (\Omega_i, \mathcal{F}_i) = (\Omega, \mathcal{F}) \) for every \( i \in I \) then \( \sup_{i \in I} \mu_i(A_i) < +\infty \) can be replaced by a more useful condition, namely, the family of finite measures \( \{\lambda_i(B) = \mu_i(B \cap A_i) : i \in I\} \) is uniformly \( \sigma \)-additive. Note that Vitali-Hahn-Saks Theorem 2.30 yields some light on this point, but the situation in general is complicated and some tools from Functional Analysis are really useful. Therefore, uniform absolutely continuity or uniform integrability or uniform \( \sigma \)-additivity for a family of measures is not completely discussed in these notes. Perhaps checking the viewpoint in Schilling [111, Chapter 16, pp. 163–175] may help.
1.2.2 Mean Convergence

When comparing the convergence almost everywhere (or in measure) with the mean convergence (i.e., in $L^1$) we encounter the following equivalence:

**Theorem 1.30 (Vitali).** Let $\{f_n\}$ be a pointwise almost everywhere Cauchy sequence of integrable functions. Then $\{f_n\}$ is a Cauchy sequence in $L^1$ if and only if $\{f_n\}$ is $\mu$-equicontinuous.

**Proof.** First, for every $\varepsilon > 0$ there exist $A$ and $\delta > 0$ such that for any $F \in \mathcal{F}$ with $\mu(F) < \delta$ we have

$$\int_F |f_n| \, d\mu + \int_{A^c} |f_n| \, d\mu + \delta \mu(A) < \frac{\varepsilon}{4}, \quad \forall n.$$

Thus, the estimate

$$\int_\Omega |f_n - f_k| \, d\mu \leq \int_{A^c} |f_n - f_k| \, d\mu + \int_{A \cap \{ |f_n - f_k| \geq \delta \}} |f_n - f_k| \, d\mu + \delta \mu(A)$$

shows that

$$\int_\Omega |f_n - f_k| \, d\mu < \frac{\varepsilon}{2} + \int_{A \cap \{ |f_n - f_k| \geq \delta \}} |f_n - f_k| \, d\mu.$$

Since $\{f_n\}$ is a almost everywhere Cauchy sequence and $\mu(A) < \infty$, there exists an index $n_\varepsilon$ such that $\mu(\{ A \cap |f_n - f_k| \geq \delta \}) < \delta$, for every $n, k \geq n_\varepsilon$. Hence, taking $F = \{ A \cap |f_n - f_k| \geq \delta \}$ we have

$$\int_{A \cap \{ |f_n - f_k| \geq \delta \}} |f_n - f_k| \, d\mu < \frac{\varepsilon}{2},$$

i.e., $\|f_n - f_k\|_1 < \varepsilon$, for every $n_\varepsilon$.

Assuming that $\{f_n\}$ is a Cauchy sequence in $L^1$, given $\varepsilon > 0$ there exists an index $n_\varepsilon$ such that $\|f_n - f_k\|_1 \leq \varepsilon/2$ for every $n, k \geq n_\varepsilon$. Thus for any $A \in \mathcal{F}$

$$\int_{A^c} |f_n| \, d\mu \leq \int_{A^c} |f_{n_\varepsilon}| \, d\mu + \frac{\varepsilon}{2}, \quad \forall n \geq n_\varepsilon. \tag{1.7}$$

Since each $f_i$ is integrable, for every $\delta > 0$ the set $F_{i,\delta} = \{|f_i| \geq \delta\}$ has finite measure,

$$\int_{F_{i,\delta}} |f_i| \, d\mu = \int_{\{0 < |f_i| < \delta\}} |f_i| \, d\mu \to 0 \quad \text{as} \quad \delta \to 0,$$

and

$$\int_F |f_i| \, d\mu \to 0 \quad \text{as} \quad \mu(F) \to 0,$$
for any fixed $i$. If $A_\delta = \bigcup_{i=1}^{n_\varepsilon} F_{i,\delta}$ then for every $k = 1, \ldots, n_\varepsilon$ we deduce
\[
\int_{A_\delta^c} |f_k| \, d\mu \leq \int_{F_{k,\delta}^c} |f_k| \, d\mu \leq \max_{1 \leq i \leq n_\varepsilon} \int_{F_{i,\delta}^c} |f_i| \, d\mu \leq \frac{\varepsilon}{2},
\]
provided $\delta$ is sufficiently small. Thus, there is $\delta$ such that for $A = A_\delta$ we have
\[
\int_{A^c} |f_k| \, d\mu \leq \varepsilon, \quad \forall i \geq 1, \text{ and } \mu(A) < \infty.
\]
Similarly,
\[
\max_{1 \leq i \leq n_\varepsilon} \int_{F_i} |f_i| \, d\mu \to 0 \quad \text{as } \mu(F) \to 0,
\]
and with $A^c = F$ in (1.7), we complete the proof of the $\mu$-equicontinuity.

Note that in the proof of the above result we have shown that if a Cauchy sequence in $L^0 \cap L^1$ (i.e., in measure) is $\mu$-equicontinuous then it is also a Cauchy sequence in $L^1$. Moreover, we may assume that the sequence $\{f_n\}$ of integrable functions is a Cauchy sequence in measure for every measurable set of finite measure, i.e., for every $\varepsilon > 0$ and every $A \in \mathcal{F}$ with $\mu(A) < \infty$ we have
\[
\mu(\{x \in A : |f_n(x) - f_k(x)| \geq \varepsilon\}) \to 0 \quad \text{as } n, k \to \infty, \quad (1.8)
\]
to deduce that $\{f_n\}$ is a Cauchy sequence in $L^1$ if and only if $\{f_n\}$ is $\mu$-equicontinuous. For instance, Lebesgue dominate convergence Theorem 1.18 can be restated as
\[
\lim_n \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu,
\]
for any $\mu$-equicontinuous sequence $\{f_n\}$ of measurable (necessarily integrable) functions which converges to a almost everywhere finite function $f$, in measure for every measurable set of finite measure.

Actually, it is a good exercise to revise the proof of Vitali Theorem 1.30 and to deduce the following generalization

**Proposition 1.31.** Let $\{f_n\}$ be a Cauchy sequence of measurable functions, in measure for every measurable set of finite measure, i.e., (1.8). Then $\{f_n\}$ is a $p$-Cauchy sequence, $0 < p < \infty$, i.e., for every $\varepsilon > 0$ there exists $n_\varepsilon$ such that
\[
\int_{\Omega} |f_n - f_m|^p \, d\mu < \varepsilon, \quad \forall n, m \geq n_\varepsilon,
\]
if and only if $\{|f_n|^p\}$ is $\mu$-equicontinuous. \qed

There are several application of Vitali Theorem 1.30, namely,
Corollary 1.32. Let \( \{f_n\} \) be a sequence of integrable functions which converges to an integrable function \( f \), in measure on every measurable set of finite measure. If
\[
\lim_n \int_\Omega f_n^+ \, d\mu = \int_\Omega f^+ \, d\mu \quad \text{and} \quad \lim_n \int_\Omega f_n^- \, d\mu = \int_\Omega f^- \, d\mu \quad (1.9)
\]
then \( \{f_n\} \) is \( \mu \)-uniform integrable and \( f_n \to f \) in \( L^1 \).

Proof. From the elementary inequality
\[
|a^+ - b^+| \lor |a^- - b^-| \leq |a - b| \leq |a^+ - b^+| + |a^- - b^-|, \quad \forall a, b \in \mathbb{R},
\]
we deduce that (1) \( f_n^+ \to f^+ \) and \( f_n^- \to f^- \) in measure (on every measurable set of finite measure), and that (2) \( f_n \to f \) in \( L^1 \) if and only if \( f_n^+ \to f^+ \) and \( f_n^- \to f^- \) in \( L^1 \). Hence we may assume that \( f_n \) and \( f \) are nonnegative, without any lost of generality.

Now, the dominate convergence implies that
\[
\lim_n \int_\Omega (f_n \wedge f) \, d\mu = \int_\Omega f \, d\mu
\]
and by assumption
\[
\lim_n \int_\Omega (f_n + f) \, d\mu = 2 \int_\Omega f \, d\mu.
\]
Hence, the equality
\[
|f_n - f| = f_n \lor f - f_n \wedge f = f_n + f - 2(f_n \wedge f),
\]
shows that \( \|f_n - f\|_1 \to 0 \).

Finally, the condition (1.9) implies that \( \sup_n \|f_n\|_1 < \infty \) and Vitali Theorem 1.30 (actually Proposition 1.31 with \( p = 1 \)) yields the \( \mu \)-equicontinuity of \( \{f_n\} \), and we deduce the \( \mu \)-uniform integrability condition. \( \square \)

Note that if
\[
\lim_n \int_\Omega f_n \, d\mu = \int_\Omega f \, d\mu \quad \text{and} \quad \lim_n \int_\Omega |f_n| \, d\mu = \int_\Omega |f| \, d\mu
\]
then the relation \( a^\pm = (|a| \pm a)/2 \), for every real number \( a \), and the linearity of the integral show that (1.9) holds.

Proposition 1.33. If \( \{f_n\} \) is a sequence of \( \mu \)-uniformly integrable functions such that the negative part of the superior limit (\( \limsup_n f_n^- \)) is an integrable function then
\[
\lim \sup_n \int_\Omega f_n \, d\mu \leq \int_\Omega \lim \sup_n f_n \, d\mu. \quad (1.10)
\]
Chapter 1. Abstract Integration

Proof. Since \( \{f_n\} \) are \( \mu \)-uniformly integrable functions, for a given \( \varepsilon > 0 \) there exists \( \delta > 0 \) and \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) such that

\[
\int_{A \cap \{|f_n| > 1/\delta\}} |f_n| \, d\mu + \int_{A^c} |f_n| \, d\mu \leq \varepsilon, \quad \forall n,
\]

Now, decompose the integral

\[
\int_{\Omega} f_n \, d\mu = \int_{A^c} f_n \, d\mu + \int_{A \cap \{|f_n| > 1/\delta\}} f_n \, d\mu + \int_{A \cap \{|f_n| \leq 1/\delta\}} f_n \, d\mu \]

to check that for every \( n \) we have

\[
\int_{\Omega} f_n \, d\mu \leq \varepsilon + \int_{\Omega} g_n \, d\mu, \quad \text{where} \quad g_n = 1_{A \cap \{|f_n| \leq 1\}} f_n,
\]

with \( |g_n| \leq (1/\delta) 1_A \). Thus, by means of Lebesgue dominate convergence theorem we obtain

\[
\limsup_n \int_{\Omega} f_n \, d\mu \leq \varepsilon + \limsup_n g_n \, d\mu.
\]

Hence, if \( \limsup_n f_n \geq 0 \) then \( \limsup_n g_n \leq \limsup_n f_n \) and we deduce (1.10). Otherwise, because \( (\limsup_n f_n)^- = g \) is an integrable function, we may replace \( f_n \) with \( f_n + g \) to obtain the desired inequality. \( \square \)

Let us comment on the above Proposition 1.33. First, for a measure space \((\Omega, \mathcal{F}, \mu)\), take a measurable set \( A \in \mathcal{F} \) with \( 0 < \mu(A) \leq 1 \) and find a finite partition \( A = \bigcup_{k=1}^{K} A_{k,i} \) with \( 0 < \mu(A_{k,i}) \leq 1/k \), for every \( i \). If \( \{a_k\} \) and \( \{b_k\} \) are two sequences of real numbers then we construct a sequence of functions \( \{f_n\} \) as follows: the sequence of integers \( \{1, 2, 3, \ldots, 10, 11, \ldots\} \) is grouped as \( \{(1); (2, 3); (4, 5, 6); (7, 8, 9, 10); \ldots\} \) where the \( k \) group has exactly \( k \) elements, i.e., for any \( n = 1, 2, \ldots \), we select first \( k = 1, 2, \ldots \), such that \( (k-1)k/2 < n \leq k(k+1)/2 \) and we write (uniquely) \( n = (k-1)k/2 + i \) with \( i = 1, 2, \ldots, k \) to define

\[
f_n(x) = \begin{cases} 
a_k & \text{if } x \in A \setminus A_{k,i}, \\
b_k & \text{if } x \in A_{k,i}. \end{cases}
\]

Now, we may construct a sequence of nonnegative function \( \{f_n\} \) with \( a_k = 0 \) and \( b_k = \sqrt{k} \), for every \( k \) so that

\[
\int_{\Omega} f_n \, d\mu = b_k \mu(A_{k,i}) \leq \sqrt{k} \leq \frac{2}{\sqrt{n}}.
\]

Because for every \( x \in A \) there exist \( i, k \) such that \( x \in A_{k,i} \) and then \( f_n(x) = b_k \), we deduce that \( \limsup_n f_n(x) = \infty \), for every \( x \) in \( A \). Since the sequence \( \{f_n\} \)
converges (to 0) in $L^1$, this is an example of the strict inequality $0 < \infty$ in (1.10). More general, if we choose $a_k = a$ and $b_k \to b$ with $\lim_k b_k/k = 0$ then

$$
\int_{\Omega} f_n \, d\mu = a\mu(A \setminus A_{k,i}) + b_k\mu(A_{k,i}) \to a\mu(A), \quad \text{as} \quad n \to \infty,
$$

while $\lim sup_n f_n(x) = a \vee b$. This sequence $\{f_n\}$ is also $\mu$-uniformly integrable and the inequality (1.10) becomes $a\mu(A) \leq (a \vee b)\mu(A)$. For instance, if $a = -2$ and $b_k = -1$ we have the strict inequality $(-2)\mu(A) < (-1)\mu(A)$.

Another example, if the sequence $\{f_n\}$ admits a sub-sequence $\{f_{n_k}\}$ convergence almost everywhere to some function $f$ then

$$
\limsup_n f_n \geq \limsup_k f_{n_k} = f, \quad \text{a.e.}
$$

and therefore integrability of $(\limsup_n f_n)^-$ is guarantee. On the other hand, for a given $n = 1, 2, \ldots$, divide the interval $[0, 1]$ into $I_{k,n} = [(k-1)2^{-n}, k2^{-n}]$ with $k = 1, 2, \ldots, 2^n$ and define the functions $f_n(x) = (-1)^k$ for every $x$ in $I_{k,n}$ to check that

$$
|\{x : |f_n(x) - f_m(x)| \geq 1\}| = \frac{1}{2}, \quad \forall n \neq m,
$$

where $|\cdot|$ denotes the Lebesgue measure. Because $|f_n(x)| \leq 1$, this yields an example (in the Lebesgue space measure space $[0, 1]$) of an uniformly integrable sequence with no convergence (almost everywhere) sub-sequences, but with $\liminf_n f_n(x) = -1$ and $\limsup_n f_n(x) = 1$ for all $x$ in $[0, 1] \setminus \{k2^{-n} : k = 1, 2, \ldots, 2^n, \text{ and } n = 1, 2, \ldots\}$. Again, in this case, the inequality (1.10) is satisfied strictly, $0 < 1$.

**Theorem 1.34.** Let $\{f_n\}$ be a sequence of $\mu$-uniformly integrable functions and consider the limits $\overline{f} = \limsup_n f_n$ and $\underline{f} = \liminf_n f_n$. Then

$$
\int_{\Omega} \overline{f} \, d\mu \leq \liminf_n \int_{\Omega} f_n \, d\mu \leq \limsup_n \int_{\Omega} f_n \, d\mu \leq \int_{\Omega} \underline{f} \, d\mu, \quad (1.11)
$$

where the positive part $(\overline{f})^+$ and the negative part $(\underline{f})^-$ are both integrable.

**Proof.** Since the sequence $\{f_n\}$ is $\mu$-integrable, we obtain that the numerical sequence $\{\|f_n\|_1\}$ is bounded, and in view of the above inequality (1.11), we deduce that $(\overline{f})^+$ and $(\underline{f})^-$ are both integrable.

Now, the point is to check that the extra assumption on the integrability of the limit $(\limsup_n f_n)^-$ is not necessary in Proposition 1.33.

Indeed,\(^1\) because $-f_n^- \leq f_n$ we obtain $-\liminf_n f_n^- = \limsup_n (-f_n^-) \leq \limsup_n f_n$ and therefore $(\limsup_n f_n^-) \leq \liminf_n f_n^-$. Hence, by Fatou lemma we deduce

$$
\int_{\Omega} \liminf_n f_n^- \, d\mu \leq \liminf_n \int_{\Omega} f_n^- \, d\mu < \infty,
$$

\(^1\)a personal communication of N. Krylov
which implies that \( (\limsup_n f_n)^- \) is integrable.

Hence, we can apply Proposition 1.33 for the sequences \( \{f_n\} \) and \( \{-f_n\} \) to deduce the inequality (1.11).

Note that without assuming quasi-integrability for the limits \( f \) and \( \overline{f} \) (i.e., \( (f)^+ \) and \( (\overline{f})^- \) are both integrable), the \( \mu \)-uniform integrability cannot be replaced by \( \mu \)-equicontinuity. Indeed, similar to example in (5) after Definition 1.28, for a finite measure \( \mu \) with two atoms \( A_1 \) and \( A_2 \), \( \Omega = A_1 \cup A_2 \), and the sequence of functions with \( f_i = (i/\mu(A_1))1_{A_1} \) and \( f_i = (-i/\mu(A_2))1_{A_2} \), \( i \geq 1 \) is \( \mu \)-equicontinuous, but the limit \( f(A_k) = \lim_i f_i(A_k) \) is \( = +\infty \) for \( k = 1 \) and \( = -\infty \) for \( k = 2 \),

\[
f(A_k) = \lim_i f_i(A_k), \quad f(A_1) = +\infty, \quad f(A_2) = -\infty, \quad \int_\Omega f_i \, d\mu = 0,
\]

for any \( i \geq 1 \), and \( f \) is not quasi-integrable.

### 1.2.3 Convergence in Norm

The following result, which is also an application of Vitali Theorem 1.30 makes a connection with \( p \)-integrable functions.

**Proposition 1.35.** Let \( \{f_n\} \) be a bounded sequence in \( L^p(\Omega, \mathcal{F}, \mu) \), for some \( 0 < p < \infty \). If \( f_n \) converges to \( f \) in measure for every measurable set of finite measure then

\[
\lim_n \int_\Omega \left| |f_n|^p - |f_n - f|^p - |f|^p \right| \, d\mu = 0.
\]

**Proof.** Firstly, for every \( \varepsilon > 0 \) there exists a constant \( C_\varepsilon > 0 \) such that for every numbers \( a \) and \( b \) we have

\[
|a + b|^p - |b|^p \leq \varepsilon |b|^p + C_\varepsilon |a|^p. \quad (1.12)
\]

Indeed, if \( 0 < p \leq 1 \) the the simple estimate \( |a + b|^p \leq |a|^p + |b|^p \) yields estimate (1.12). Now, for \( 1 < p < \infty \), the function \( t \mapsto |t|^p \) is convex; and so \( |a + b|^p \leq (|a| + |b|)^p \leq (1 - \lambda)^{-p}|a|^p + \lambda^{1-p}|b|^p \), for any \( \lambda \) in \( (0, 1) \). Hence, by taking \( \lambda = (1 + \varepsilon)^{1/(1-p)} \) we deduce (1.12) with \( p > 1 \).

Secondly, by assumption

\[
\int_\Omega |f_n|^p \, d\mu \leq C < \infty, \quad \forall n,
\]

and since \( |f_n - f|^p \leq 2^p (|f_n|^p + |f|^p) \), we obtain

\[
\int_\Omega |f_n - f|^p \, d\mu \leq 2^{p+1} C, \quad \forall n,
\]

for every \( 0 < p < \infty \).
Next, estimate (1.12) implies
\[ |f_n|^p - |f_n - f|^p - |f|^p | \leq |f_n|^p - |f_n - f|^p + |f|^p \leq \varepsilon |f_n - f|^p + (1 + C\varepsilon)|f|^p. \]

Hence, by setting \( g_n = (|f_n|^p - |f_n - f|^p - |f|^p | - \varepsilon |f_n - f|^p)^+ \), we have \( 0 \leq g_n \leq (1 + C\varepsilon)|f|^p \) and so, Vitali Theorem 1.30 yields
\[ \lim_n \int_{\Omega} g_n \, d\mu = 0. \]

Therefore
\[ \limsup_n \int_{\Omega} |f_n|^p - |f_n - f|^p - |f|^p | \, d\mu \leq \varepsilon 2^{p+1}C, \]
i.e., the desired result. \( \square \)

- **Remark 1.36.** In particular, if \( f_n \) converges to \( f \) in measure for every measurable set of finite measure and \( \|f_n\|_p \to \|f\|_p \) then \( \|f_n - f\|_p \to 0 \). \( \square \)

Also, we may generalize Definition 1.28 to \( L^p \), with \( 1 \leq p < \infty \), as follows:

**Definition 1.37.** Let \( \{f_i : i \in I\} \) be a family of measurable functions almost everywhere finite (or elements in \( L^0 \)). If for every \( \varepsilon > 0 \) there exist \( \delta > 0 \) and \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) such that
\[ \int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \, d\mu + \int_{A^c} |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I, \]
then the family is called \( \mu \)-uniformly integrable of order \( p \), for \( 0 < p < \infty \). \( \square \)

Actually, this means that a family \( \{f_i : i \in I\} \) of measurable functions almost everywhere finite is \( \mu \)-uniformly integrable (or \( \mu \)-equicontinuous) of order \( p \) if and only if \( \{\|f_i\|^p : i \in I\} \) is \( \mu \)-uniformly integrable (or \( \mu \)-equicontinuous).

**Theorem 1.38.** Let \( \{f_i : i \in I\} \) be a family of measurable functions almost everywhere finite in a measure space \( (\Omega, \mathcal{F}, \mu) \). Then the following statements are equivalent:

(1) \( \{f_i : i \in I\} \) are \( \mu \)-uniformly integrable of order \( p \);

(2) for any \( \varepsilon > 0 \) there exists a nonnegative \( p \)-integrable function \( g \) such that
\[ \int_{\{|f_i| \geq g\}} |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I; \]

(3) (a) there exists a constant \( C > 0 \) such that
\[ \int_{\Omega} |f_i|^p \, d\mu \leq C, \quad \forall i \in I, \]
and (b) for every $\varepsilon > 0$ there exist a constant $\delta > 0$ and a nonnegative $p$-integrable function $h$ such that for every $F \in \mathcal{F}$

$$\int_F h^p \, d\mu < \delta \quad \text{implies} \quad \int_F |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I.$$ 

Proof. (1) $\Rightarrow$ (2): Given $\varepsilon > 0$ choose $\delta > 0$ and $A \in \mathcal{F}$ as in Definition 1.37 and set $g = (1/\delta)1_A$. By means of the inequality

$$1_{\{|f_i| \geq g\}} |f_i|^p \leq 1_A |f_i|^p + 1_{\{|f_i| \geq 1/\delta\}} |f_i|^p,$$

we obtain

$$\int_{\{|f_i| \geq g\}} |f_i|^p \, d\mu \leq \int_{A^c} |f_i|^p \, d\mu + \int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \, d\mu \leq \varepsilon, \quad \forall i \in I.$$

and because $\mu(A) < \infty$ the function $g$ is $p$-integrable.

(2) $\Rightarrow$ (3): For every nonnegative $p$-integrable function $g$ and every $F \in \mathcal{F}$ we have

$$\int_F |f_i|^p \, d\mu = \int_{F \cap \{|f_i| \geq g\}} |f_i|^p \, d\mu + \int_{F \cap \{|f_i| < g\}} |f_i|^p \, d\mu \leq$$

$$\leq \int_{\{|f_i| \geq g\}} |f_i|^p \, d\mu + \int_F |g|^p \, d\mu.$$

Hence, for $F = \Omega$ and $g$ as in (2) for $\varepsilon = 1$ we get (3) (a). Similarly, taking $g$ as in (2) for $\varepsilon/2$ and $h = g$, we deduce (3) (b).

(3) $\Rightarrow$ (1): Given $\varepsilon > 0$ choose $\delta > 0$ and $h \geq 0$ as in (3) (b). Define $A_r = \{h \leq r\}$ to check that

$$r^p \mu(A_r) \leq \int_{A_r} h^p \, d\mu \leq \int_{\Omega} h^p \, d\mu < \infty,$$

which means that $A_r$ has finite measure for every $r > 0$. Moreover, on the complement,

$$\int_{A_r^c} h^p \, d\mu = \int_{\Omega} h^p 1_{h > r} \, d\mu \to 0 \quad \text{as} \quad r \to \infty.$$

Hence, if $r$ is sufficiently large then take $A = A_r$ to deduce that the condition (3) (b) yields

$$\int_{A^c} h^p \, d\mu < \delta \quad \text{implies} \quad \int_{A^c} |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I,$$

i.e., one part Definition 1.37 of $\mu$-integrability of order $p$. Next, because $h$ is $p$-integrable, there exists $\delta' > 0$ such that

$$\mu(F) < \delta' \quad \text{implies} \quad \int_F h^p \, d\mu < \delta.$$
Thus, take $C$ as in (3) (a) to check the inequality
\[ r\mu(\{|f_i| \geq r\}) \leq \int_{\{|f_i| \geq r\}} |f_i| \text{d}\mu \leq \int_{\Omega} |f_i| \text{d}\mu \leq C, \quad \forall i \in I. \]

Now, if $r$ sufficiently large so that $C/r \leq \delta'$ then $\mu(\{|f_i| \geq r\}) < \delta'$, and the condition (3) (b) with the set $F = \{|f_i| \geq r\}$ yields
\[ \int_{\{|f_i| \geq r\}} |f_i| \text{d}\mu \leq \varepsilon, \]
proving the $\mu$-integrability of order $p$.

Alternatively, the proof may continuous as follows:

(3) $\Rightarrow$ (2): Given $\varepsilon > 0$ choose $\delta > 0$ and $h \geq 0$ as in (3) (b). If $C$ is as in (3) (a), then the inequality
\[ a^p \int_{\{|f_i| \geq ah\}} |h|^p \text{d}\mu \leq \int_{\{|f_i| \geq ah\}} |f_i|^p \text{d}\mu \leq \int_{\Omega} |f_i|^p \text{d}\mu \leq C, \quad \forall a > 0, \]
shows we can select $a$ sufficiently large so that
\[ \int_{\{|f_i| \geq ah\}} |h|^p \text{d}\mu \leq \frac{C}{a^p} \leq \delta. \]

Hence, the condition (3) (b) with $F = \{|f_i| \geq ah\}$ yields
\[ \int_{\{|f_i| \geq ah\}} |f_i|^p \text{d}\mu \leq \varepsilon, \quad \forall i \in I, \]
i.e., we deduce (2) with $g = ah$.

(2) $\Rightarrow$ (1): Given $\varepsilon > 0$ find $g$ as in (2). Thus, the inequality
\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \text{d}\mu = \int_{\{|f_i| \geq 1/\delta\} \cap \{|f_i| \geq g\}} |f_i|^p \text{d}\mu + \\
+ \int_{\{|f_i| \geq 1/\delta\} \cap \{|f_i| < g\}} |f_i|^p \text{d}\mu \leq \\
\leq \int_{\{|f_i| \geq g\}} |f_i|^p \text{d}\mu + \int_{\{g \geq 1/\delta\}} g^p \text{d}\mu.
\]
proves that
\[ \int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \text{d}\mu \leq \varepsilon + \int_{\{g \geq 1/\delta\}} |g|^p \text{d}\mu, \quad \forall i \in I, \]
and since
\[ \lim_{\delta \to 0} \int_{\{g \geq 1/\delta\}} g^p \text{d}\mu = 0, \quad \forall g \in L^p, \]
we can find $\delta > 0$ such that
\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i|^p \, d\mu \leq 2\varepsilon, \quad \forall i \in I.
\]
Also the set $A_r = \{ g \geq r \}$ has finite measure for every $r > 0$, and
\[
\int_{A^c} |f_i|^p \, d\mu = \int_{A^c \cap \{|f_i| \geq g\}} |f_i|^p \, d\mu + \int_{A^c \cap \{|f_i| < g\}} |f_i|^p \, d\mu \leq
\int_{\{|f_i| \geq g\}} |f_i|^p \, d\mu + \int_{\{g < r\}} g^p \, d\mu
\]
ensures that there exists $A = A_r$, for some $r > 0$, such that
\[
\int_{A^c} |f_i|^p \, d\mu \leq 2\varepsilon.
\]
Hence, the family $\{f_i : i \in I\}$ is $\mu$-uniformly integrable of order $p$.

Remark 1.39. Note that a measure space $(\Omega, \mathcal{F}, \mu)$ is $\sigma$-finite if and only if there exists a strictly positive integrable function $h$. Indeed, if $\mu$ is $\sigma$-finite then there exists an increasing sequence $\{\Omega_k\} \subset \mathcal{F}$ such that $\Omega = \bigcup_k \Omega_k$ and $0 < \mu(\Omega_k) < \infty$. Thus, the function $h = \sum_k (2^{-k} / \mu(\Omega_k)) 1_{\Omega_k} > 0$ is integrable, for every $p$. Conversely, if there exists a strictly positive integrable function $h$ then the sets $\Omega_k = \{ h \geq 1/k \}$ satisfy the required condition. Moreover, if $h > 0$ and integrable then $h^{1/p}$ is strictly positive and $p$-integrable.

The following result applies for $\sigma$-finite measure spaces.

Corollary 1.40. Let $h$ be a strictly positive $p$-integrable function on a measure space $(\Omega, \mathcal{F}, \mu)$. Then we can revise the statements in Theorem 1.38 as follows: (2) becomes: for every $\varepsilon > 0$ there exists $\alpha > 0$ such that
\[
\int_{\{|f_i| \geq \alpha h\}} |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I;
\]
and (3) (b) reads as: for every $\varepsilon > 0$ there exist a constant $\delta > 0$ such that for every $F \in \mathcal{F}$
\[
\int_F h^p \, d\mu < \delta \quad \text{implies} \quad \int_F |f_i|^p \, d\mu < \varepsilon, \quad \forall i \in I.
\]
The equivalence among properties (1), (2) and (3) remains true.

Proof. The inequalities
\[
\int_{\{|f_i| \geq \alpha h\}} |f_i|^p \, d\mu = \int_{\{|f_i| \geq \alpha h\} \cap \{|f| \geq g\}} |f_i|^p \, d\mu + \int_{\{|f_i| \geq \alpha h\} \cap \{|f| < g\}} |f_i|^p \, d\mu \leq
\int_{\{|f| \geq g\}} |f| \, d\mu + \int_{\{|g| \geq \alpha h\}} |g|^p \, d\mu
\]
and
\[ \int_F |f_i|^p \, d\mu \leq \int_{\{|f_i| \geq \alpha h\}} |f_i|^p \, d\mu + \alpha^p \int_F |h|^p \, d\mu \]
provide the equivalence.

Note that if \( \mu(\Omega) < \infty \) then we can take \( A = \Omega \) in the Definition 1.37, i.e., \( g = 1/\delta \) and \( h = 1 \) in conditions (2) and (3) of Theorem 1.38 and Corollary 1.40. For instance, the interested reader may consult the books by Bauer [15, Section 21, pp. 121–131], among others.

Also we have a practical criterium to check the \( \mu \)-uniformly integrability of order \( q \), compare with (9) in Section 1.2.1.

**Proposition 1.41.** Let \( \{f_i : i \in I\} \) be a family of measurable functions equi-bounded on \( L^p(\Omega, \mathcal{F}, \mu) \) and \( L^p(\Omega, \mathcal{F}, \mu) \), for some \( 0 < r < p < \infty \), i.e., there exists a constant \( C > 0 \) such that
\[
\left( \int_{\Omega} |f_i|^p \, d\mu \right) \leq C^p, \quad \left( \int_{\Omega} |f_i|^r \, d\mu \right) \leq C^r \quad \forall i \in I.
\]
Then for any \( q \) in \( (r, p) \) and for every \( \varepsilon > 0 \) there exist \( \delta > 0 \) and a measurable set \( A \) with \( \mu(A) < \infty \) such that
\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i|^q \, d\mu + \int_{\Omega \setminus A} |f_i|^q \, d\mu < \varepsilon, \quad \forall i \in I,
\]
i.e., the family \( \{f_i : i \in I\} \) is \( \mu \)-uniformly integrable of order \( q \).

**Proof.** Indeed, because the family is bounded in \( L^p \), it has to be a family of measurable functions taking finite values almost everywhere and the set \( \{|f_i| \geq 1/\delta\} \) has finite \( \mu \)-measure for every \( \delta > 0 \) and \( i \) in \( I \).

Now, write \( q = sp \) with \( 0 < s < 1 \) to deduce
\[
|f_i|^q |f_j|^{(1-s)r} \mathbb{1}_{\{|f_j| \geq 1/\delta\}} \leq |f_i|^{sp} |f_j|^{(1-s)r}, \quad \forall i, j \in I,
\]
and then apply Hölder inequality to obtain
\[
\delta^{-(1-s)r} \int_{\{|f_j| \geq 1/\delta\}} |f_i|^q \, d\mu \leq \left( \int_{\Omega} |f_i|^p \, d\mu \right)^s \left( \int_{\Omega} |f_j|^r \, d\mu \right)^{1-s} \leq C.
\]
Hence, for a given \( \varepsilon > 0 \) choose \( \delta > 0 \) so that \( C\delta^{(1-s)r} < \varepsilon/2 \). Next, fix \( j \) in \( I \) and take \( A = \{|f_j| \geq 1/\delta\} \), which has finite \( \mu \)-measure and satisfies
\[
\int_{\Omega \setminus A} |f_i|^q \, d\mu = \int_{\{|f_j| \geq 1/\delta\}} |f_i|^q \, d\mu \leq C\delta^{(1-s)r} < \varepsilon/2.
\]
Similarly, take \( i = j \) to deduce
\[
\int_{\{|f_i| \geq 1/\delta\}} |f_i|^q \, d\mu \leq C\delta^{(1-s)r} < \varepsilon/2,
\]
proving the \( \mu \)-uniform integrability of order \( q \). □
To complete this section, we show a relation of totally bounded (or pre-compact) sets in $L^p$ and uniformly integrable sets of order $p$. Recall that a family of functions $\{f_i: i \in I\}$ is a totally bounded subset of $L^p(\Omega, \mathcal{F}, \mu)$ if for every $\varepsilon > 0$ there exists a finite subset of indexes $J \subset I$ such that for every $i$ in $I$ there exists $j$ in $J$ satisfying $\|f_i - f_j\|_p < \varepsilon$, i.e., any element in $\{f_i: i \in I\}$ is within a distance $\varepsilon$ from the finite set $\{f_j: j \in J\}$. Sometimes $\{f_j: j \in J\}$ is called an $\varepsilon$-net relative to $\{f_i: i \in I\}$. This concept of totally bounded sets is equivalent to pre-compact set on a complete metric space, in particular, this also applied to the topological vector space $L^p(\Omega, \mathcal{F}, \mu)$ with $0 < p < 1$ and the distance $d(f,g) = \|f - g\|_p^p$.

**Proposition 1.42.** Let $\{f_i: i \in I\}$ be a totally bounded subset of $L^p(\Omega, \mathcal{F}, \mu)$, with $0 < p < \infty$. Then $\{f_i: i \in I\}$ is $\mu$-uniformly integrable of order $p$.

**Proof.** For a given $\varepsilon > 0$, denote by $J_\varepsilon \subset I$ the finite subset of indexes obtained from the totally boundedness property. We assume $1 \leq p < \infty$ to be able to use the triangular inequality $\|f - g\|_p \leq \|f\|_p + \|g\|_p$. The case where $0 < p < 1$ is treated analogously, by means of the inequality $\|f - g\|_p^p \leq \|f\|_p^p + \|g\|_p^p$.

Since the finite family $\{f_j: j \in J_\varepsilon\}$ is $\mu$-equicontinuous (also $\mu$-uniformly integrable) of order $p$, for this same $\varepsilon > 0$ there exists $\delta > 0$ and $A$ in $\mathcal{F}$ such that

$$F \in \mathcal{F} \text{ with } \mu(F) < \delta \implies \int_F |f_j|^p \, d\mu \leq \varepsilon, \quad \forall j \in J_\varepsilon,$$

and

$$\int_{A\varepsilon} |f_j|^p \, d\mu < \varepsilon, \quad \forall j \in J_\varepsilon,$$

which combined with the inequality

$$\inf \{\|f_i - f_j\|_p: j \in J_\varepsilon\} \leq \varepsilon, \quad \forall i \in I,$$

show that $\{|f_i|^p: i \in I\}$ is $\mu$-equicontinuous.

Now, we redo the argument to show that $\mu$-equicontinuity plus uniformly bounded is equivalent to $\mu$-uniformly integrability. Indeed, the family of functions $\{f_i: i \in I\}$ is uniformly bounded in $L^p$, namely

$$\|f_i\|_p \leq \|f_i - f_j\|_p + \|f_j\|_p \leq \varepsilon + \sup \{\|f_j\|_p: j \in J_\varepsilon\} < \infty,$$

and the inequality

$$\mu(\{|f_i| \geq c\}) \leq \frac{1}{c^p} \int_\Omega |f_i|^p \, d\mu \leq \frac{1}{c^p} \sup_{i \in I} \|f_i\|_p^p,$$

shows that for every $\delta > 0$ there exists $c > 0$ (sufficiently large) so that the set $F_{i,c} = \{|f_i| \geq c\}$ satisfies $\mu(F_{i,c}) < \delta$, for every $i$. Hence, by taking $F = F_{i,c}$ within the $\mu$-equicontinuity condition for the whole family $\{|f_i|^p: i \in I\}$, we deduce that $\{f_i: i \in I\}$ is also $\mu$-uniformly integrable of order $p$. \qed
Certainly, the converse of Proposition 1.42 fails. For instance, the sequence \(\{f_n(x) = \sin nx\}\) on \(L^2(\pi, \pi]\) satisfies \(||f_n||_2 = 2\pi\) so that any \(L^2\)-convergent subsequence would converge to a some function \(f\) with \(||f||_2 = 2\pi\). However, Riemann-Lebesgue Theorem (e.g., see part I) affirms that \(\langle f_n, g \rangle \to 0\) for every \(g\) in \(L^1\), which means that the sequence cannot be totally bounded in \(L^2(\pi, \pi]\). Nevertheless, because \(|f_n(x)| \leq 1\), this sequence is \(\mu\)-uniformly integrable of any order \(p\).

An important role is played by the weak convergence in \(L^1\), i.e., when \(\langle f_n, g \rangle \to \langle f, g \rangle\) for every \(g\) in \(L^\infty\). Actually, we have the Dunford-Pettis criterion: a set \(\{f_i : i \in I\}\) is sequentially weakly pre-compact in \(L^1(\Omega, \mathcal{F}, \mu)\) if and only if it is \(\mu\)-uniformly integrable (a partial proof for the case of a finite measure can be found in Meyer [90, Section II.2, T23, pp. 39-40]). However, any bounded set in \(L^p(\Omega, \mathcal{F}, \mu), 1 \leq p < \infty\), is weakly pre-compact. This is a general result (Alaoglu’s Theorem) valid for any reflexive Banach space, e.g., see Conway [29, Section V.3 and V.4, pp. 123–137]. We delay this discussion until later, in Section 2.4.

**Exercise 1.8.** Consider the Lebesgue measure on the interval \((0, \infty)\) and define the functions \(f_i = (1/i)\mathbf{1}_{(i, 2i]}\) and \(g_i = 2^i\mathbf{1}_{(2^{i-1}, 2^i)}\) for \(i \geq 1\). Prove that (a) the sequence \(\{f_i : i \geq 1\}\) is uniformly integrable of any order \(p > 1\), but not of order \(0 < p \leq 1\). On the contrary, show that (b) the sequence \(\{g_i : i \geq 1\}\) is uniformly integrable of any order \(0 < p < 1\), but the sequence is not equi-integrable of any order \(p \geq 1\). \(\square\)

### 1.3 Vector-valued Integrals

First let take a look at vector-valued measures. As mentioned early, the \(\sigma\)-additivity concept can be extended to vector-valued set functions, i.e., for a set function \(\mu : \mathcal{A} \to \mathbb{R}^n\), we can extend the \(\mu(\sum_n A_n) = \sum_n \mu(A_n)\) for any disjoint sequence \(\{A_n\}\) in the \(\sigma\)-algebra \(\mathcal{A}\), where the convergence of the series is one of the assumptions. Notice that in the above definition, we may use any other topological vector space \(V\) instead of \(\mathbb{R}^d\) and that the vector-valued measure \(\mu\) does not take any “infinite” values. If \(V = \mathbb{R}^d\) with norm \(||\cdot||\) then we may define the variation of \(\mu\) as \(|\mu|(A) = \sup\{\sum_n \|\mu(A_n)\| : A = \sum_n A_n\}\). As an exercise, with the technique of signed-measures (and looking at each coordinate of \(\mathbb{R}^d\)), we can show that (1) the series \(\sum_n \mu(A_n)\) is absolutely convergent for any disjoint sequence \(\{A_n\}\) of \(\mathcal{A}\); (2) the variation of \(\mu\) is a \(\sigma\)-additive finite (real-valued) measure on \(\mathcal{A}\). For instance, the reader may check the books Dinculeanu [33], Ma [84] or Panchapagesan [98] for greater details and applications. Also the reader may find convenient to read Cohn [28, Appendix E, pp. 350-357] and Diekel and Uhl [32].

Given two measurable spaces \((\Omega, \mathcal{F})\) and \((E, \mathcal{E})\) we may consider the space of measurable (or \(\mu\)-measurable) functions from \(\Omega\) into \(E\), denoted by \(L^0(\Omega, \mathcal{F}; E)\) or its complete version \(L^0(\Omega, \mathcal{F}^\mu; E)\), where \(\mathcal{F}^\mu\) is the \(\mu\)-completion of \(\mathcal{F}\). When \(E\) is a nonseparable metric space, we add the condition “separable range” into
the definition of measurability. In this way, we are able to approximate measurable functions by a sequence of simple functions (i.e., measurable functions assuming a finite number of values) converging almost everywhere.

In general we ignore the subtle differences between \((\mu\text{-})\) measurable functions and the classes of functions defining the (quotient) space \(L^0(\Omega, \mathcal{F}, \mu; E)\) of almost measurable \(E\)-valued functions which is a complete metrizable space under the convergence in measure (provided \(E\) is a metrizable space), and a vector space (provided \(E\) is so), but not quite a topological vector space if \(\mu\) is not a finite measure. Similarly, \(S^0(\Omega, \mathcal{F}; E)\) or \(S^0(\Omega, \mathcal{F}, \mu; E)\) denotes the spaces of \(\mathcal{F}\)-measurable or almost measurable \(E\)-valued simple functions.

Sometimes, a \(E\)-valued function \(f\) is called strongly measurable if it is almost everywhere limit of a sequence of simple functions, and weakly measurable if the real-valued function \(x \mapsto \langle e, f \rangle\) is measurable for any \(e\) belonging to the topological dual of \(E\). It is proven (Yosida [135, Section V.4, pp. 130-132]) that a function is strongly measurable if and only if it is weakly measurable with separable range.

1.3.1 Metric Space of Measurable Functions

Clearly, we use \(E = \mathbb{R}^n, n \geq 1\) or \(\mathbb{R} = [-\infty, +\infty]\), or in general a (complete) metric (or Banach) space \(E\) with its Borel \(\sigma\)-algebra \(\mathcal{E}\).

**Proposition 1.43.** Let \((\Omega, \mathcal{F}, \mu)\) be a finite measure space and define \(d\) by

\[
d(f, g) = \int_{\Omega} \frac{d_E(f, g)}{1 + d_E(f, g)} \, d\mu \quad \text{or} \quad d(f, g) = \int_{\Omega} \min\{d_E(f, g), 1\} \, d\mu.
\]

Then we can show that (1) the function \((f, g) \mapsto d(f, g)\) is a metric on \(L^0 = L^0(\Omega, \mathcal{F}, \mu; E)\); (2) we have \(d(f_n, f) \to 0\) if and only if \(f_n \to f\) in measure; (3) the metric \(d\) is complete in \(L^0\) if \(d_E\) is complete in \(E\). Moreover, (4) if \(\mu\) is only \(\sigma\)-finite, then the convergence in measure on every finite subset is metrizable.

**Proof.** (1) We need to verify that \(d_E(f, g) \geq 0\), the equality only when \(f = g\) a.e., and the triangular inequality. Only the last point requires some consideration. Since the function \(q : r \mapsto r/(1+r)\) (or \(q : r \mapsto \min\{r, 1\}\), \(r \geq 0\), is nonnegative, increasing, bounded (by 1) and sub-linear; for any measurable functions \(f, g, h\), almost everywhere defined, the inequality

\[
d_E(f(x), g(x)) \leq d_E(f(x), h(x)) + d_E(h(x), g(x))
\]

implies

\[
q[d_E(f(x), g(x))] \leq q[d_E(f(x), h(x)) + d_E(h(x), g(x))] \leq q[d_E(f(x), h(x))] + q[d_E(h(x), g(x))],
\]

which yields the triangular inequality \(d(f, g) \leq d(f, h) + d(h, g)\) after integration.
(2 \iff) Let \( f_n \to f \) in measure then for every \( \delta > 0 \) and \( \varepsilon > 0 \) there exists \( N \) such that \( n > N \) implies
\[
\mu(\{d_E(f_n,f) > \delta\}) = \mu(\{x : d_E(f_n(x), f(x)) > \delta\}) < \varepsilon.
\]
Thus, for any \( n > N \) we have
\[
d(f_n, f) = \int_{\Omega} q[d_E(f_n(x), f(x))] \mu(dx) \leq
\leq \int_{\Omega} 1_{\{d_E(f_n(x), f(x)) > \delta\}} \mu(dx) + \delta \int_{\Omega} 1_{\{d_E(f_n(x), f(x)) \leq \delta\}} \mu(dx) \leq
\leq \mu(\{d_E(f_n,f) > \delta\}) + \delta \mu(\{d_E(f_n,f) \leq \delta\}) < \varepsilon + \delta \mu(\Omega).
\]
Since \( \varepsilon \) and \( \delta \) are arbitrary, we deduce that \( d(f_n, f) \to 0 \).

(2 \implies) If \( d(f_n, f) \to 0 \) then for any \( \delta > 0 \) we use the fact that the function \( r \to q(r) \) is increasing to get
\[
d(f_n, f) \geq \int_{\Omega} q[d_E(f_n(x), f(x))] 1_{\{d_E(f_n(x), f(x)) > \delta\}} \mu(dx) \geq
\geq q(\delta) \mu(\{x : d_E(f_n(x), f(x)) > \delta\}).
\]
Since \( d(f_n, f) \to 0 \) we deduce \( \mu(\{x : d_E(f_n(x), f(x)) > \delta\}) \to 0 \).

(3) We can apply Lebesgue dominate convergence Theorem 1.18 to conclude.

(4) If \((\Omega, \mathcal{F}, \mu)\) is a \( \sigma \)-finite measure space then \( \Omega = \sum_n \Omega_n \) for a disjoint sequence \( \{\Omega_n : n \geq 1\} \) of measurable sets with \( \mu(\Omega_n) < \infty \). Thus, we can define
\[
d(f,g) = \sum_n \frac{2^{-n}}{\mu(\Omega_n)} \int_{\Omega_n} q[d_E(f(x), g(x))] \mu(dx),
\]
to get a metric in \( L^0 \) which yields a convergence equivalent to the convergence in measure on every set of finite measure.

\[
\square
\]

1.3.2 With Values in a Banach Space

If \((E, |\cdot|_E)\) is a Banach space then we can consider the spaces (1) \( S^1 = S^1(\Omega, \mathcal{F}, \mu; E) \subset L^0 \) the subspace of all simple functions with finite-measure support, i.e., almost measurable functions \( f \) assuming a finite number of values and satisfying \( \mu(\omega \in \Omega : f(\omega) \neq 0) < \infty \); and (2) \( L^\infty = L^\infty(\Omega, \mathcal{F}, \mu; E) \subset L^0 \) of all almost measurable, with \( \sigma \) finite support, and almost bounded functions, i.e., an equivalence class can be regarded as a function \( f \) defined outside of a negligible set \( N \) with \( f(\Omega \setminus N) \) contained into the closure of a countable bounded subset of \((E, |\cdot|_E)\) and \( \{\omega \in \Omega \setminus N : f(\omega) \neq 0\} \) is a \( \sigma \)-finite set (i.e., a countable union of sets of finite \( \mu \)-measure). The elements in \( L^\infty \) are called essentially bounded measurable functions and with the essential sup-norm defined by
\[
\|f\|_\infty = \inf \{C \geq 0 : |f(\omega)|_E \leq C, \text{ a.e. } \omega\},
\]

it is a Banach space.

Now, define the vector spaces \( \mathcal{L}^1 = \mathcal{L}^1(\Omega, \mathcal{F}, \mu; E) \) of all functions \( f \) in \( \mathcal{L}^0(\Omega, \mathcal{F}, \mu; E) \) such that
\[
\|f\|_1 = \int_\Omega |f(\omega)|_E \mu(d\omega) < \infty.
\]

Similarly, using the equivalence classes, we define the vector space of classes of functions \( L^1(\Omega, \mathcal{F}, \mu; E) \subset L^0(\Omega, \mathcal{F}, \mu; E) \).

**Proposition 1.44.** Let \( (\Omega, \mathcal{F}, \mu) \) be a measure space and \( (E, |\cdot|_E) \) be a Banach space. Then \( (L^1(\Omega, \mathcal{F}, \mu; E), |\cdot|_1) \) is Banach space. Moreover, the set \( S^1(\Omega, \mathcal{F}, \mu; E) \) is a dense subspace. Furthermore, if \( \mu \) is a regular Borel measure, \( \Omega \) is a Polish space and \( E \) is separable then the space \( L^1(\Omega, \mathcal{F}, \mu; E) \) is also separable.

**Proof.** Let \( \{f_n : n \geq 1\} \) be a Cauchy sequence in \( L^1 \). Based on the estimate
\[
\varepsilon \mu(\{x \in \Omega : |f(x)|_E \geq \varepsilon\}) \leq \int_\Omega |f(x)|_E \mu(dx), \quad \forall \varepsilon > 0,
\]
we obtain that \( \{f_n : n \geq 1\} \) is also a Cauchy sequence in measure, and Lebesgue dominate convergence Theorem 1.18 yields a measurable function \( f \) such that \( f_n \to f \) in measure (and almost everywhere through a subsequence). Thus, for every \( m \) and \( n \), integrate the inequality
\[
|f_n(\omega) - f(\omega)|_E \leq |f_n(\omega) - f_m(\omega)|_E + |f_m(\omega) - f(\omega)|_E, \quad \text{a.e. } \omega \in \Omega,
\]
to get
\[
\|f_n - f\|_1 \leq \varepsilon + \int_\Omega |f_m(\omega) - f(\omega)|_E \mu(d\omega),
\]
for any \( \varepsilon > 0 \), provided \( m, n \geq N(\varepsilon) \).

Now, if \( \|f_n - f\|_1 \) does not vanish as \( n \to 0 \) then there exists a sequence of integers \( \{n'\} \) and \( \delta > 0 \) such that \( \|f_{n'} - f\|_1 \geq \delta \), for every \( n' \). Again, by the Lebesgue dominate Theorem 1.18, there is a subsequence, denote by \( \{f_{n''}\} \) such that \( |f_{n''}(\omega) - f(\omega)|_E \to 0 \) almost everywhere in \( \omega \). Hence, by means of Fatou’s Lemma 1.17, we may take \( m = n'' \to \infty \) to deduce \( \|f_n - f\|_1 \leq \varepsilon \), for every \( n \geq N(\varepsilon) \), i.e., \( L^1 \) is complete.

To check that \( S^1 \) is dense in \( L^1 \), we construct a sequence \( \{f_n : n \geq 1\} \subset S^1 \) such that \( |f_n(\omega) - f(\omega)|_E \to 0 \) and \( |f_n(\omega) - f(\omega)|_E \leq |f(\omega)|_E \), almost everywhere in \( \omega \). Hence, Lebesgue dominate convergence Theorem 1.18 shows that \( \|f_n - f\|_1 \to 0 \).

If \( \Omega \) is separable and \( \mu \) a regular Borel measure, then that there exists a countable basis of open sets \( \mathcal{O} \) such that for every \( \varepsilon > 0 \) and any \( F \in \mathcal{F} \) with \( \mu(F) < \infty \) there exists \( O \in \mathcal{O} \) such that \( \|1_F - 1_O\|_1 < \varepsilon \). Choose a countable dense set \( \{e_0, e_1, \ldots\} \in E \) with \( e_0 = 0 \), and consider the family \( S_0 \) of all functions \( \varphi \) in \( S^1 \) taking a.e. finite number of values \( \{e_0, \ldots, e_n(\varphi)\} \).
satisfying \( \varphi^{-1}(e_i) \in \mathcal{O} \), for \( i \geq 1 \). The family \( S_0 \) is countable and for any \( f \) in \( S^1 \) there exists a sequence \( \{ \varphi_n \} \subset S_0 \) such that \( \| \varphi_n - f \|_1 \to 0 \). Hence, \( S_0 \) is dense in \( S^1(\Omega, \mathcal{F}, \mu; E) \), which dense in \( L^1(\Omega, \mathcal{F}, \mu; E) \).

Now, everything is in place to give a meaning to the integral of \( f \) in the space \( L^1(\Omega, \mathcal{F}, \mu; E) \) as an element in the Banach space \( E \). For an almost everywhere simple function \( \varphi = \sum_{i=1}^{n} e_i 1_{A_i} \), with \( A_i \) disjoint, the expression

\[
\int_{\Omega} \varphi(\omega) \mu(d\omega) = \sum_{i=1}^{n} \mu(A_i) e_i
\]

is an element in \( E \), which satisfies

\[
\left| \int_{\Omega} \varphi(\omega) \mu(d\omega) \right|_E = \left| \sum_{i=1}^{n} \mu(A_i) e_i \right|_E \leq \sum_{i=1}^{n} \mu(A_i) |e_i|_E = \int_{\Omega} |\varphi(\omega)|_E \mu(d\omega).
\]

This map is initially defined from \( S^1(\Omega, \mathcal{F}, \mu; E) \subset L^1(\Omega, \mathcal{F}, \mu; E) \) into \( E \), and by linearity and continuity, it can be extended to the whole \( L^1(\Omega, \mathcal{F}, \mu; E) \), which satisfies

\[
\left\| \int_{\Omega} f(\omega) \mu(d\omega) \right\| \leq \int_{\Omega} |f(\omega)|_E \mu(d\omega), \quad \forall f \in L^1,
\]

and is called Bochner’s integral. Actually, we could use this map to define the space \( L^1(\Omega, \mathcal{F}, \mu; E) \), i.e., an \( E \)-valued function \( f \) is called integrable (in the Bochner’s sense) if there exists a sequence \( \{ f_n : n \geq 1 \} \subset S^1(\Omega, \mathcal{F}, \mu; E) \) of almost everywhere simple functions such that \( \|f_n - f\|_1 \to 0 \) and a posteriori, by means of the estimate

\[
\left| \int_{\Omega} f_n(\omega) \mu(d\omega) - \int_{\Omega} f(\omega) \mu(d\omega) \right|_E \leq \|f_n - f\|_1,
\]

the limit

\[
\int_{\Omega} f(\omega) \mu(d\omega) = \lim_{n} \int_{\Omega} f_n(\omega) \mu(d\omega), \quad \text{in } E,
\]

defines the integral. In any case, an almost measurable \( E \)-valued function \( f \) is integrable if and only if \( |f|_E \) is integrable.

**Remark 1.45.** Sometimes, we prefer to define the integral indirectly as follows. Let \( e \) be in \( E \) and \( e' \) be an element in the dual space \( E' \). If \( E'' \) denotes the double dual then the mapping \( J : E \to E'', \quad \langle e', J(e) \rangle = \langle e, e' \rangle \), defines a continuous inclusion. For any simple function \( \varphi = \sum_{i=1}^{n} e_i 1_{A_i} \) the expression

\[
\langle e', I(\varphi) \rangle = \int_{\Omega} \langle e', \varphi(\omega) \rangle \mu(d\omega) = \sum_{i=1}^{n} \langle e', e_i \rangle \mu(A_i)
\]
defines \( I(\varphi) \) as an element in the image \( J(E) \). Since \( J(E) \subset E'' \) is closed and
\[
|I(\varphi)|_{E''} \leq \sum_{i=1}^n |e_i|_E \mu(A_i) = \|\varphi\|_1.
\]
there exists a unique extension to a mapping \( I : L^1 \to J(E) \). Hence the integral of an element \( f \) in \( L^1(\Omega, \mathcal{F}, \mu; E) \) is defined as
\[
\int_{\Omega} f(\omega) \, d\mu(\omega) = J^{-1}(I(f)), \quad \text{and satisfies}
\]
\[
\left\| \int_{\Omega} f(\omega) \, d\mu(\omega) \right\| \leq \int_{\Omega} |f(\omega)|_E \, d\mu(\omega) = \|f\|_1.
\]
This methods works even for weakly almost measurable functions, i.e., when \( \omega \mapsto \langle e', f(\omega) \rangle \) is almost measurable, for every \( e' \) in the dual space \( E' \).

Given a metrizable space \( \Omega \) and a Banach space \( (E, |\cdot|_E) \), denote by \( C^0_b = C^0_b(\Omega; E) \) the space of continuous and bounded functions from \( \Omega \) into \( E \) endowed with the sup-norm
\[
\|f\|_\infty = \sup \{|f(\omega)|_E : \omega \in \Omega\}.
\]
It is clear that we have \( C^0_b \subset L^\infty \). The support of a continuous function \( f \) is the closure of the set \( \{f(\omega) \neq 0\} \) denoted by \( \text{supp}(f) \). The subspace \( C^0 = C^0(\Omega; E) \) of all continuous functions with compact support makes perfectly sense when \( \Omega \) is also locally compact, and Tietze’s extension can be used to show that given any compact set \( K \subset \Omega \) and a continuous function defined on \( f|_K \) defined on \( K \) we can extend \( f|_K \) to an element \( f \) in \( C^0(\Omega, E) \). In general, if \( \Omega \) is not locally compact then the extension function may belongs only to \( C^0_0 \), with a support not necessarily compact. In any case, it is clear that \( C^0_0 \subset L^1(\Omega, \mathcal{F}, \mu) \) if \( \mu \) is a Radon measure, i.e., if \( \mu(K) < \infty \) for any compact set \( K \) of \( \Omega \).

**Proposition 1.46.** Let \( \Omega \) be a Polish space, \( \mu \) be a regular Borel measure on \( (\Omega, \mathcal{F}) \), and \( E \) a Banach space. Then \( C^0(\Omega; E) \cap L^1 \) is dense in \( L^1 = L^1(\Omega, \mathcal{F}, \mu; E) \). Moreover, if \( \Omega \) is locally compact and \( \mu \) is a Radon measure then \( C^0_0(\Omega; E) \) is dense in \( L^1(\Omega, \mathcal{F}, \mu; E) \).

**Proof.** Essentially by construction (or by means of Proposition 1.44) the space \( S^1(\Omega, \mathcal{F}, \mu; E) \) of almost simple functions with finite-measure support (namely, \( \mu(\varphi \neq 0) < \infty \)) is dense in \( L^1(\Omega, \mathcal{F}, \mu; E) \). Hence to show that \( C^0 \cap L^1 \) is dense we have to approximate integrable functions of the type \( 1_F \) with \( \mu(F) < \infty \).

Since \( \Omega \) is also a Polish space, the Borel measure \( \mu \) is inner regular, i.e., there exists a compact set \( K \subset F \) and an open set \( F \subset O \) such that \( \mu(O \setminus K) < \varepsilon \). As in Urlyshon’s extension arguments the function
\[
k_\varepsilon(\omega) = \frac{d_\Omega(\omega, \Omega \setminus O)}{d_\Omega(\omega, K) + d_\Omega(\omega, \Omega \setminus O)}
\]
is continuous and \( \|1_F - k_\varepsilon\|_1 \leq \mu(O \setminus K) \leq \varepsilon \). If \( \Omega \) is locally compact then we may assume that the open set \( O \) has a compact closure \( \overline{O} \), i.e., \( f_\varepsilon \) has a compact support.

**Remark 1.47.** Going back to the dual norm, if \( E \) is a Banach space with dual space \( E' \) (for simplicity first take \( E = \mathbb{R} \) or \( E = \mathbb{R}^d \)) then we consider \( L^p(\Omega, \mathcal{F}, \mu; E) \) and \( L^q(\Omega, \mathcal{F}, \mu; E') \) with \( \frac{1}{p} + \frac{1}{q} = 1 \). The duality pairing becomes

\[
\langle f, g \rangle = \int_\Omega \langle f, g \rangle_{E, E'} \, d\mu, \quad \forall f \in L^p, \ g \in L^q, \ \frac{1}{p} + \frac{1}{q} = 1,
\]

where \( \langle \cdot, \cdot \rangle_{E, E'} \) denotes the duality pairing between \( E \) and \( E' \). In this case, we use the duality mapping \( J: E \rightarrow E' \) satisfying \( |J(f)|_{E'} = |f|_E \). For instance, if \( E \) is a (real) Hilbert space then \( J \) is the identity and \( \langle \cdot, \cdot \rangle_{E, E'} \) is the inner product \( \langle \cdot, \cdot \rangle_{E} \). Therefore, this shows that \( J_p(f) = J(f)|f|^{p-2}\|f\|_p^{1-p} \) maps \( L^p(\Omega, \mathcal{F}, \mu; E) \) into \( L^q(\Omega, \mathcal{F}, \mu; E') \) and \( \|J_p(f)\|_q = \|f\|_p \), for every \( f \). It is clear that the space \( L^q(\Omega, \mathcal{F}, \mu; E') \) is contained into the dual space of \( L^p(\Omega, \mathcal{F}, \mu; E) \), but the equality (valid for \( 1 \leq p < \infty \) follows form Riesz Representation Theorem B.63, any linear continuous function on \( L^p(\Omega, \mathcal{F}, \mu; E) \) can be represented as the duality pairing above, for some \( g \) in \( L^q(\Omega, \mathcal{F}, \mu; E') \).

For instance, the reader may check the book by Cohn [28] and Yosida [135]
Chapter 2

Basic Functional Analysis

There are many well known books on functional analysis at the introductory level, each with particular objective and orientation, e.g., Bachman and Narici [14], Brezis [22] Conway [29], Eidelman et al. [41], Hutson et al. [69], Riesz and Nagy [107], Rudin [109], Swartz [119], Taylor and Lay [123], Yosida [135], among many others. Moreover, for instance, the comprehensive guide (to infinity dimensional analysis) Aliprantis and Border [6] is valuable to some readers. In this chapter, only very elementary concepts are discussed, but trying not to over simplify the material.

2.1 Background and Introduction

Most of this section is usually skipped, but it may serve as a quick review. As expected, several concepts and ideas of the $d$-dimensional Euclidean space are generalized to infinite dimension. For instance, the reader may find useful take a quick look at the book by Carothers [25], for a short course on Banach space theory, among other textbooks.

Perhaps the simplest one are the metric spaces $\langle X, d \rangle$ where the idea of the distance $d$ is used to derived the topology, which includes all the notions of convergence and neighborhood. If the vector structure and the concept of orthogonality are required then an inner product $\langle \cdot, \cdot \rangle$ must be defined. This inner product yields a norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$, with a key property so-called parallelogram identity, namely,

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2, \forall x, y \in X.$$ 

Next, a norm induces a metric $d(x, y) = \|x - y\|$ and so, the topology is generated. Besides completeness (i.e., the property that any Cauchy sequence is convergent), the concept of density (i.e., a subset with closure equals to the whole space) and separability (i.e., existence of a countable dense set) are two highly desirable properties. Completeness (which is a characteristic of the metric, instead of the topology) is essentially required for a metric space. There are
standard ways of completing metric spaces, practically the same argument used to obtain the representation of real numbers as convergence Cauchy sequences.

A vector space with a norm is called a normed space, a complete normed space (or vector space with a complete norm) is called a Banach space and a vector space with a complete norm satisfying the parallelogram identity (or with a complete inner product) is called a Hilbert space. The vector space could be relative to the complex numbers if necessary. Clearly, a Hilbert space is also a Banach space. A vector space with a topology such that the scalar multiplication and the addition are continuous is called topological vector space. Certainly, there are metric vector spaces which are not topological vector space, and yet they may be useful for a certain analysis. As seen in detail later, a vector space with a sufficient large family of seminorms becomes a so-called locally convex topological vector space (in short, lctvs), in particular, a normed space is also a lctvs.

Linear operators are linear mapping between normed spaces (i.e., preserving the linear structure), while linear functional are linear mapping from a normed space into the real (or the complex) numbers. Now, given a continuous linear operator $T : X \to Y$ between two normed spaces $(X, \| \cdot \|_X)$ and $(Y, \| \cdot \|_Y)$ then the expression

$$\|T\| = \sup \left\{ \|Tx\|_Y : \|x\|_X = 1 \right\} = \sup \left\{ \|Tx\|_Y : \|x\|_X \leq 1 \right\}$$

defines the so-called operator norm. Thus, a linear operator is continuous if and only if its (operator) norm $\|T\|$ is finite. In view of this, a continuous linear operator is also called a linear bounded operator. Actually, for a linear operator the following conditions are equivalent: (a) the pre-image of an open set is open, i.e., $\{ x \in X : Tx \in V \}$ is an open set in $X$ for every open set $V$ in $Y$; (b) the pre-image of a neighborhood of the origin in $X$ is a neighborhood of the origin in $Y$; (c) if $x_n \to x$ in $X$ then $Tx_n \to Tx$ in $Y$, i.e., if $\|x_n-x\|_X \to 0$ then $\|Tx_n-Tx\|_Y \to 0$; (d) the condition (c) holds only for $x = 0$, i.e., if $\|x_n\|_X \to 0$ then $\|Tx_n\|_Y \to 0$; (e) $T$ preserves bounded sequences, i.e., if $\sup_n \|x_n\|_X < \infty$ then $\sup_n \|Tx_n\|_Y < \infty$; (f) $T$ maps bounded sets into bounded sets, i.e., if $\sup_{x \in B} \|x\|_X < \infty$ then $\sup_{x \in B} \|Tx\|_Y < \infty$.

The space of all continuous linear operators from $X$ into $Y$ is denoted by $L(X,Y)$, which becomes a normed space with the operator norm. In particular, the case of $Y = \mathbb{R}$ (or $Y = \mathbb{C}$, if the vector space is complex) is called the dual space of $X$ and denoted by $X'$. This process can be repeated to obtain the double dual space $X''$. It is not hard to realize that $X$ can be embedded into $X''$, and if a isomorphism can be established between $X$ and $X''$, then the space $X$ is called reflexive. The Lebesgue spaces $L^p$ are the typical examples, which are reflexive only when $1 < p < \infty$.

Now let us recall the contraction principle for nonlinear mapping $f$ on a complete metric space $(X,d)$, i.e., if $f : X \to X$ satisfies $d(f(x),f(x')) \leq \alpha d(x,x')$ for every $x, x'$ in $X$ and some constant $0 < \alpha < 1$ then (a) there exists a unique fixed point, namely, $x^*$ in $X$ such that $f(x^*) = x^*$, (b) and for any initial point $x_0$, the sequence $\{x_n\}$ defined by induction $x_n = f(x_{n-1})$, converges
to $x^*$. For this to work, metric space $X$ must be complete, but the contraction assumption is only needed eventually, i.e., if $f^n(x) = f^{n-1}(f(x))$ with $f^1 = f$, then the above principle holds true assuming $d(f^n(x), f^n(x')) \leq \alpha_n d(x, x')$ for every $x, x'$ in $X$ with $\sum_n \alpha_n < \infty$.

At this point, it may be interesting to take a quick look at the books by Brown and Pearcy [23], Friedman [46], Halmos [64], Lindenstrauss and Tzafriri [83] or MacCluer [85], among others.

### 2.1.1 Simple Spectral Analysis

The space $L(X)$ of linear continuous operators from a Banach space $X$ into itself endowed with the operator norm $\| \cdot \|$ becomes a Banach space. In this context, it seems better to use $| \cdot |$ instead of $\| \cdot \|_X$ to symbolize the norm in the space $X$, i.e., $\|T\| = \sup_{\|x\| \leq 1} |Tx|$. Moreover, it is also convenient to work with complex-valued linear continuous operator acting on a Banach space over the complex number $\mathbb{C}$, i.e., $| \cdot |$ is a loaded operation in the sense that $|z|$ means the modulus of the complex number $z$ and also $|x| = \| \cdot \|_X$ means the $X$-norm when $x$ is an element of Banach space $X$.

The convergence in the operator norm is sometime called uniform convergence (better say, uniform on bounded sets) in contrast with the so-called strong convergence, which is the pointwise convergence in the $X$-norm, i.e., $|T_n x - Tx| \to 0$ for every $x$ in $X$ means that $T_n \to T$ strongly (or in the strong topology) as $n \to \infty$, while that $\|T_n - T\| \to 0$ means the convergence in norm. It is clear that uniform convergence implies strong convergence, and based on the linearity, if $T_n e_i \to T e_i$ in $X$ for any $e_i$ in $X$, $i = 1, \ldots, n$ then $T_n x \to Tx$ uniformly for any $x$ of the form $x = \sum_{i=1}^n r_i e_i$, $0 \leq r_i \leq 1$, which means that both convergence in norm (or uniformly on bounded set) and strongly (or pointwise in norm) are equivalent if the space $X$ has finite-dimension. Similarly, the fact that the addition and the scalar multiplication are continuous operations in a normed space $X$, any linear operator from $X$ into itself is necessarily continuous over any finite-dimensional subspace. This is the case of $d$-dimensional square matrices regarded as linear continuous operators from $\mathbb{R}^d$ in itself, but the other hand, linear operators densely defined and non necessarily continuous are of key interest in infinite dimensions.

The symbol $I$ denotes the identity operator on $X$ and, for a densely defined linear operator $T$, the resolvent set $\rho(T)$ is the set of complex numbers $\lambda$ for which the range of $\lambda I - T$ is (1) dense in $X$, (2) one-to-one, and (3) the inverse operator $(\lambda I - T)^{-1} = R(\lambda, T)$ (which is called the resolvent operator, a priori only densely defined on $X$) can be extended as a linear continuous operator from $X$ into itself. The spectrum of $T$, denoted by $\sigma(T)$, is the set of all complex numbers not in $\rho(T)$, i.e., $\sigma(T)$ is the complement of $\rho(T)$.

The finite-dimensional spectral analysis is what in algebra is called the canonical representation of a $d$-dimensional square matrix $A$, i.e., first the equation $\det(\lambda I - A) = 0$ gives $k$ distinct roots $\lambda_1, \ldots, \lambda_k$ with multiplicities $m_1, \ldots, m_k$, $m_1 + \cdots + m_k = d$, i.e., each $\lambda_i$ is an eigenvalue with multiplicity $m_k$ and its corresponding eigenspace $M_i$ of dimension $1 \leq n_i \leq m_i$, and next,
each the subspace $M_i$ is invariant under $A$ and the canonical (or Jordan) matrix of $A$ is composed by Jordan blocks $J_i$, each block is a matrix with $\lambda_i$ in the main diagonal, with 1 in $m_i - n_i$ cells below the main diagonal, and with 0 in any other places.

**Proposition 2.1.** If $T$ is a densely defined linear operator $T$ from a Banach space $X$ into itself then the resolvent set $\rho(T)$ is open in the complex plane and the function $\lambda \mapsto (\lambda I - T)^{-1}$ is continuous from $\rho(T)$ into $L(X)$. Moreover, if $T$ is also continuous then the spectrum $\sigma(T)$ is a compact set.

**Proof.** First, if $S$ is linear continuous operator with norm $\|S\| < 1$ then the series $\sum_{n=0}^{\infty} S^n$ converges in the operator norm, this limit belongs to $L(X)$ and is indeed the inverse of $(I - S)$, and denoted by $(I - S)^{-1}$. Indeed, the formal equality $(I - S) \sum_{n=0}^{\infty} S^n = I$ and the inequality

$$\| \sum_{n \geq k} S^n \| \leq \sum_{n \geq k} \| S^n \| \leq \sum_{n \geq k} \| S \|^n \to 0 \quad \text{as} \quad k \to \infty$$

show the previous assertions.

To check that $\rho(T)$ is open, take $\lambda_0$ in $\rho(T)$ and use the equality $\lambda I - T = \lambda_0 I - T - (\lambda_0 - \lambda) I$ to get

$$(\lambda I - T)(\lambda_0 I - T)^{-1} = I - S, \quad \text{with} \quad S = (\lambda_0 - \lambda)(\lambda_0 I - T)^{-1}.$$

Hence, if $|\lambda - \lambda_0|$ is sufficiently small so that $\|(\lambda_0 - \lambda)(\lambda_0 I - T)^{-1}\| < 1$ then the operator $I - S$ is invertible, and the equality $(\lambda I - T) = (I - S)(\lambda_0 I - T)$ shows that $(\lambda I - T)^{-1} = (\lambda_0 I - T)^{-1}(I - S)^{-1}$ is the inverse of $\lambda I - T$.

The previous calculation also shows that if $|\lambda - \lambda_0| < 1/\|(\lambda_0 I - T)^{-1}\|$ then

$$(\lambda I - T)^{-1} = (\lambda_0 I - T)^{-1} \sum_{n=0}^{\infty} [(\lambda_0 - \lambda)(\lambda_0 I - T)^{-1}]^n,$$

which implies the continuity of the function $\lambda \mapsto (\lambda I - T)^{-1}$ at any $\lambda_0$ in $\rho(T)$.

Finally, if $|\lambda| > \|T\|$ then $S = \|T\|/\lambda < 1$ and the equality

$$\lambda^{-1}(\lambda I - T) = I - S$$

shows that $(\lambda I - T)^{-1}$ exists, i.e., $\sigma(T)$ is compact since it is closed and it is contained in a closed ball with center at the origin and radius $\|T\|$.

As mentioned in Section 1.3.2 integrals of functions with values in a Banach space makes perfectly sense, for either measurable or continuous functions. Moreover, analytic functions with values in a Banach space and contour integrals can also be defined. For instance, just working with the Riemann-Stieltjes contour integrals, if $\lambda \mapsto A(\lambda)$ is a continuous functions on an connected open $\Omega \subset \mathbb{C}$ with values in $L(X)$ and $C$ is a rectifiable piecewise $C^1$ curve in $\Omega$ then the contour integral of $A(\cdot)$ along the curve $C$ defines an element in $L(X)$. 

[Preli
dinary]  Menaldi  November 11, 2016
Proposition 2.2. If $T$ is a densely defined linear operator $T$ from a Banach space $X$ into itself, $\Omega$ is a simply connected open inside $\rho(T)$ and $\lambda \mapsto f(\lambda)$ is an analytic function in $\Omega$ then

$$\int_C f(\lambda)(\lambda I - T)^{-1}d\lambda = 0,$$

for any simple closed rectifiable piecewise $C^1$ curve contained in $\Omega$.

**Proof.** First, since the domain $\Omega$ is simply connected and $C$ is a simple closed rectifiable piecewise $C^1$ curve, for any given $\varepsilon > 0$ the $C$ can be decomposed into an even number of smaller simple closed rectifiable piecewise $C^1$ curves $C_i$, $i = 1, \ldots, 2n$, each with diameter smaller than $\varepsilon$, in such a way that the common parts between $C_i$ and $C$ have the same orientation, the other parts of $C_{2k-1}$ and $C_{2k}$ are the same curves with opposite orientation for $k = 1, \ldots, n$, and the equality

$$\int_C f(\lambda)(\lambda I - T)^{-1}d\lambda = \sum_{i=1}^n \int_{C_i} f(\lambda)(\lambda I - T)^{-1}d\lambda$$

holds true.

Hence, each small curve $C_i$ is contained in a ball with center $\lambda_i$ and radius $\varepsilon$, which can be chosen so small that $|\lambda - \lambda_i| \leq \varepsilon$ implies

$$(\lambda I - T)^{-1} = \sum_{k=0}^{\infty} (\lambda_i - \lambda)^k (\lambda_i I - T)^{-(k+1)},$$

for any $\lambda$ in $C_i$. Therefore

$$\int_{C_i} f(\lambda)(\lambda I - T)^{-1}d\lambda = (\lambda_i I - T)^{-(k+1)} \sum_{k=0}^{\infty} \int_{C_i} f(\lambda)(\lambda_i - \lambda)^k d\lambda,$$

and because $f$ is analytic in $\Omega$, each term in the series vanishes and the whole contour integral vanishes are desired. \(\square\)

At this point, let $T$ be a densely define linear operator on a Banach space $X$ with spectrum set $\sigma(T) \subset \mathbb{C}$ and let $f$ be an analytic function defined on an open set $\Omega \supset \sigma(T)$. In view of Proposition 2.2, for any simple closed rectifiable piecewise $C^1$ curve $C$ inside $\Omega$, encircling $\sigma(T)$ (i.e., the open set $D$ of which $C$ is its boundary contains the spectrum of $T$), and oriented in a positive direction (i.e., leaving $D$ on the left), the contour integral

$$\frac{1}{2\pi i} \int_C f(\lambda)(\lambda I - T)^{-1}d\lambda = f(T)$$

(2.1)

defines an element in $L(X)$, independent of the chosen curve $C$. The notation $f(T)$ is justified by the following assertion: if the analytic function $f$ is expressed as $f(\lambda) = \sum_k a_k \lambda^k$ for any $|\lambda| < r_0$, and $\|T\| < r_0$ then

$$f(T) = \frac{1}{2\pi i} \int_C f(\lambda)(\lambda I - T)^{-1}d\lambda = \sum_k a_k T^k.$$
Indeed, by means of Proposition 2.1 the curve $C$ can be chosen inside the open set \{ $\lambda : \|T\| < |\lambda| < r_0$ \} and the geometric series $(\lambda I - T)^{-1} = (1/\lambda) \sum_h (T/\lambda)^h$ is convergent for any $\lambda$ in $C$. Hence

$$\frac{1}{2\pi i} \int_C f(\lambda)(\lambda I - T)^{-1} \, d\lambda = \sum_h \sum_k a_k T^k \frac{1}{2\pi i} \int_C \lambda^{k-h-1} \, d\lambda = \sum_k a_k T^k,$$

since the contour integral inside the double series vanishes for $h \neq k$ and is equal to 1 when $h = k$.

As a Corollary, if $f(\lambda) = \lambda$ then $f(T) = T$ for any $T$ in $L(X)$, and Proposition 2.2 shows that the spectrum of any linear continuous operator $T$ is nonempty.

Certainly, a densely defined linear operator $T$ is not sufficient to develop the spectral theory, commonly, besides $T$ being defined on a domain $D(T)$ dense in $X$, the operator $T$ is assumed to be a closed operator (i.e., with graph \{(x, Tx) : x \in D(T)\} closed as a subspace of $X \times X$). However, in this section, $T$ belongs to $L(X)$, and a typical example is when $X$ is finite dimensional, i.e., $T$ is a matrix. In this case a matrix $T = A$, the spectrum contains only eigenvalues, and by means of the previous results it can be deduced that $\sigma(f(A)) = f(\sigma(A))$, i.e., $v$ is an eigenvector corresponding to the eigenvalue $\lambda$ of $A$ iff $v$ is an eigenvector corresponding to the eigenvalue $f(\lambda)$ of $f(A)$.

For instance, of particular interest is the expression $\exp(tA) = \sum_k (tA)^k / k!$, which is the fundamental solution to the Cauchy problem $y' = Ay$, $y(0) = I$. However, relations like $\exp(A + B) = \exp(A) \exp(B)$ holds when $AB = BA$, but not in general. There are many more properties of the spectrum of a linear operator that should be studied, e.g., the operational calculus and the case when $X$ is a Hilbert space and $T$ is symmetric or compact. For instance, the reader is referred to the textbook Conway [29] or Taylor and Lay [123], among other, while a comprehensive study can be found in Dunford and Schwartz [39] or Reed and Simon [103].

### 2.1.2 Three Basic Results

What follows is an early presentation of the so-called three essential principles (uniformly-bounded, open-mapping and closed-graph) in the context of normed spaces. This simplified version of most of the arguments used in the remaining sections can certainly be skipped.

In a complete metric space $(X,d)$, the diameter of a set $F$ is $d(F) = \sup\{d(x,y) : x, y \in X\}$ and a set $E$ is called nowhere dense if the interior of its closure is empty, i.e., $\text{Int} \, E = \emptyset$.

**Theorem 2.3.** Let $(X,d)$ be a complete metric space. (1) If $\{F_n\}$ is a sequence of non-empty closed sets such that $F_n \supset F_{n+1}$ for every $n$ and $d(F_n) \to 0$ then there is a point $x$ in $X$ satisfying $\bigcap_n F_n = \{x\}$. (2) If $\{E_n\}$ is a sequence of nowhere dense sets then $\bigcup_n E_n \neq X$. 
Proof. (1) If $F_n = F_{n_0}$ for every $n \geq n_0$ then $\bigcup_n F_n = F_0$ and because $d(F_n) \to 0$, the set $F_0$ must be only one point $\{x\}$. On the contrary, if there is a subsequence $\{F_{n_k}\}$ such that $F_{n_k} \neq F_{n_{k+1}}$ then choose a sequence $\{x_k\}$ of points such that $x_k$ belongs to $F_{n_k} \setminus F_{n_{k+1}}$ to obtain $d(x_k, x_j) \leq d(F_{n_k}) \to 0$, for every $j \geq k$, i.e., $\{x_k\}$ is a Cauchy sequence. Therefore $x_k \to x$ and because $F_{n_k}$ is closed and contains all $x_j$ with $k \geq j$, the limit point $\bar{x}$ must belongs to $F_{n_k}$, for every $k \geq 1$, i.e., $\bar{x}$ belongs to $\bigcup_k F_{n_k}$. Since the sequence is monotone decreasing and the intersection $\bigcap_k F_{n_k}$ is at most one point, the equality $\bigcup_n F_n = \{\bar{x}\}$ is proved.

(2) Consider a sequence $\{E_n\}$ of nowhere dense sets. Given an open ball $B(x_0, 1) = \{x \in X : d(x, x_0) < 1\}$, because the interior of $\overline{E_1}$ is empty, the ball $B(x_0, 1)$ cannot be contained in $\overline{E_1}$, and so, there exists a point $x_1$ in $B(x_0, 1)$ such that $x_1$ does not belong to $\overline{E_1}$. This means that there is a ball $B(x_1, r_1)$ of radius $0 < r_1 < 1/2$ such that $B(x_1, r_1) \subset B(x_0, 1)$ and $B(x_1, r_1) \cap \overline{E_1} = \emptyset$. Again, repeating the argument, there exists a point $x_2$ and a number $0 < r_2 < 1/3$ such that $B(x_2, r_2) \subset B(x_1, r_2)$ and $B(x_2, r_2) \cap \overline{E_2} = \emptyset$. By induction, there are two sequences $\{x_k\}$ and $\{r_k\}$ such that $r_k \to 0$, $B(x_k, r_k) \subset B(x_{k-1}, r_k)$ and $B(x_k, r_k) \cap \overline{E_k} = \emptyset$. Therefore, invoking part (1), there is a point $\bar{x}$ such that $\bigcap_k B(x_k, r_k) = \{\bar{x}\}$, i.e., $\bar{x}$ does not belongs to $\bigcup_n E_n$, which means that $\bigcup_n E_n \neq X$. \hfill \Box

There are three main results (which are developed later in a more general setting) that can be proved in a relative simple way for normed spaces. First, the (Banach-Steinhaus) principle of uniform boundedness

**Theorem 2.4.** Let $\{T_i : i \in I\}$ be a family of continuous linear operators from a Banach space $X$ into a normed space $Y$. If $\sup_{i \in I} \|T_i x\|_Y < \infty$, $\forall x \in X$ then there exits a constant $C > 0$ such that $\|T_i x\|_Y \leq C \|x\|_X$, $\forall x \in X$, $\forall i \in I$.

**Proof.** We need to show that for some ball $B(x_0, r_0) = \{x \in X : \|x - x_0\|_X < r\}$ and some constant $K$ such that $\|T_i x\|_Y \leq K$ for every $x$ in $B(x_0, r)$ and for any $i$ in $I$. Indeed, by contradiction suppose that no such a ball exists. Begin with a ball $B(x_0, r_0)$, the family $\{T_i : i \in I\}$ is not uniformly bounded, so that there exists a point $x_1$ in $B(x_0, r_0)$ and an index $i_1$ in $I$ such that $\|T_{i_1} x_1\|_Y > 1$. Now, by continuity, there exists $0 < r_1 < 1$ such that $\|T_{i_1} x\|_Y > 1$, for every $x$ in the ball $B(x_1, r_1) \subset B(x_0, r_0)$. Again, in this ball $B(x_1, r_1)$, the family $\{T_i : i \in I\}$ is not uniformly bounded, so that there exists a point $x_2$ in $B(x_1, r_1)$ and an index $i_2$ in $I$ such that $\|T_{i_2} x_2\|_Y > 2$. Now, by continuity, there exists $0 < r_2 < 1/2$ such that $\|T_{i_2} x\|_Y > 2$, for every $x$ in the ball $B(x_2, r_2) \subset B(x_1, r_1)$. Thus, by induction, there exists sequences $\{x_k\} \subset X$, $\{r_k\} \subset (0,1)$ and $\{i_k\} \subset I$ such that $B(x_{k+1}, r_{k+1}) \subset B(x_k, r_k)$, $r_k \to 0$, $\|T_{i_k} x_k\|_Y > k$, for every $x$ in $B(x_k, r_k)$. Hence, by means of Theorem 2.3, part (1), there exists a point $\bar{x}$ in $\bigcap_n B(x_k, r_k)$. Therefore, $\|T_{i_k} \bar{x}\|_Y > k$, which contradict the assertion $\sup_{i \in I} \|T_i x\|_Y < \infty$, $\forall x \in X$. \hfill \Box

Second, the (Banach-Schauder) open mapping
Theorem 2.5. If $T$ is a continuous linear operator from a Banach spaces $X$ onto another Banach space $Y$ then $T$ maps open sets onto open sets, i.e., for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $y$ in $Y$ satisfying $\|y\|_Y < \delta$ there exists $x$ in $X$ satisfying $y = Tx$ and $\|x\|_X < \varepsilon$. In particular, if $T$ is also one-to-one then $T^{-1}$ is a continuous linear operator.

Proof. First, if $B_X(0, r)$ and $B_Y(0, r)$ denote the ball centered at the origin 0 with radius $r$ on $X$ and $Y$ then we show that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $B_Y(0, \varepsilon) \subset TB_X(0, 2\varepsilon)$. Indeed, the equality $X = \bigcup_n nB_X(0, \varepsilon)$, for given any $\varepsilon > 0$, and assumption that $T$ is surjective, yield that $Y = \bigcup_n nTB_X(0, \varepsilon)$. Hence, because $Y$ is a complete metric space, Theorem 2.3, part (2), implies that for some $n$ the closure $nTB_X(0, \varepsilon)$ must contains some ball $B_Y(y_0, r)$. Therefore with $\delta = r/n$ we deduce

$$B_Y(0, \delta) \subset \{ y = y_1 - y_2 : y_1, y_2 \in B_Y(y_0, \delta) \} \subset \overline{T\{ x = x_1 - x_2 : x_1, x_2 \in B_X(0, \varepsilon) \}} \subset TB_X(0, 2\varepsilon)$$

as desired.

Next, we show that that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $B_Y(0, \delta) \subset TB_X(0, 2\varepsilon)$. Indeed, based on the first part, for any sequence of positive numbers $\{\varepsilon_k\}$ there exists another sequence of positive numbers $\{\delta_k\}$ such that $B_Y(0, \delta_k) \subset \overline{TB_X(0, \varepsilon_k)}$ and $\delta_k \to 0$. Given $\varepsilon > 0$, choose $\sum_k \varepsilon_k < \varepsilon$ and $\delta = \delta_1$. Now, if $y \in B_Y(0, \delta)$ then there exists $x_1 \in B_X(0, \varepsilon_1)$ such that $\|y - Tx_1\|_Y < \delta_2$, i.e., $y - Tx_1 \in B_Y(0, \delta_2)$. Repeating the argument, there exists $x_2 \in B_X(0, \varepsilon_2)$ such that $\|y - Tx_1 - Tx_2\|_Y < \delta_3$, and by induction, there is a sequence $\{x_k\} \subset X$ such that

$$\|y - Tx_1 - Tx_2 - \cdots - Tx_k\|_Y < \delta_{k+1} \to 0.$$

Because $\|x_k\| < \varepsilon_k$ the series of partial sums are a Cauchy sequence in $X$, which is a complete (normed) space, the series $\sum_k Tx_k$ converges to some $x$ in the closure $\overline{B_X(0, \varepsilon)}$. Thus, $y = Tx$ with $x \in B_X(0, 2\varepsilon)$.

Finally, the last statement says that $T$ preserves neighborhood of the origin, i.e., $T$ maps a neighborhood of the origin in $X$ into another neighborhood of the origin in $Y$. Because $T$ is linear and the scalar multiplication and addition of vectors are continuous operations, the conclusion follows. □

Third, the closed graph theorem can be regarded as a convenient way of checking continuity of a linear operator. The analysis of linear closed operators (i.e. linear operator with a closed graph) has several applications, mainly referring to operators defined only in a dense subspace. We are only concerned with linear defined everywhere, i.e., in the whole space.

Theorem 2.6. Let $T$ be a linear operator between two Banach spaces $X$ and $Y$. If $T$ satisfies, for any sequence $\{x_n\}$ in $X$,

$$x_n \to 0 \text{ and } Tx_n \to y \text{ imply } y = 0$$

then $T$ is continuous.
Proof. Note that the above condition is equivalent to assume that for any sequence \( \{x_n\} \) in the domain of \( T \) such that \( x_n \to x \) and \( Tx_n \to y \) we have \( Tx = y \), which is the definition of a closed linear operator with domain \( D_T \subset X \) into \( Y \).

Consider the graph of \( T \), namely, \( \text{graph}(T) = \{(x,Tx) : x \in X\} \), as a subset of the Cartesian product \( X \times Y \), which is a Banach space with the norm \( \| (x,y) \| = \| x \|_X + \| y \|_Y \). The assumptions on \( T \) imply that \( \text{graph}(T) \) is a closed linear subspace, and so, a Banach space in itself. The linear mapping \( S : (x,Tx) \mapsto x \) from \( \text{graph}(T) \) into \( X \) is continuous and one-to-one. Hence, invoking the open-mapping Theorem 2.5, its inverse \( S^{-1} : x \mapsto (x,Tx) \) is a continuous linear operator. Thus, by looking at the second component, we deduce that \( T \) is a continuous linear operator.

Another key result that could be included in this list is Hahn-Banach Theorem, and its various generalizations and variations, e.g., see the accessible book by Friedman [46, Chapter 4, pp. 123–185]. However, this will be discussed later, in a more general context. Let us also mention, that for instance, the reader interested in applications may take a look at the textbooks Aubin [12], Griffel [59], Hutson et al. [69], Oden and Demkowicz [95], Zeidler [137, 138, 139], among many others.

2.1.3 Examples and Comments

By now, the reader must be familiar with the concepts of metric, norm, scalar (or inner) product, and topological vector spaces. Besides the \( L^p \) spaces corresponding to a measure space \((\Omega, \mathcal{F}, \mu)\) we have many other useful Banach spaces:

- \( C_b(X) \): for \( X \) a Hausdorff topological (usually a complete separable metrizable, i.e., Polish) space, this is the Banach space of real-valued (or complex-valued) continuous and bounded functions on \( X \), with the sup-norm, \( \| f \| = \sup\{|f(x)| : x \in X\} \). If \( X \) is compact then the suffix \( b \) is not necessary (because a continuous function on a compact set is bounded) and we write \( C_b(X) = C(X) \), the space of continuous functions. Usually, the base space \( X \) is an open or closed subset of \( \mathbb{R}^d \).

- \( C_0(X) \): for \( X \) a locally compact (but not compact) Hausdorff topological (usually a complete separable metrizable, i.e., Polish) space, this is the separable Banach space of real-valued (or complex-valued) continuous functions vanishing at infinity on \( X \), i.e., a continuous function \( f \) belongs to \( C_0(X) \) if for every \( \varepsilon > 0 \) there exists a compact subset \( K = K_\varepsilon \) of \( X \) such that \( |f(x)| \leq \varepsilon \) for every \( x \in X \setminus K \). This is a proper subspace of \( C_b(X) \) with the sup-norm. Usually, \( X \) is an open subset of \( \mathbb{R}^d \). Note that depending on the context, sometimes this space is denoted by either \( C_*(X) \) or \( C^\infty_0(X) \).

- \( C_0(X) \): for \( X \) a compact subset of a locally compact Hausdorff topological (usually a Polish) space, this is the separable Banach space of real-valued (or
complex-valued) continuous functions vanishing on the boundary of $X$, with the sup-norm. Usually, $X$ is a closed and bounded subset of $\mathbb{R}^d$.

- $C^k_b(E)$: for $E$ a domain in the Euclidean space $\mathbb{R}^d$ (i.e., the closure of the interior of $E$ is equal to the closure of $E$, and usually connected) and $k$ a nonnegative integer, this is the subspace of $C^k_b(E)$ of functions $f$ such that all derivatives up to the order $k$ belong to $C^k_b(E)$ with the natural norm or seminorms. For instance, if $E$ is compact then any continuous function on $E$ is also bounded and $C^k_b(E) = C^k_b(E)$ is a separable Banach space with the sup-norm for the function and all derivatives up to the order $k$ included. Clearly, this is extended to the case $k = \infty$, but $C^\infty_b(E)$ is not a normed space.

For instance, the interested reader may consult the book Kufner et al. [76].

**Exercise 2.1.** Given a domain $E$ in the Euclidean space $\mathbb{R}^d$ and $0 < \alpha < 1$ we say that a function $f : E \to \mathbb{R}$ is Hölder continuous in $E$ with exponent $\alpha$ if there exists a constant $C$ such that $|f(x) - f(y)| \leq C|x - y|^{\alpha}$, for every $x, y$ in $E$ (and the limiting case $\alpha = 1$ is called Lipschitz continuous), and the smallest of all those constant $C$ is denoted by $[f]_{\alpha,E},$ i.e.,

$$[f]_{\alpha,E} = \sup_{x,y \in E, x \neq y} \{|f(x) - f(y)| |x - y|^{-\alpha}\}.$$ 

For the limiting case $\alpha = 0$, we use $C^0(E) = C(E)$. Now, denote by $C^{0,\alpha}(E)$ the space of all Hölder (Lipschitz) continuous functions on $E$. Sometime, the notation $C^{0,\alpha}(E) = C^\alpha(E)$, with $0 < \alpha < 1$, could be used. Assume $E$ a bounded set and prove that $C^{0,\alpha}(E)$ are Banach spaces with the norm

$$\|f\|_{\alpha,E} = [f]_{\alpha,E} + \sup_{x \in E} |f(x)|, \quad 0 < \alpha \leq 1$$

Consider also the case when $E$ is unbounded and discuss the spaces $C^{n,\alpha}_b(E)$ defined as a combination of $C^n_b(E)$ and $C^{0,\alpha}(E)$.

Sometimes, the interest is on functions of one variable such as the typical space $C([0,T];\mathbb{R}^d)$ of continuous $\mathbb{R}^d$-valued functions defined on the compact (time-) interval $[0,T]$. It is relative simple to verify that this is a separable Banach space with the sup-norm. If an unbounded interval $[0,\infty]$ replace $[0,T]$ then either continuous and bounded functions are used to preserve the sup-norm and to have a Banach space, or more general spaces are necessary, like Polish spaces. Variation of these spaces are frequently discussed, for instance:

(1) the space $C^\alpha([0,T];\mathbb{R}^d)$ of $\alpha$-Hölder continuous functions with $0 < \alpha < 1$, where $\alpha = 1$ is referred to as Lipschitz-continuous functions and denoted either $C^{0,1}([0,T];\mathbb{R}^d)$ or Lip$([0,T];\mathbb{R}^d)$. These are Banach spaces, but they are non-separable. The inclusion from $C^\alpha([0,T];\mathbb{R}^d)$ into $C([0,T];\mathbb{R}^d)$ is continuous, and $C^\alpha([0,T];\mathbb{R}^d)$ is a subspace dense in $C([0,T];\mathbb{R}^d)$, this is referred to continuous and dense.

(2) the space $D([0,T];\mathbb{R}^d)$ of cad-lag functions, i.e., functions $t \mapsto f(t)$ which are continuous from the right at each time $t$ in $[0,T)$ and have limit from the left at each time $t$ in $(0,T]$. This space is of great importance in study of processes in probability theory, usually refer to as canonical space, and it is Polish space, i.e., a complete separable metrizable space, not necessarily a topological vector space. The inclusion from $C([0,T];\mathbb{R}^d)$ into $D([0,T];\mathbb{R}^d)$ is continuous and $C([0,T];\mathbb{R}^d)$ is a closed subspace of $D([0,T];\mathbb{R}^d)$.

(3) the space $BV([0,T];\mathbb{R}^d)$ of functions having bounded variation with the variation-norm. These functions are not necessarily continuous and they form a Banach space, which is non-separable. Because functions with bounded variation are expressible as a difference of two monotone nondecreasing functions from the right and having limits from the left, at any time $t$ in $[0,T)$, or equivalently cag-lad, or even, consider continuous functions with bounded variation, i.e., the spaces $BV([0,T];\mathbb{R}^d) \cap D([0,T];\mathbb{R}^d)$ or $BV([0,T];\mathbb{R}^d) \cap C([0,T];\mathbb{R}^d)$. Again, these are non-separable Banach spaces.

(4) the space $AC([0,T];\mathbb{R}^d)$ of absolutely continuous functions. The norm given to this space involves the property that any absolutely continuous function $f$ can be written as the integral of its almost everywhere pointwise derivative, i.e.,

$$ f(t) = f(0) + \int_0^t \dot{f}(s)ds, \quad \forall t \in [0,T], $$

where $\dot{f}$ belongs to the Lebesgue space $L^1([0,T],\mathbb{R}^d)$. Thus, the norm $f$ in $AC([0,T];\mathbb{R}^d)$ is defined as $|f(0)| + \|\dot{f}\|_{L^1}$. This space is a separable Banach space. The inclusion from $AC([0,T];\mathbb{R}^d)$ into $C([0,T];\mathbb{R}^d)$ is continuous and dense. Also, inclusion from $AC([0,T];\mathbb{R}^d)$ into $BV([0,T];\mathbb{R}^d)$ is continuous (under the variation norm), but $AC([0,T];\mathbb{R}^d)$ is closed as a subspace of $BV([0,T];\mathbb{R}^d)$.

(5) the Sobolev space $W^{1,p}([0,T];\mathbb{R}^d)$ could be defined as the subspace of $AC([0,T];\mathbb{R}^d)$ of all functions $f$ such that $\dot{f}$ belongs to $L^p([0,T];\mathbb{R}^d)$, for any $1 \leq p \leq \infty$. Thus $AC = W^{1,1}$ and $\text{Lip} = W^{1,\infty}$. These are Banach spaces, separable if $1 \leq p < \infty$, non-separable for $p = \infty$, and a Hilbert space for $p = 2$. Clearly, the inclusion of $W^{1,p}([0,T];\mathbb{R}^d)$ into $W^{1,q}([0,T];\mathbb{R}^d)$ with $q > p$ is continuous and dense. Usually, the Sobolev spaces $W^{1,p}$ are defined and considered subspaces of the Lebesgue spaces $L^p$, i.e., functions are classes of equivalence, but all this is not necessary for the one-dimensional case of functions defined on $[0,T]$. Absolutely continuous functions are not necessarily Hölder continuous, but the Sobolev space $W^{1,p}([0,T];\mathbb{R}^d)$ is continuously embedded in the Hölder space $C^{1-1/p}([0,T];\mathbb{R}^d)$, for $1 < p < \infty$.

The reader could take a look at the book by Friz and Victoir [47] for more details on the previous statements (1), . . . , (5).

For $X$ compact, the dual space of $C_b(X) = C(X)$ is the space of finite regular Borel measures on $X$, e.g. see Dunford and Schwartz [39, Vol 1, Section
VI.6, Theorem 3, pp. 265]. Hence, if $X$ is locally compact then the Banach space $C_0(X)$ of continuous functions vanishing at infinity on $X$ (i.e., continuous functions such that for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon$ in $X$ satisfying $|f(x)| \leq \varepsilon$ for every $x$ in $X \setminus K_\varepsilon$) can be regarded as a subspace of $C(X \cup \{\infty\})$, via the one-point compactification. This means that the dual of the space of $C_r(X)$ is the space of finite regular Borel (or Radon) measures on $X$.

However, our interest is on $C_0^0(\Omega)$, the space of real (or complex) continuous functions with compact support on an open domain $\Omega$ of $\mathbb{R}^d$. As it is clear from the above spaces, $C_0^0(\Omega)$ is not a normed space, but for any compact set $K$, the subspace $C_0^0(K)$ of all functions in $C_0^0(\Omega)$ with support in $K$ is a Banach space isomorphic to $C_0^0(K)$, the space of continuous functions on $K$ vanishing on the boundary $\partial K$ of $K$. Certainly, we can write $\Omega = \bigcup_n K_n$, with $K_n$ a compact subset of the interior of $K_{n+1}$, and then we could use the sup-norm. Thus, $C_0^0(\Omega)$ would be a separable normed space, $C_0^0(K)$ would be closed subspace, but $C_0^0(\Omega)$ is not complete with this metric, actually, its completion is the above space $C_0(\Omega)$. As we have seen early, $C_0^0(\Omega) \subset C_0(\Omega)$ is a dense subspace of $L^p(\Omega)$, for any $1 \leq p < \infty$, thus we desire to endow $C_0^0(\Omega)$ with a topology that make it a “good” complete space, this topology is called inductive limits and $C_0^0(\Omega)$ becomes a complete locally convex topological space. Certainly, more complicate is the space $C_\infty^0(\Omega)$.

A couple of points we need to remember. For a metric space $(X, d)$ the following properties are equivalent: (1) $S$ is compact, i.e., every open cover has a finite subcover; (2) $S$ is both complete and totally bounded, i.e., every Cauchy sequence is convergent and for every $\varepsilon > 0$ there exists a finite subset $\{x_1, \ldots, x_n\} \subset X$ such that $\min_i d(x, x_i) < \varepsilon$ for every $x$ in $X$; (3) $S$ is sequentially compact, i.e., every sequence has a convergent subsequence. Because a subset of a metric space is itself a metric space, the previous properties apply to subsets of metric spaces with the relative topology. For a detailed proof of these facts see, e.g., Dudley [36, Theorem 2.3.1, pp. 45–47]. Therefore, a compact metric space is separable, and also, a compact space is metrizable.

## 2.2 Compactness and Separability

Functionals refer to functions with real (or complex) values, usually regarding continuous nonlinear functionals on a metric space. Perhaps, it may be convenient to take a look at some pieces of the basic arguments regarding function spaces, e.g., Carothers [24, Chapters 10–13, pp. 139–213].

### 2.2.1 Linear Functionals

However, in many circumstances the interest is on continuous linear functionals, which are necessarily defined on a topological vector space (tvs), i.e., a vector space with a compatible (Hausdorff) topology, meaning such that the addition and the scalar multiplication are continuous operations. Thus if $X$ is a tvs then a (sub)base (of open sets) for the origin (0) determines (or defines) the topology.
Our interest is in the space \( X' \) of continuous linear functionals defined on \( X \). Note that a linear functional is continuous if and only if it is continuous at the origin. This vector space \( X' \) (the dual space of \( X \)) can be endowed with several topologies. Perhaps the most immediate is the weak* topology, namely, a subbase of open sets for the origin is the family of sets \( \{ f \in X' : |\langle f, x \rangle| < r \} \), indexed by \( x \) in \( X \) and \( r > 0 \), i.e., weak* convergence means pointwise convergence. Certainly, this dual space \( X' \) is a subspace of the product space \( \mathbb{R}^X \) (or \( \mathbb{C}^X \)) (i.e., all functional defined on \( X \)) with requires not topology on \( X \). This product space can be endowed with the product topology, which is not necessarily a vector space. However, if \( g \) is in \( \mathbb{R}^X \) then subbase of open sets for the point \( g \) is the family of sets \( \{ f \in \mathbb{R}^X : |f(x) - g(x)| < r \} \), indexed by \( x \) in \( X \) and \( r > 0 \), i.e., again this means the pointwise convergence. Moreover, Tychonoff’s Theorem affirms that a subset of \( \mathbb{R}^X \) (or \( \mathbb{C}^X \)) is compact in the product topology if and only if all projections are compacts in \( \mathbb{R} \) (or \( \mathbb{C} \)), i.e., a set \( K = \{ f : f \in \mathbb{R}^X \} \) is compact if and only if the set \( \{ f(x) : f \in K \} \) is compact in \( \mathbb{R} \) (or \( \mathbb{C} \)) for every \( x \) in \( X \).

Now we are ready to state and prove a criterion for compactness of continuous linear functionals under the weak* topology, which is known as Alaoglu or Banach-Alaoglu Theorem.

**Theorem 2.7.** If \( U \) is a neighborhood of the origin in a topological vector space \( X \) then the set \( F = \{ f \in X' : |\langle f, x \rangle| \leq 1, \forall x \in U \} \) is weak* compact

**Proof.** Since the scalar multiplication is a continuous operation and \((1/\lambda)x \to 0 \) as \( \lambda \to \infty \), we deduce that if \( U \) is a neighborhood of 0 in \( X \) then there exists \( \lambda(x) > 0 \) such that \( x/\lambda(x) \) belongs to \( U \). Thus, the inequality \( |\langle f, x \rangle| = \lambda(x)|\langle f, x/\lambda(x) \rangle| \leq \lambda(x) \) implies that

\[
|\langle f, x \rangle| \leq \lambda(x), \quad \forall f \in K, \ x \in X,
\]

i.e., the set \( \{ \langle f, x \rangle : f \in F \} \) is a bounded set in \( \mathbb{R} \) (or \( \mathbb{C} \)), for every \( x \) in \( X \). Thus, via Tychonoff’s Theorem, the proof is completed if the set of scalar \( \{ \langle f, x \rangle : f \in F \} \) is closed, for every \( x \) in \( X \).

To this end, suppose that \( g \) belongs to the closure of \( F \) and show that \( g \) is in \( X' \), i.e., a continuous linear functional and that \( |\langle f, x \rangle| \leq 1 \), for any \( x \) in \( U \). Indeed, take \( \varepsilon > 0 \), any two points \( x, y \) and any two scalars \( \alpha, \beta \) to consider the set \( \{ f : |f(z) - g(z)| < \varepsilon, \ z = x, y, \alpha x + \beta y \} \), which is an open set in \( \mathbb{R}^X \) (or \( \mathbb{C}^X \)) containing \( g \). This neighborhood must contains some element \( f \) in \( F \), and therefore,

\[
|\alpha g(x) + \beta g(y) - g(\alpha x + \beta y)| = |\alpha(f(x) - g(x)) + \beta(f(y) - g(y)) - (f(\alpha x + \beta y) - g(\alpha x + \beta y))| \leq (|\alpha| + |\beta| + 1)\varepsilon,
\]

which shows that \( g \) is a linear functional.

Essentially the same argument can be used to prove that \( g \) is continuous, i.e, for every \( x \) in \( X \), \( \varepsilon > 0 \), the open set \( \{ f : |f(x) - g(x)| < \varepsilon/2 \} \) must intercept \( F \) in some point \( f \). Since \( f \) is continuous, there exists a neighborhood \( V \) of \( x \)
such that \(|\langle f, y \rangle - \langle f, x \rangle| \leq \varepsilon/2\), for every \(y\) in \(V\), which yields \(|g(y) - g(x)| \leq \varepsilon\), for every \(y\) in \(V\).

Similarly, choose \(x\) in \(U\) and \(\varepsilon > 0\) to claim again that open set \(\{f : |f(x) - g(x)| < \varepsilon\}\) must intercept \(F\) in some point \(f\), which implies that

\[
|g(x)| \leq |g(x) - f(x)| + |\langle f, x \rangle| < \varepsilon + 1,
\]

i.e., \(g\) belongs to \(F\).

\[\square\]

**Remark 2.8.** If the topological vector space \(X\) is separable then the weak\(^*\) topology is sequential, e.g., the unit ball is weakly\(^*\) sequentially compact, and a subset \(C\) of \(X\) is weakly\(^*\) closed if and only if any converging sequence \(\{x_n\}\) of element on \(C\) converges to a limit \(x\) belonging to \(C\). The reader may check the book by Brezis [22].

If \(X\) is a normed space then a typical neighborhood is the closed unit ball \(V = \{x \in X : \|x\|_X \leq 1\}\), and therefore, the weak\(^*\) compact set \(F\) becomes the dual closed unit ball in \(X'\), i.e.,

\[
F = \{f \in X' : |\langle f, x \rangle| \leq 1, \forall x \in X, \|x\|_X < 1\},
\]
or equivalently \(F = \{f \in X' : \|f\|_{X'} \leq 1\}\) where \(\|f\|_{X'} = \sup\{\|\langle f, x \rangle\| : x \in X, \|x\|_X \leq 1\}\) is actually a norm on \(X'\). Moreover, if \(X\) is also separable then a more constructive proof without the use of Tychonoff’s Theorem can be given, namely, the set \(F\) is sequentially compact based on the Cantor’s diagonal argument. Indeed, if \(\{x_i\}\) is a dense sequence in \(X\) and \(\{f_n\}\) is a sequence in \(F\) then the numerical double sequence \(\{f_n(x_i)\}\) is bounded and therefore, there exists a subsequence \(\{f_{nk}\}\) such that \(f_{nk}(x_i)\) converges to some scalar denoted by \(f_0(x_i)\), for every \(i \geq 1\). Thus, for every \(x\) in \(X\) with \(\|x\|_X < 1\) there exists a subsequence \(\{x_{ik}\}\) such that \(\|x_{ik}\| < 1\) and \(x_{ik} \to x\). Since \(\{f_n\} \subset F\), the inequality

\[
|f_{nk}(x) - f_{nk}(x_{ik})| \leq \|f_{nk}\|_{X'}\|x - x_{ik}\|_X \leq \|x - x_{ik}\|_X
\]

show that the subsequence \(f_{nk}(x)\) converges, and the linearity implies that \(f_{nk}(x) = \lambda f_{nk}(x/\lambda)\) also converges to \(f_0(x)\), for every \(x\) in \(X\). Hence \(f_0\) is a functional and a variation of the above argument show that \(f_0\) is linear and continuous, moreover, \(f_0\) belongs to \(F\) and \(f_{nk}\) converges weakly\(^*\) to \(f_0\). For instance, the interested reader may consult the textbooks Conway [29, Section V.1-3, pp. 124–137], Rudin [109, Sections 3.15-19, pp. 66–70], Swartz [119, Section III.15, pp. 199–208].

### 2.2.2 Nonlinear Functional

Now our interest turn into the nonlinear functional. A family \(\{f_i : i \in I\}\) of real-valued (or complex-valued) continuous functions on a metric space \((X, d)\) is called:

(a) equi-continuous if for every $x_0$ in $X$ and any $\varepsilon > 0$ there exists $\delta > 0$ such that $d(x, x_0) < \delta$ implies $|f_i(x) - f_i(x_0)| < \varepsilon,$ for every $i$ in $I$, or equivalently, for any sequences $\{i_n\} \subset I$ and $\{x_n\} \subset X$ with $x_n \to x_0$ we have $|f_{i_n}(x_n) - f_{i_n}(x_0)| \to 0$;

(b) uniformly bounded if there exists a constant $C > 0$ such that $|f_i(x)| \leq C$ for every $x$ in $X$ and $i$ in $I$;

(c) pre-compact if any sequence contains a uniformly convergent sequence, i.e., if $\{f_i : i \in J\}$ with $J$ a sequence of indices in $I$ then there exists a subsequence $J'$ of $J$ and a function $f(x)$ such that for every $\varepsilon > 0$ there is finite subset $J'_{\varepsilon}$ satisfying $|f_i(x) - f(x)| < \varepsilon$ for every $x$ in $X$ and $i$ in $J' \setminus J'_{\varepsilon}$.

It is clear that uniformly bounded means equi-bounded in the space $C_b(X)$. Similarly, a subset of $C_b(X)$ is pre-compact (also called relatively compact) if its closure is compact. Note that uniformly bounded is equivalent to pointwise pre-compact, i.e., the set $\{f_i(x) : i \in I\}$ is pre-compact for every fixed $x$ in $X$.

Recall that a subset of a complete metric space is pre-compact if and only if it is totally bounded, e.g., Yosida [135, Section 0.2, pp. 13–15]. In particular, since the space $C_b(X)$ is complete, the family of functions $\{f_i : i \in I\}$ is pre-compact if it is totally bounded, namely, if for every $\varepsilon > 0$ there exists a finite subset $J \subset I$ such that for every $i$ in $I$ there exists $j$ in $J$ satisfying $|f_i(x) - f_j(x)| < \varepsilon,$ for every $x$ in $X$, i.e., any element in $\{f_i : i \in I\}$ is within a distance $\varepsilon$ from the finite set $\{f_j : j \in J\}$.

It is clear that a finite family of continuous and bounded functions is equi-continuous, uniformly bounded and pre-compact. The following is a typical version of Arzela-Ascoli compactness argument

**Theorem 2.9.** Let $\{f_i : i \in I\}$ be a family of real-valued (or complex-valued) continuous functions on a compact metric space $(X, d)$. Then the family is pre-compact if and only if it is equi-continuous and uniformly bounded, and in this case, the family is uniformly equi-continuous, i.e., for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|f_i(x) - f_i(y)| < \varepsilon,$ for every $i \in I$ and for every $x, y$ in $X$ satisfying $d(x, y) < \delta$.

**Proof.** Suppose the family $\{f_i : i \in I\}$ is pre-compact and proceed in two steps. 

(1) For the possible infinite supremum $C = \sup\{|f_i(x)| : x \in X, i \in I\}$ there exit sequences $\{i_n, x_n\}$ such that $|f_{i_n}(x_n)| \to C$. Since $X$ is compact, there exists a subsequence $\{x_{n'}\}$ convergent to $x_0$. Because the family is pre-compact, there exists a subsequence of $\{i_{n'}\}$, denoted by $\{i_{n''}\}$, such that $\{f_{i_{n''}}(x)\}$ converges uniformly to some $f(x)$, in particular, $|f_{i_{n''}}(x_{n''}) - f_{i_{m''}}(x_{n''})| \to 0$ as $n''$ and $m''$ go to infinite. Hence

$$|f_{i_{n''}}(x_{n''}) - f(x_{0})| \leq \varepsilon + |f_{i_{m''}}(x_{n''}) - f_{i_{m''}}(x_{0})| + |f_{i_{m''}}(x_{0}) - f(x_{0})|,$$

for every $n''$ and $m''$ sufficiently large. By means of the continuity of $f_{i_{m''}}$, the right-hand term converges to zero, which implies that $f_{i_{m''}}(x_{n''}) \to f(x_0)$ and then $C = f(x_0) < \infty$, i.e., the family is uniformly bounded. (2) By contradiction, if the family was not equi-continuous at some $x_0$ then there exist
\( \varepsilon > 0 \) and a sequence \( \{x_n\} \) such that \( x_n \to x_0 \) and \( |f_{i_n}(x_n) - f_{i_n}(x_0)| \geq \varepsilon \). Because the family is pre-compact, there exists a subsequence of \( \{i_n\} \), denoted by \( \{i_{n'}\} \), such that \( \{f_{i_{n'}}(x)\} \) converges uniformly to some \( f(x) \). As above, this would imply that \( f_{i_{n'}}(x_{n''}) \to f(x_0) \), which is a contradiction.

Conversely, now suppose that the family \( \{f_i : i \in I\} \) is equi-continuous and bounded. Because the family is uniformly bounded, for every \( x \) and any sequence of indices \( J \), we can find a sequence of indices \( \{i_n\} \subset J \) such that the numerical sequence \( \{f_{i_n}(x)\} \) converges to some value denoted by \( f(x) \). Thus, for \( D \) is a countable dense set in \( X \) the diagonal Cantor procedure shows that we can get a sequence of indices \( \{i_{n'}\} \) such that \( f_{i_{n'}}(x) \to f(x) \), for every \( x \) in \( D \). Because \( D \) is dense in \( X \), for every \( x_0 \) there exists a sequence \( \{x_k\} \subset D \) such that \( x_k \to x_0 \); and the equi-continuity implies that the double sequence \( f_{i_{n'}}(x_k) \) is also convergent to some value, denoted by \( f(x_0) \). To show that that convergence is uniformly in \( X \), we use the compactness of \( X \). Indeed, by contradiction, if the convergence is not uniformly then there exists \( \varepsilon > 0 \) and a sequence \( \{x_k\} \) such that \( |f_{i_{n'}}(x_k) - f_{i_{n'}}(x_0)| \geq \varepsilon \) for any \( n', m' \) and \( k \) sufficiently large. Because \( X \) is compact, there exists a subsequence of \( \{x_k\} \), denoted \( \{x_{k'}\} \), such that \( x_{k'} \to x_0 \). Hence, by means of the equi-continuity and the inequality

\[
|f_{i_{n'}}(x_{k'}) - f_{i_{n'}}(x_{k'})| \leq |f_{i_{n'}}(x_{k'}) - f_{i_{n'}}(x_0)| + |f_{i_{n'}}(x_0) - f(x_0)| + |f_{i_{n'}}(x_0) - f_{i_{n'}}(x_0)| + |f(x_0) - f_{i_{n'}}(x_0)|,
\]

we obtain a contradiction.

To verify that the family is uniformly equi-continuous, we argue by contradiction as previously. Indeed, if the family \( \{f_i : i \in I\} \) was not uniformly equi-continuous then there would exist \( \varepsilon > 0 \) and sequences \( \{i_n\} \), \( \{x_n\} \) and \( \{y_n\} \) such that \( d(x_n, y_n) \to 0 \) and \( |f_{i_n}(x_n) - f_{i_n}(y_n)| > \varepsilon \). Because the space \( X \) is compact, there exist subsequences \( \{x_{n'}\} \) and \( \{y_{n'}\} \) such that \( x_{n'} \to x_0 \) and \( y_{n'} \to y_0 \), with \( x_0 = y_0 \). Hence, the inequality

\[
|f_{i_{n'}}(x_{n'}) - f_{i_{n'}}(y_{n'})| \leq |f_{i_{n'}}(x_{n'}) - f_{i_{n'}}(x_0)| + |f_{i_{n'}}(y_0) - f_{i_{n'}}(y_{n'})|,
\]

and the equi-continuity at the point \( x_0 = y_0 \) yields a contradiction.

\[ \square \]

\textbf{Remark 2.10.} A point of importance in the above arguments is the assumption that \( X \) is a compact space and \( f_i \) takes values inside another compact space. Thus, Arzelà-Ascoli Theorem 2.9 can be restated as a family of functions from a compact metric space into a complete metric space is pre-compact (or equivalently totally bounded) if and only if it is equi-continuous and pointwise pre-compact (or equivalently pointwise totally bounded). For instance, the reader may consult the book Dudley \cite[Theorem 2.4.7, pp. 52-53]{Dudley} or Dunford and Schwartz \cite[Vol 1, Section VI.6, Theorem 7, pp. 266-267]{DunfordSchwartz}, for a more detailed account.

\[ \square \]

\textbf{Exercise 2.2.} With the notation of Exercise 2.1, let \( \{f_n\} \) be a bounded sequence in the Hölder space \( C^{0, \alpha}(K) \) with \( K \subset \mathbb{R}^d \) and \( 0 < \alpha \leq 1 \). Prove that if \( 0 < \alpha' < \alpha \) and \( K \) is compact then there exists a subsequence \( \{f_{n_k}\} \) and a function \( f \) in \( C^{0, \alpha}(K) \) such that \( f_{n_k} \to f \) in \( C^{0, \alpha'}(K) \).

\[ \square \]
2.2.3 Baire Category Arguments

As mentioned early, a subset $F$ of a topological space $X$ is called a $F_\sigma$-set (of $X$) if $F$ can be expressed as a countable (i.e., denumerable or finite) union of closed sets. Similarly, a subset $G$ of $X$ is called a $G_\delta$-set (of $X$) if $G$ can be expressed as an intersection of open sets. Certainly, the complement of a $F_\sigma$-set is a $G_\delta$-set, and conversely. Any countable intersection of $G_\delta$-sets is itself a $G_\delta$-set, and any union of $F_\sigma$-sets is itself a $F_\sigma$-set. It is also clear that any closed set is a $F_\sigma$-set and any open set is a $G_\delta$-set. However, on only some topological spaces (e.g., metric spaces) any closed is also $G_\delta$-set, this property was used in Part I, with Borel measures. This is related to the concept of dense subsets (or everywhere dense), complete and locally compact spaces, as well as the following definition.

A subset $A$ of a topological space $X$ is nowhere dense if its closure $\overline{A}$ has empty interior, in other words, the complement of the closure $\overline{A}$ is dense set in $X$. Any subset of topological space $X$ that can written as a countable union of nowhere dense sets is called sets of first category, and any subset which is not of first category is said to be of second category. In particular, a topological space $X$ is of second category (or a Baire topological space) if it cannot be written as a countable union of nowhere dense sets.

It is immediate that (1) any subset of a set of first category is of first category, (2) any countable union of sets of first category is of first category, (3) any closed set with empty interior is of first category, (4) a closed set is nowhere dense if and only if its interior is empty, (5) if $A$ is nowhere dense then for any open set $U$ the interior of $U \setminus \overline{A}$ must be nonempty (6) if $A$ is nowhere dense and open then $\overline{A} \setminus A$ is nowhere dense, (7) if $A$ is nowhere dense and closed then $A \setminus \overline{A}$ is nowhere dense, (8) if a topological space contains a subset of second category then the whole space is of second category, (9) categories are preserved by homeomorphism, i.e., if $h: X \to Y$ is a continuous one-to-one open mapping then $h(A)$ has the same category as $A$. Note that a homeomorphism satisfies: the image of the interior of the closure of $A$ is equal to the interior of the closure of the image of $A$. Thus, if $\{D_i\}$ is a sequence of nowhere dense sets such that $h(A) = \bigcup_i D_i$ then $A = \bigcup_i h^{-1}(D_i)$ and $\{h^{-1}(D_i)\}$ is also a sequence of nowhere dense sets.

For instance, the Cantor set (which is an uncountable union of nowhere dense sets) is nowhere dense in $[0,1]$. However, the rational number (which are a countable union of nowhere dense sets) is not nowhere dense. Hence, a countable union of nowhere dense sets might be not nowhere dense and an uncountable union of nowhere dense sets might be nowhere dense.

**Theorem 2.11.** If $X$ is either a complete metric space or a locally compact Hausdorff space then the intersection of every countable family of dense open subsets of $X$ is also dense. Consequently, if $\{A_i\}$ is a sequence of nowhere subsets of $X$ then $D_i = X \setminus \overline{A_i}$ is a sequence of open dense subsets of $X$ and therefore $\bigcap_i D_i \neq \emptyset$, i.e., $X \neq \bigcup_i A_i$, meaning that $X$ is second category. Furthermore, any nonempty open set $U$ in $X$ must be of second category.

**Proof.** Let $\{D_i\}$ be a sequence of dense open subsets of $X$ and $O$ be an arbitrary
nonempty open subset of $X$. Because $O_i$ is dense and open, the set $D_i \cap O$ is a nonempty open set, thus, we can choose a sequence $\{O_i : i \geq 1\}$ of nonempty open sets satisfying $O_i \subset D_i \cap O_{i-1}$, with $O_0 = O$. Moreover, for $i \geq 1$, if $X$ is a metric space then we can choose $O_i$ to be open balls of smaller than radius $1/i$ while if $X$ is a locally compact space then we can choose $O_i$ such that $O_i$ is compact. In any case, we have constructed a sequence of either nested closed balls with radius going to zero or compact sets, and so, necessarily $\bigcap_i O_i = K \neq \emptyset$ and $K \subset D_i \cap O$, for every $i \geq 1$. Hence $O$ intersects the countable union $\bigcap_i D_i$, i.e., $\bigcap_i D_i$ is a dense set in $X$.

Finally, we show that if $A$ is a set of first category then its complement $X \setminus A$ is dense in $X$. Indeed, suppose $A = \bigcup_i A_i$ with $A_i$ nowhere dense, since $D_i = X \setminus A_i$ forms a sequence of dense open sets, the intersection

$$\bigcap_i D_i = X \setminus \bigcup_i A_i \subset X \setminus \bigcup_i A_i = X \setminus A$$

is also dense in $X$. In particular, if $U$ is an open subset of first category in $X$ then its complement $X \setminus U$ is dense, and because its is also closed, we deduce that $U$ must be empty, i.e., all open sets are of second category in $X$. \hfill \square

• Remark 2.12. The concepts of a space of second category and a set of second category in a given topological space may differ one of the other. For instance, as mentioned early, a closed set $A$ with an empty interior of a topological space $X$ is a set of first category (actually nowhere dense) in $X$. However, if the metric space $X$ is complete then any closed set $A$ can be regarded as a complete metric space, and the previous Theorem 2.11 affirms that $A$ is a space of second category, when considered with the relative topology. \hfill \square

The reader may want to check the book Bachman and Narici [14, Chapter 6, pp 74–84] for a quick context on this section, e.g., it is deduced that any everywhere dense $G_\delta$-set is also a set of second category. Based on Baire category arguments we show three far-reaching principle relative to continuous linear mappings, namely: uniformly boundedness, interior (or open) mapping, and closed graph principles.

### 2.3 Three Essential Principles

Recall that a **topological vector space** $X$ is a vector space with a Hausdorff topology for which the addition of vector and the scalar multiplication are continuous operations. In particular, the topology results invariant under translation, and a basis has the form $\{x + U_i : i \in I, x \in X\}$ with $\{U_i : i \in I\}$ a base of open sets around the origin. Moreover, in view of the continuity of the scalar multiplication, the open sets of the base can be chosen balanced, i.e., satisfying $tU_i \subset U_i$ for every $|t| \leq 1$ and any index $i$ in $I$. Indeed, if $U$ is a neighborhood of zero then by continuity, there exist $\delta > 0$ and a neighborhood $V$ of zero such that $rV \subset U$, for any $|r| < \delta$. The set $W = \bigcup_{|r| < \delta} rV$ satisfies $sW \subset W$, for every $|s| \leq 1$, and so $W$ is a balanced neighborhood of zero contained in $U$.
Clearly, a normed space is a topological vector space. Usually, an invar-
ant metric on topological vector space is constructed from a sequence \( \{d_k\} \) of translation invariant pseudo-metrics by means of \( d(x, y) = \sum_k 2^{-k}d_k(x, y)/(1 + d_k(x, y)) \). Sometimes, the attention is directed to the topology and the topolog-
al vector space is called metrizable. If the sequence of pseudo-metrics is induced by seminorms, i.e., \( d_k(x, y) = p_k(x - y) \), then the topological vector space is called a seminormed (Hausdorff) space or quasi-normed space, but this is not a standard terminology; see locally convex topological spaces later on.

- **Remark 2.13.** The continuity of the scalar multiplication and the addition in a topological vector space \( X \) implies that for every open set \( U \) containing zero there exists another open set \( V \) containing zero such that the closure \( \overline{V} \subset U \). Indeed, because \( 0 + 0 + 0 - 0 = 0 \), given an open set \( U \) containing zero, there exists a balanced open set \( V \) containing zero such that \( 0 + V + V - V \subset U \). This implies that \( 0 + V + V \cap ((X \setminus U) + V) \), i.e., the closure \( \overline{V} \subset U \). Actually, if \( K \) is a compact set and \( C \) is a closed set such that \( K \cap C = \emptyset \), then there exists an open set \( V \) containing zero such that \( (K + V) \cap (C + V) = \emptyset \), e.g., see Rudin [109, Theorem 1.10, pp. 9–10].

It is clear that a sequence \( \{x_n\} \) of vectors in \( X \) is a Cauchy sequence if for every neighborhood \( U \) of the origin there exists an index \( N = N(U) \) such that \( x_n - x_m \in U \), for every \( n, m \geq N \). Thus, a topological vector space \( X \) is called (a) complete if any Cauchy sequence is convergent, and (b) \( F \)-space if it is complete and metrizable (with an invariant metric), i.e., the topology is given by a metric \( d \) with the property that \( d(x, y) = d(x - y, 0) \), for every \( x, y \) in \( X \). Perhaps a typical example of a \( F \)-space is the \( L^p \), with \( 0 < p < 1 \), of all measurable functions \( f \) such that

\[
\|f\|_p = \int_{\Omega} |f(x)|^p \, dx < \infty,
\]

with the (translation invariant, non-homogeneous) metric \( d(f, g) = \|f - g\|_p \). Note that a Polish space (i.e., a complete separable metrizable space) is not necessarily a topological vector space, but a separable \( F \)-space is a Polish space. Also, the continuity of the scalar multiplication implies that the only open linear subspace of a vector topological space is the whole space.

- **Remark 2.14.** If \( F \) is a non-null linear functional on a topological vector space \( X \) then the following assertions are equivalent: (a) \( F \) is continuous, (b) the null space or kernel \( \mathcal{N}(F) = \{x \in X : Fx = 0\} \) is closed, (c) \( \mathcal{N}(F) \) is no dense in \( X \), (d) \( F \) is bounded in some neighborhood of zero. Indeed, (a) implies (b), which implies (c) are quite obvious. Now, if (c) holds and \( x \) belongs to the interior of the complement of \( \mathcal{N}(F) \) then there is a balanced neighborhood of zero \( V \) such that \( (x + V) \cap \mathcal{N}(F) = \emptyset \). Because \( F \) is linear and \( V \) balanced, the image \( F(V) \) is also a balanced set of scalars, i.e., either \( F(V) \) is bounded (i.e., (d) holds true) or \( F(V) \) is the whole scalar field, which means that there exists \( y \) in \( V \) such that \( Fy = -Fx \), and so \( x + y \) belongs to \( \mathcal{N}(F) \), i.e., the contradiction \( (x + V) \cap \mathcal{N}(F) \neq \emptyset \). Finally, if (d) holds then there exist a constant \( C > 0 \) and a neighborhood of zero \( V \) such that \( |Fx| \leq C \), for every \( x \) in \( V \). Thus, for every
\(\varepsilon > 0\) there exists a neighborhood of zero \(W = (\varepsilon/C)V\) such that \(|Fx| < \varepsilon\), for every \(x\) in \(W\), which means that \(F\) is continuous at zero, and by linearity, continuous everywhere. \(\square\)

- **Remark 2.15.** Any one-to-one linear mapping \(T\) from the Euclidean space onto a finite-dimensional linear subspace \(Y\) of a topological vector space is necessarily bi-continuous (i.e., \(T\) and \(T^{-1}\) are continuous), and thus \(Y\) is a closed linear subspace. Indeed, the argument is by induction on the dimension of \(Y\). If \(T : \mathbb{R}^n \to Y\) then the continuity of the scalar multiplication shows that \(T\) is continuous. For \(n = 1\), the operator \(T^{-1}\) is a functional, and Remark 2.14 (b) implies its continuity. Now, assuming the assertion true for dimension \(n - 1\), if \(\{e_1, \ldots, e_n\}\) is the canonical basis in \(\mathbb{R}^n\) then \(\{u_i = Te_i, i = 1, \ldots, n\}\) is a basis on \(Y\) and there exist linear functionals \(F_i\) such that \(x = F_1(x)e_1 + \cdots + F_n(x)e_n\). The null space of each \(F_i\) is a linear subspace of dimension \(n - 1\), which is closed by the induction assumption, and again Remark 2.14 (b) yields the continuity of each \(F_i\). Hence the inverse \(T^{-1}(x) = (F_1(x), \ldots, F_n(x))\), is continuous, and the argument is completed. \(\square\)

- **Remark 2.16.** If \(N\) is a closed subspace of a topological vector space \(X\) then the quotient (vector) space \(X/N\) of elements of the form \(\bar{x} = x + N\) and with the quotient topology given by the open set of the form \(\bar{U} = U + N\) with \(U\) open in \(X\). The mapping \(x \mapsto \bar{x}\) is linear and continuous from \(X\) onto \(X/N\) and \(N\) is its null space or kernel. In general, this quotient operation preserve most properties, namely, \(X/N\) is a topological vector space of the same type as \(X\), e.g., if \(X\) is a Banach space or an \(F\)-space then so is the quotient space \(X/N\). For instance, we define

\[
\|\bar{x}\| = \inf \{\|x + n\| : n \in N\} \quad \text{and} \quad d(\bar{x}, 0) = \inf \{d(x, n) : n \in N\}
\]

to obtain a norm or an invariant metric in \(X/N\). Certainly, this also applies to the locally convex spaces discussed later. For instance, to check that the quotient space \(X/N\) of a \(F\)-space \(X\) is complete, take a Cauchy sequence \(\{\bar{x}_k\} \subset X/N\). By extracting a subsequence, we may suppose that \(d(\bar{x}_k - \bar{x}_{k+1}, 0) < 2^{-k}\), for every \(k \geq 1\), and by induction, there are points \(x_k\) in \(\bar{x}_k\) such that \(d(x_{k+1} - x_k, 0) + 2^{-k} \geq d(x_{k-1} - x_k, 0)\), for every \(k \geq 1\), with \(x_0 = 0\). Thus \(\{x_k\}\) is a Cauchy sequence in \(X\), and it must converge, \(x_k \to x\). Since \(d(\bar{x}_k, \bar{x}) \leq d(x_k, x)\), the sub-sequence \(\bar{x}_k \to \bar{x}\), and therefore the whole sequence converges. \(\square\)

### 2.3.1 Uniformly Boundedness Principle

A family of linear operators \(T_i : X \to Y, i \in I\), between two topological vector spaces is called *equi-continuous* if for any neighborhood \(V\) of 0 in \(Y\) there exists a neighborhood \(U\) of 0 in \(X\) such that \(T_i(U) \subset V\), for every \(i \in I\). Certainly any linear operator in the family must be continuous and any finite family of continuous linear operators is equi-continuous. Moreover, every equi-continuous family of linear operator is equi-bounded in the sense that the set \(\{T_i(x) : x \in B, i \in I\}\) is a bounded set in \(Y\) for any bounded set \(B\) in \(X\). Indeed, if \(B\)
is bounded then there exists an scalar \( r \) such that \( B \subseteq rU \). Hence \( T_i(B) \subseteq T_i(rU) \subseteq rV \), which show that \( \{ T_i(x) : x \in B, i \in I \} \subseteq rV \) as desired.

The converse is given by the following version of the Banach-Steinhaus Theorem uniformly boundedness principle.

**Theorem 2.17.** Let \( \{ T_i : i \in I \} \) be family of continuous linear operators \( T_i : X \to Y \) between two topological vector spaces, and denote by \( X_0 \) the set of all \( x \) in \( X \) such that \( \{ T_i(x) : i \in I \} \) is a bounded set in \( Y \). If \( X_0 \) is a second category set then \( X_0 = X \), the family \( \{ T_i : i \in I \} \) is equi-continuous, and \( \{ T_i(x) : i \in I, x \in B \} \) is a bounded set in \( Y \), for any bounded set \( B \) in \( X \).

**Proof.** Since the addition is a continuous operation (see Remark 2.13), there exist (balanced) neighborhoods \( W \) and \( V \) such that \( W - W \subseteq V \). The pre-image of a closed set by a continuous mapping is closed, and so, the set \( F = \bigcap_i T_i^{-1}(W) \) is closed.

Because the set \( T(x) \) is bounded for any \( x \) in \( X_0 \), there exists \( n \) such that \( T(x) \subseteq nW \), i.e., \( T_i(x) \subseteq nW \) or equivalently \( x \in nT_i^{-1}(W) \), for every \( i \), so \( x \in nF \), which yields \( X_0 \subseteq \bigcup_{n=1}^{\infty} nF \). Since \( X_0 \) is of second category, at least one \( nF \) should not be nowhere dense, and because \( x \mapsto nx \) is a homeomorphism from \( X \) into itself, \( F \) cannot be nowhere dense in \( X \). As mentioned early, \( F \) is closed because each \( T_i \) is continuous and therefore, there is an interior point \( x \) in \( F \). Thus, \( x - F \) contains a neighborhood \( U \) of 0 in \( X \), and

\[
T_i(U) \subseteq T_i(x) - T_i(F) \subseteq W - W \subseteq V, \quad \forall i \in I,
\]
i.e., \( \{ T_i : i \in I \} \) is equi-continuous.

Now, for any given bounded set \( B \), there exists \( r > 0 \) such that \( rU \supseteq B \). Hence

\[
T_i(B) \subseteq T_i(rU) = rT_i(U) \subseteq rV, \quad \forall i \in I.
\]

Because the neighborhood \( V \) is arbitrary, the set \( \{ T_i(x) : i \in I, x \in B \} \) is bounded.

In particular, because each point \( x \) in \( X \) is a bounded set, the set \( \{ T_i(x) : i \in I \} \) is bounded, and by definition \( x \) belongs \( X_0 \), i.e., \( X_0 = X \). \( \square \)

A typical way to apply the Banach-Steinhaus Theorem 2.17 is as follows

**Corollary 2.18.** If a sequence \( \{ T_i : i \geq 1 \} \) of continuous linear operators from a \( F \)-space \( X \) into a topological vector space \( Y \) is pointwise convergence, i.e., the limit \( T(x) = \lim_i T_i(x) \) exits in \( Y \), for every \( x \) in \( X \), then \( T \) is a continuous linear operator.

**Proof.** Recall that a \( F \)-space is a complete and metrizable topological vector space, i.e., \( X \) is of second category. Moreover, because any convergent sequence is bounded, by means of Theorem 2.17, we deduce that \( \{ T_i : i \geq 1 \} \) is equi-continuous. This is, for every neighborhood \( V \) of 0 there exist neighborhoods \( W \) and \( U \) of 0 such that \( T_i(U) \subseteq W \subseteq V \), for every \( i \geq 1 \), which implies that \( T(U) \subseteq V \) and so \( T \) is a linear continuous operator. \( \square \)
Remark 2.19. If $X$ and $Y$ are normed spaces, a linear mapping $T : X \to Y$ is continuous if and only if the quantity $\|T\| = \sup_{\|x\| \leq 1} \|Tx\|$ is finite. The space of all continuous linear operators is denoted by $L(X,Y)$ and endowed with the norm $\| \cdot \|$ is a normed space, which is complete if $Y$ is so. In this case, the uniformly boundedness principle reads as follows: If $\{T_i : i \in I\}$ is a family of continuous linear operator form a Banach space $X$ into a normed space $Y$ such that

$$\sup_{i \in I} \|T_i(x)\|_Y < \infty, \quad \forall x \in X,$$

then $\{\|T_i\| : i \in I\}$ is a bounded numerical set.

\[ \square \]

2.3.2 Open Mappings Theorem

Together with the uniformly boundedness principle, there is a couple of basic results. The first refers to Banach-Schauder or open mappings Theorem, i.e., operator between topological vector spaces transforming open sets into open sets.

**Theorem 2.20.** Let $T$ be a continuous linear operator $T : X \to Y$, where $X$ is an $F$-space, $Y$ is a topological vector space, and the image $T(X)$ is a second category set in $Y$. Then (1) $T(X) = Y$, (2) $Y$ is an $F$-space, and (3) $T$ is an open mapping, i.e., if $U$ is an open set in $X$ then $T(U)$ is an open set in $Y$.

**Proof.** Recall that a $F$-space is a complete topological vector space with an invariant metric defining its the topology, and let $d$ such a metric in $X$. To prove (3), it suffices to show that $T$ is open at $0$, i.e., for every $r > 0$ the image $T(U)$ of the ball $U = \{x \in X : d(x,0) < r\}$ is a neighborhood of $0$ in $Y$. To this purpose, define

$$U_n = \{x \in X : d(x,0) < 2^{-n}r\}, \quad \forall r' < r, \ n = 1, 2, \ldots,$$

to show first that there exists a neighborhood $V$ of $0$ in $Y$ such that $V \subset \overline{T(U_1)}$.

Indeed, $U_2 - U_2 \subset U_1$ yields

$$T(U_2) - T(U_2) \subset \overline{T(U_2)} - T(U_2) \subset \overline{T(U_1)}.$$

Let us check that $\overline{T(U_1)}$ is a neighborhood of $0$. Because $U_2$ is a neighborhood of $0$, we have the equality $T(X) = \bigcup_n nT(U_2)$. Since $T(X)$ is of second category, at least one $nT(U_2)$ has to be of second category in $Y$, and because the multiplication by a nonzero scalar is an homeomorphism, $T(U_2)$ is of second category in $Y$ and so its closure has a nonempty interior, denoted by $V_2$. Then $V = V_2 - V_2$ is an open set in $Y$ containing $0$ and $V \subset \overline{T(U_1)}$.

Now, we show that $\overline{T(U_1)} \subset T(U)$. Indeed, let $y_1$ be in $\overline{T(U_1)}$ and define $y_n$ in $\overline{T(U_n)}$ by induction, for every $n \geq 2$, as follows. The previous argument can be applied to $V_k$ ($k \geq 2$) instead of $V_1$, and so $\overline{T(U_n+1)}$ contains a neighborhood of $0$. Hence, because $y_n$ belongs to the closure of $T(U_n)$, we have $(y_n - T(U_{n+1})) \cap
$T(U_n) \neq \emptyset$, i.e., there exists $x_n$ in $U_n$ such that $Tx_n$ belongs to $y_n - \overline{T(U_{n+1})}$, which allows us to set $y_{n+1} = y_n - Tx_n$. Now, since $d(x_n, 0) < 2^{-n}r'$ for $n \geq 1$, the sum $x_1 + x_2 + \cdots + x_n$ is a Cauchy sequence and so it converges to some $x$ in $X$ with $d(x, 0) \leq r' < r$, i.e., $x$ belongs to $U$. The telescoping series

$$
\sum_{n=1}^{m} Tx_n = \sum_{n=1}^{m} (y_n - y_{n+1}) = y_1 - y_{m+1}
$$

and the continuity of the operator $T$ imply that $y_1 = Tx$, i.e., $y_1$ belongs to $U$. This completes (3).

Since the only open subspace is the whole space $Y$ and $T(X)$ is indeed a subspace, we deduce that (3) implies (1).

Finally, following Remark 2.16 with $N$ the null space of $T$, we obtain a quotient space $X/N$ which is an $F$-space. Because $X/N$ is homeomorphic to $T(X) = Y$, we deduce that $Y$ is an $F$-space. \hfill \Box

- **Remark 2.21.** Recalling that in a topological vector space the topology is given by a base of balanced open sets containing the origin, note that the key result within the proof is the following assertion: if $T : X \to Y$ is a continuous linear mapping between two topological vector spaces such that the range of $T$ is a set of second category in $Y$ and $U$ is a neighborhood of $0$ in $X$ then the closure of the image of $U$, i.e., $\overline{TU}$, contains a neighborhood of $0$ in $Y$. \hfill \Box

- **Remark 2.22.** A typical application of the open mapping Theorem 2.20 is the case of a continuous onto mapping $T$ between two $F$-spaces. This is, if $T^{-1}$ exists (i.e., $T$ is also one-to-one) then $T^{-1}$ is also continuous. By means of the quotient spaces, even when $T$ is not one-to-one, the openness of $T$ means that for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $d(Tx, 0) < \delta$ then there exists $z$ in $X$ satisfying $Tz = 0$ and $d(x - z, 0) < \varepsilon$. \hfill \Box

### 2.3.3 Closed Graph Theorem

The second result has to do with the notion of *closed operators* or *closed graphs*, i.e., the graph $\{(x, Tx) : x \in X\}$ is a closed set in the product $X \times Y$. In the case of $F$-spaces, the graph is closed if and only if for any sequence $\{x_n\}$ such that $x_n \to x$ in $X$ and $T(x_n) \to y$ in $Y$ we have $y = T(x)$.

**Theorem 2.23.** Let $T$ be a closed linear operator $T : X \to Y$, where $X$ and $Y$ are $F$-spaces. Then $T$ is also a continuous linear operator.

**Proof.** Let $G$ be the graph of $T$, i.e., $G = \{(x, Tx) : x \in X\}$ considered as a subspace of the product $X \times Y$, which is also an $F$-space. Since closed subspaces of a complete metric space is complete and because $T$ is closed, we deduce that $G$ is an $F$-space.

Consider the linear and continuous applications $a : G \to X$ and $b : X \times Y \to Y$ defined by $a(x, Tx) = x$ and $b(x, y) = y$. Since $a$ is a one-to-one mapping from the $F$-space $G$ onto the $F$-space $X$, the open mappings Theorem 2.20 can be used to deduce that $a^{-1} : X \to G$ is continuous. Hence, the composition $b \circ a^{-1} = T$ is also continuous. \hfill \Box
For instance, more details on a proof of the above open mappings and closed graphs Theorems 2.20 and 2.23 can be found in Rudin [109, Chapters 2, pp. 41–54]. In most of the cases, these three principles are very useful when dealing with Banach space, where the concept of bounded set is easier. Recall that the unit closed ball is compact only on a finite-dimensional Banach space.

**Exercise 2.3.** Let $V$ be a finite-dimensional linear subspace of a topological vector space $X$ and $p$ be a continuous seminorm on $X$ such that $p(v) = 0$ and $v$ in $V$ imply $v = 0$. Take a basis $\{v_1, \ldots, v_n\}$ in $V$ and consider the continuous linear mapping $c = (c_1, \ldots, c_n)$ from $\mathbb{R}^n$ into $X$ defined by $Tc = c_1 v_1 + \cdots + c_n v_n$. First (1) minimize the real-valued function $c \mapsto p(Tc)$ over the region $\{c : |c_1| + \cdots + |c_n| = 1\}$, and then (2) prove the estimate

$$|c_1| + \cdots + |c_n| \leq K p(Tc), \quad \forall c \in \mathbb{R}^n$$

for some constant $K > 0$. Finally, (3) deduce that $T^{-1} : V \to \mathbb{R}^n$ is also continuous and therefore $V$ is closed in $X$. \hfill \Box

**Exercise 2.4.** On a given topological vector space $X$, (1) recall the definition of sequentially compact and bounded sets, and (2) prove that any sequentially compact set $A \subset X$ is also a bounded set. Next, (3) show that every topological vector space $X$ having a compact neighborhood of zero is finite dimensional. \hfill \Box

### 2.3.4 Hahn-Banach Theorem

Sometimes, the open mappings and closed graphs Theorems 2.20 and 2.23 are tied together, leaving room for another principle.

Recall that a real linear functional $f$ on a (linear) vector space $X$ is a linear mapping from $X$ into $\mathbb{R}$, i.e., satisfying $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$, for every $x, y$ in $X$ and $\alpha, \beta$ in $\mathbb{R}$. By making use of the axiom of choice or Zorn’s Lemma, we have

**Lemma 2.24** (Hahn-Banach). *Let $X$ be a real linear vector space and $p$ be a function from $X$ into $[0, \infty)$ satisfying

$$p(x + y) \leq p(x) + p(y), \quad p(\lambda x) \leq \lambda p(x), \quad \forall \lambda \geq 0, \ x, y \in X.$$*

If $X_0$ is a linear subspace of $X$ and $f_0 : X_0 \to \mathbb{R}$ is a linear mapping satisfying $f_0(x) \leq p(x)$ for every $x$ in $X_0$, then there exists a real linear function $f$ on $X$ such that $f(x) \leq p(x)$ for every $x$ in $X$ and $f(x) = f_0(x)$, for any $x$ in $X_0$.

**Proof.** Consider the family $\mathcal{A}$ of all pairs $(X_0, f_0)$, where $X_0$ is a linear subspace containing $X_0$ and $f_0$ is a real linear functional defined on $X_0$ such that $f_0(x) \leq p(x)$ for every $x$ in $X_0$ and $f_0(x) = f_0(x)$, for any $x$ in $X_0$. Therefore, an order relation can be given in $\mathcal{A}$ by the condition $(X_0, f_0) \prec (X_0, f_0)$ if $X_0 \subset X_0$ and $f_0 = f_0$ on $X_0$.

Since any chain or totally ordered subset of $\{(X_0, f_0)\}$ has $(\bigcup_\beta X_0, f_0')$, with $f_0' = f_0$ on $X_0$, as an upper bound, we can use Zorn’s Lemma to find a maximal
element \((\hat{X}, \hat{f})\). We show that \(\hat{X} = X\) by contradiction. Indeed, if \(x_1\) belongs to \(X \setminus \hat{X}\) then the subspace \(X_2 = \{x + \lambda x_1 : x \in \hat{X}, \lambda \in \mathbb{R}\}\) strictly contains the subspace \(\hat{X}\), and based on the inequality \(-p(-y-x_1) - f_1(y) \leq p(x+x_1) - f_1(x)\), for every \(x, y\) in \(X\), we can find a constant \(c\) such that
\[
\sup_{y \in X} \{ -p(-y-x_1) - f_1(y) \} \leq c \leq \inf_{x \in X} \{ p(x+x_1) - f_1(x) \}.
\]
Hence, we can define the linear function \(f_2(x + \lambda x_1) = f_1(x) + \lambda c\), which satisfies
\[
f_2(x + \lambda x_1) = f_1(x) + \lambda c \leq p(x + \lambda x_1), \quad \forall x \in \hat{X}, \lambda \in \mathbb{R},
\]
to obtain a contradiction. \(\Box\)

- **Remark 2.25.** Since the real and imaginary parts of a complex number satisfy \(z = \Re(z) + i\Im(z)\) and \(\Im(z) = -\Re(iz)\), if \(Z\) is a vector space over the complex number \(\mathbb{C}\) and \(F\) is a linear functional on \(Z\) then the real part \(f(z) = \Re(F(z))\) and the imaginary part \(g(z) = \Im(F(z))\) are \(\mathbb{R}\)-linear functional on \(Z\) (i.e., when \(Z\) is considered as a vector space on the real number \(\mathbb{R}\)) satisfying \(g(z) = -if(iz)\). Conversely, if \(f\) is a \(\mathbb{R}\)-linear functionals on \(Z\) then \(F(z) = f(z) - if(iz)\) is a linear functional on \(Z\). With this argument, Hahn-Banach Lemma 2.24 can be extended to vector spaces on the complex number, e.g., see Swartz [119, Chapter 8, Lemma 3, p. 76]. \(\Box\)

Based on the above observation, the vector space \(X\) could be on the complex number, and the functional \(f_0\) could be complex-valued if \(|f_0(x)| \leq p(x)\) and \(p(x) \leq |\lambda|p(x)\) for every \(x\) in \(X\). Most of the time, the notation \(\langle f, x \rangle = f(x)\) is used with linear functionals.

Recall that functions \(p\) with the property required in Lemma 2.24 are better understood in a stronger form, namely, seminorms, i.e., \(p : X \to [0, \infty)\), where one requires the triangular inequality \(p(x + y) \leq p(x) + p(y)\), for every \(x, y\) in \(X\) and the homogeneity condition \(p(\lambda x) = \lambda p(x)\). It is also clear that the triangular inequality can be rewritten as \(|p(x) - p(y)| \leq p(x - y)|\), for every \(x, y\) in \(X\) and that necessarily \(p(0) = 0\), but \(p(x)\) may vanish for some \(x \neq 0\). If \(X\) is a topological vector space then the interest is on continuous seminorms.

The following version of the Hahn-Banach Theorem is a direct consequence of Lemma 2.24, applied for the real-valued functional \(x \mapsto \Re(\langle f, x \rangle)\), or equivalently, writing \(\langle f, x \rangle = |\langle f, x \rangle|e^{i\theta}\).

**Theorem 2.26.** Let \(X_0\) be a subspace of a topological vector space \(X\) and \(p\) be a continuous seminorm on \(X\). If \(f_0\) is a linear functional on \(X_0\) satisfying \(|\langle f_0, x \rangle| \leq p(x)\), for every \(x\) in \(X_0\), then \(f_0\) can be extended to a continuous linear functional on \(X\), namely, there exists a continuous linear functional on \(X\) such that (a) \(\langle f, x \rangle = \langle f_0, x \rangle\) for every \(x\) in \(X_0\) and (b) \(|\langle f, x \rangle| \leq p(x)\) for every \(x\) in \(X\).

- **Remark 2.27.** A direct consequence is the separation of convex sets, namely, if \(A\) and \(B\) are two disjoint non-empty convex subsets of a real normed vector space \(X\), and one of them has a non-empty interior, then \(A\) and \(B\) can be
called lower semi-continuous (d) if
satisfying \( \ell \) above, a seminorm on topological vector space \( X \) is a functional on \( X \) such that (a) \( p(x) \geq 0 \), (b) \( |p(x) - p(y)| \leq p(x - y) \), (c) \( p(\lambda x) = |\lambda| p(x) \).
Moreover, \( p \) is called lower semi-continuous (d) if \( x_n \to x \) then \( p(x) \leq \lim \inf_n p(x_n) \), or equivalently, if the set \( \{ p \leq 1 \} = \{ x : p(x) \leq 1 \} \) is closed. It is clear that if \( p_1, \ldots, p_n \) are seminorms then \( x \mapsto \max_i p_i(x) \), \( x \mapsto \sum_i a_ip_i(x) \), with scalar \( a_i \geq 0 \), are also seminorms.

Furthermore, a subset \( B \) of \( X \) is called bounded if for any neighborhood of zero \( U \) there exists a constant \( \alpha > 0 \) such that \( B \subset \alpha U \), for every \( |\beta| \geq \alpha \).
Now, we may re-phase the open mapping Theorem 2.20 as follows (of which Banach space version is sometimes known as Gelfand Theorem):

**Proposition 2.28.** If \( p \) is a seminorm on a \( F \)-space \( (X, d) \) then each of the following conditions are equivalent: (1) \( p \) is continuous at zero, (2) \( p \) is continuous, (3) \( p \) is lower semi-continuous, (4) \( \{ x : p(x) \leq 1 \} \) is a neighborhood of zero, (5) \( p \) maps bounded set into bounded sets.

**Proof.** Since \( |p(x) - p(y)| \leq p(x - y) \), we deduce that (1) \( \Rightarrow \) (2). It is clear that (2) \( \Rightarrow \) (3).
(3) \( \Rightarrow \) (4) is essentially the key argument in open mapping Theorem 2.20. If \( p \) is a lower semi-continuous semi-norm on \( X \) then the set \( \{ p \leq 1 \} \) is a neighborhood of zero. Indeed, since \( \{ p \leq n \} = n \{ p \leq 1 \} \) is a closed set and \( X = \bigcup_n \{ p \leq n \} \), the Baire Category Theorem 2.11 implies the existence of an open set \( U \) inside some set \( \{ p \leq n \} \). This means that \( U - U \) is inside \( \{ p \leq 2n \} \), i.e., \( \{ p \leq 2n \} \) (or equivalently \( \{ p \leq 1 \} \)) is a neighborhood of zero.

To check that (4) \( \Rightarrow \) (5), for any bounded set \( B \) find a constant \( \alpha > 0 \) such that \( B \subset \alpha \{ p \leq 1 \} \), which means that \( p(B) \) is inside the bounded interval \( [0, \alpha] \).

Finally, (5) \( \Rightarrow \) (1), to check that \( p \) is continuous at zero, take a sequence \( \{ x_n \} \) converging to zero and consider the set \( \{ d < r \} = \{ x : d(x) < r \} \) with \( r = \sup_n d(x_n, 0) < \infty \). Because \( \{ d < r \} \) is a bounded set the image \( p(\{ d < r \}) \) is also a bounded set (in \( \mathbb{R} \)), i.e., \( 0 \leq p(x) \leq C \) for every \( x \) in \( \{ d < r \} \) and some constant \( C > 0 \). For every \( \varepsilon > 0 \) the set \( (C/\varepsilon) \{ d < r \} \) is open and thus, there exists \( N \) such that \( x_n \) belongs to \( (C/\varepsilon) \{ d < r \} \) for every \( n \geq N \), i.e., \( (\varepsilon/C)x_n \) is in \( \{ d < r \} \). Hence \( p((C/\varepsilon)x_n) \leq C \), i.e., \( p(x_n) \leq \varepsilon \), which proves that \( p(x_n) \to 0 \).

**Remark 2.29.** The arguments in the proof of Proposition 2.28 are simpler when \( X \) is a Banach with norm \( \| \cdot \| \). For instance, from (4) follows that \( \{ p \leq 1 \} \supset \{ x : \| x \| \leq \alpha \} \), for some \( \alpha > 0 \), i.e., \( \| x \| \leq 1 \) implies \( p(x) \leq \alpha \), so (5).
Next, to check that (5) \( \Rightarrow \) (1), applied (5) to the bounded set \( x/\| x \| \) to obtain \( p(x) \leq C \| x \| \) for every \( x \) and some constant \( C > 0 \), which yields (1). In any way, since any bounded set in normed space \( X \) is covered by a ball, we deduce that (5) and (6) are equivalent.
For instance, the reader interested in a more convex analysis oriented course may take a look at Clarke [27, Part I, pp. 1–169], among many other books.

### 2.4 More on Lebesgue Spaces

The following theorem is perhaps the most important result relative to the theory of set functions, and a proof can be found in Dunford and Schwartz [39, Vol 1, Section III.7, pp. 155–164] or Yosida [135, Section II.2, pp. 70–72]. An additive set function \( \lambda : \mathcal{F} \to \mathbb{R}^d \) is called \( \mu \)-continuous if for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that for every \( F \in \mathcal{F} \) with \( \mu(F) < \delta \) we have \( |\lambda(F)| < \varepsilon \), see Definition 1.28.

**Theorem 2.30 (Vitali-Hahn-Saks).** Let \( (\Omega, \mathcal{F}, \mu) \) be a measure space and \( \{\lambda_n\} \) be a sequence of \( \mu \)-continuous additive set functions \( \mathbb{R}^d \)-valued. If the limit \( \lim_{n} \lambda_n(A) \) exists and is finite for every \( A \) in \( \mathcal{F} \) then \( \{\lambda_n\} \) is \( \mu \)-uniformly continuous, i.e., for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that for every \( F \) in \( \mathcal{F} \) with \( \mu(F) < \delta \) we have \( \sup_{n} |\lambda_n(F)| < \varepsilon \).

**Proof.** First, recall that the set \( L^1(\Omega, \mathcal{F}, \mu) \) of all real-valued (or complex-valued) integrable functions is a Banach space and the subset \( \mathfrak{F}_0(\mu) \) of all functions integrable functions \( f = \mathds{1}_F \), with \( F \) in \( \mathcal{F} \) is a closed (just use the fact that from any convergence sequence in \( L^1 \) we can a convergent subsequence almost everywhere), but certainly, \( \mathfrak{F}_0(\mu) \) is not a vector subspace. Thus \( \mathfrak{F}_0(\mu) \) is a complete metric space and the space of simple functions (almost measurable functions taking a finite number of values) \( S^1(\Omega, \mathcal{F}, \mu) \) is the linear vector space generated by \( \mathfrak{F}_0(\mu) \) is a dense subspace in \( L^1(\Omega, \mathcal{F}, \mu) \).

Moreover, the complete metric space \( \mathfrak{F}_0(\mu) \) can be also regarded as the sets in \( \mathcal{F} \) with finite \( \mu \)-measure and identified \( \mu \)-almost everywhere, where the distance is given by \( d(A, B) = \mu(A \cup B \setminus A \cap B) \). If \( \mathfrak{F}(\mu) \) denotes the elements in \( \mathcal{F} \) with the \( \mu \)-almost everywhere equality and the distance \( d(A, B) = \arctan \left( \mu(A \cup B \setminus A \cap B) \right) \), then \( \mathfrak{F}(\mu) \) is also a complete metric space.

Since \( \lambda_n \) is \( \mu \)-continuous, we may consider \( \lambda_n \) as a \( \mathbb{R}^d \)-valued continuous function on \( F(\mu) \). Thus

\[
\mathcal{F}_{n,\varepsilon} = \left\{ A \in \mathfrak{F}(\mu) : \sup_{k \geq 1} |\lambda_n(A) - \lambda_{n+k}(A)| \leq \varepsilon \right\}, \quad \forall n \geq 1, \forall \varepsilon > 0,
\]

is a closed subset of \( \mathfrak{F}(\mu) \) and because the limit \( \lim_n \lambda(A) \) exists and is finite, we have the equality \( \mathfrak{F}(\mu) = \bigcup_n \mathcal{F}_{n,\varepsilon} \) for every \( \varepsilon > 0 \).

Any complete metric space is a second category set, in particular \( \mathfrak{F}(\mu) \) is a second category set and thus, at least one \( \mathcal{F}_{m,\varepsilon} \) must has nonempty interior (see Section 2.2.3 on Baire category arguments). Hence, there exists \( \delta > 0 \) and \( A_0 \) in \( \mathfrak{F}(\mu) \) such that

\[
d(A, A_0) < \delta \quad \text{implies} \quad \sup_{k \geq 1} |\lambda_m(A) - \lambda_{m+k}(A)| \leq \varepsilon.
\]
Thus, for any \( A \) in \( \mathcal{F}(\mu) \) with \( \mu(A) < \delta \) we take \( A_1 = A \cup A_0 \) and \( A_2 = A_0 \setminus A \cap A_0 \) to have \( A = A_1 \setminus A_2 \) and therefore

\[
|\lambda_n(A)| \leq |\lambda_m(A)| + |\lambda_m(A) - \lambda_n(A)| \leq \\
\leq |\lambda_m(A)| + |\lambda_m(A_1) - \lambda_n(A_1)| + |\lambda_m(A_2) - \lambda_n(A_2)| \leq \\
\leq |\lambda_m(A)| + 2\varepsilon, \quad \forall n \geq m,
\]

which shows that the sequence \( \{\lambda_n\} \) is \( \mu \)-uniformly continuous.

\( \square \)

**Corollary 2.31.** Let \( \{f_n\} \) be a bounded sequence in \( L^1(\Omega, \mathcal{F}, \mu) \) such that the limit

\[
I_n(A) = \int_A f_n \, d\mu, \quad \lim_n I_n(A) = I(A), \quad \forall A \in \mathcal{F}
\]

exists and is finite. Then \( I \) is \( \sigma \)-additive real-valued set function and the finite measures

\[
A \mapsto \nu_n(A) = \int_A |f_n| \, d\mu, \quad \forall A \in \mathcal{F}
\]

are uniformly \( \sigma \)-additive, i.e., if \( A_k \in \mathcal{F}, A_k \subset A_{k-1} \) and \( \bigcap_k A_k = \emptyset \) then \( \sup_n \nu_n(A_k) \rightarrow 0 \) as \( k \rightarrow \infty \). Moreover, for every \( \varepsilon > 0 \) there exists \( A \in \mathcal{F} \) with \( \mu(A) < \infty \) such that \( \nu_n(A^c) < \varepsilon \).

**Proof.** Define the finite measure

\[
\lambda(A) = \sum_{n=1}^{\infty} 2^{-n} \lambda_n(A) \quad \text{with} \quad \lambda_n(A) = \frac{1}{\|f_n\|_1} \int_A |f_n| \, d\mu,
\]

to get that \( \lambda(A_k) \rightarrow 0 \) as \( k \rightarrow \infty \).

Since \( I_n \) is \( \lambda \)-continuous and \( \lambda \) is finite, Vitali-Hahn-Saks Theorem 2.30 implies that \( \{I_n\} \) are uniformly \( \sigma \)-additive and thus, \( I \) is \( \sigma \)-additive. Actually, for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( \lambda(E) < \delta \) implies \( |I_n(E)| < \varepsilon \), for any \( n \). Now, consider

\[
I_n^+(A) = \int_A f_n^+ \, d\mu, \quad I_n^-(A) = \int_A f_n^- \, d\mu, \quad \forall A \in \mathcal{F}, \forall n.
\]

Since \( \lambda(A_k \cap B) \leq \lambda(A_k) \) for every \( B \in \mathcal{F} \) we have \( I_n(A_k \cap B) < \varepsilon \) if \( \lambda(A_k) < \delta \), and in particular for \( B = 1_{\{f_n > 0\}} \) we deduce \( I_n^+(A_k) < \varepsilon \). Similarly, we obtain \( I_n^-(A_k) < \varepsilon \), and therefore \( \nu_n(A) = I_n^+(A_k) + I_n^-(A_k) < 2\varepsilon \) if \( \lambda(A_k) < \delta \). Hence, \( \{\nu_n\} \) are uniformly \( \sigma \)-additive.

Finally, because each \( f_n \) is integrable, the set \( E = \bigcup_n \{f_n \neq 0\} \) is \( \sigma \)-finite, i.e., \( E = \bigcup_k E_k \) with \( E_k \subset E_{k+1} \), \( \mu(E_k) < \infty \) and also \( \nu_n(E^c) = 0 \). Therefore \( \nu_n(E^c) = \nu_n(E \setminus E_k) < \varepsilon \) if \( \lambda(E \setminus E_k) < \delta \), which must hold for \( k \) sufficiently large since \( \bigcap_k (E \setminus E_k) = \emptyset \). Thus, we choose \( A = E_k \) to conclude. \( \square \)
2.4. More on Lebesgue Spaces

2.4.1 Weak Convergence

Comparing with Definition 1.28, we see that the condition on the set \( A \) with finite measure is assured when the sequence \( \{f_n\} \) is weakly convergent. Thus, we make some comments on weak convergence.

**Definition 2.32.** A sequence \( \{f_n\} \) in \( L^p(\Omega, \mathcal{F}, \mu) \), \( 1 \leq p \leq \infty \), converges weakly to \( f \) if

\[
\lim_n \int_{\Omega} f_n g \, d\mu = \int_{\Omega} f g \, d\mu, \quad \forall g \in L^q(\Omega, \mathcal{F}, \mu),
\]

where \( 1/p + 1/q = 1 \), and with brackets this is written as \( \langle f_n, g \rangle \to \langle f, g \rangle \), even sometimes, the notation \( f_n \rightharpoonup f \) is used. In this context, the convergence in norm (i.e., when \( \|f_n - f\|_p \to 0 \)) is called **strong convergence** and usually denoted by \( f_n \to f \).

Recall that \( q = \infty \) when \( p = 1 \) and that any function (actually equivalence class) \( f(x) \) belonging to the Banach space \( L^\infty(\Omega, \mathcal{F}, \mu) \) includes the condition of \( \sigma \)-finite non-zero range, i.e., the set \( \{x \in \Omega : |f(x)| \neq 0\} \) is a countable union of sets with \( \mu \)-finite measure. Also note that technically, the weak convergence defined above for \( L^\infty \) is actually called **weak* convergence**, in the general context of Banach and dual spaces. On the other hand, as expected, by means of Hölder inequality,

\[
|\langle f_n - f, g \rangle| \leq \|f_n - f\|_p \|g\|_q,
\]

we show that weak convergence implies strong convergence.

**Proposition 2.33.** If \( \{f_n\} \) is a sequence in \( L^p(\Omega, \mathcal{F}, \mu) \) weakly convergent to \( f \) then

\[
\|f\|_p \leq \liminf_n \|f_n\|_p, \tag{2.2}
\]

for any \( 1 \leq p \leq \infty \), i.e., the norm \( \|\cdot\|_p \) is a weakly lower semi-continuous function.

**Proof.** Assume first \( 1 \leq p < \infty \). Since the function \( g = |f|^{p/q}\text{sign}(f) \) belongs to \( L^q \), with \( 1/p + 1/q = 1 \), we have \( \langle f_n, g \rangle \to \langle f, g \rangle = \|f\|_p^p \). However, Hölder inequality implies

\[
|\langle f_n, g \rangle| \leq \|f_n\|_p \|g\|_p = \|f_n\|_p \|f\|_p^{p/q}
\]

and (2.2) for \( p < \infty \).

For \( p = \infty \), we may assume that \( \|f\|_p > 0 \) and that \( f \) vanishes outside of a set of \( \sigma \)-finite measure, namely, \( \bigcup_k \Omega_k \) with \( \Omega_k \subset \Omega_{k+1} \) and \( 0 < \mu(\Omega_k) < \infty \). Thus, for any \( \varepsilon \) in the interval \( (0, \|f\|_\infty) \), the set \( \Omega_{k,\varepsilon} = \{x \in \Omega_k : |f(x)| \geq \|f\|_\infty - \varepsilon\} \) must have a positive measure for \( k \) sufficiently large. Therefore, define \( g = \text{sign}(f)\mathbb{1}_{\Omega_{k,\varepsilon}} \) to have

\[
\langle f_n, g \rangle \to \langle f, g \rangle \geq (\|f\|_\infty - \varepsilon)\mu(\Omega_{k,\varepsilon}).
\]
Again, Hölder inequality yields
\[ |\langle f_n, g \rangle| \leq \|f_n\|_p \mu(\Omega_{k,e}), \quad \text{with} \quad 0 < \mu(\Omega_{k,e}) < \infty, \]
and, we deduce
\[ \liminf_n \|f_n\|_\infty \geq \|f\|_\infty - \varepsilon, \]
i.e., (2.2).

\[ \Box \]

- **Remark 2.34.** Related to Remark 1.36 we have the following result: if \( f_n \rightharpoonup f \) weakly in \( L^p(\Omega,F,\mu) \) with \( 1 < p < \infty \) and \( \|f_n\|_p \to \|f\|_p \) then \( \|f_n - f\|_p \to 0 \) as \( n \to \infty \). This assertion fails for \( p = 1 \) or \( p = \infty \), e.g., see DiBenedetto [31, Section V.11, pp. 236–238].

It is clear that Banach-Steinhaus Theorem 2.17 (or uniformly boundedness principle) proves that any weakly convergence sequence is bounded. The converse (i.e., that a bounded sequence contains a convergent subsequence) holds true for any dual space of Banach space (Alaoglu’s Theorem 2.7, e.g., see Conway [29, Section V.3 and V.4, pp. 123–137]). The proof is similar to the one given below, valid for any separable reflexive space (recall that this fails for \( L^1 \)). On the other hand, a so-called density argument shows that if a bounded sequence \( \{x_n\} \) in a normed space \( X \) (say \( L^p \)) is such that \( f(x_n) \to f(x) \) for every \( f \) in a dense set \( D \) of the dual space \( X' \) then \( x_n \rightharpoonup x \), weakly in \( X \).

- **Remark 2.35.** Consider the space \( \ell^p \) of all sequence \( x = \{x_k\} \) with finite sum
\[ \|x\|_p = \sum p \leq \infty, \quad \text{i.e., the space } \ell^p \text{ with the discrete measure } \mu(A) = \sum_{1 \in A}, A \subset \mathbb{N}. \]
On this space \( \ell_p \) with \( 1 < p < \infty \), the weak convergence can be characterized as follows: a sequence \( x^{(n)} \) converges weakly to \( x \) if and only if (a) it is bounded, i.e., there exists a constant \( C \) such that \( \|x^{(n)}\|_p \leq C \) for every \( n \) and (b) each coordinate converges, i.e., for every \( k \), \( x^{(n)}_k \to x_k \) as \( n \to \infty \), e.g., see Bachman and Narici [14, Section 14.1, pp. 231–238].

A \( \sigma \)-algebra \( F \) is called \( \mu \)-separable if the algebra \( F_0 = \{ F \in F : \mu(F) < \infty \} \) is the completion of a countable generated algebra, i.e., there exists a countable subset \( Q \) of \( F_0 \) such that for any set \( F \) in \( F_0 \) there is a sequence \( \{F_n\} \) in \( Q \) satisfying \( \mu((F \setminus F_n) \cup (F_n \setminus F)) \to 0 \). In this case, the space \( L^q(\Omega,F,\mu) \) is separable for any \( 1 \leq q < \infty \). Certainly, this includes the case where \( \Omega \) is a Polish space (i.e., a separable and complete metrizable space) and \( \mu \) is a \( \sigma \)-finite regular Borel measure.

**Proposition 2.36.** Let \( F \) be \( \mu \)-separable and \( \{f_n : n \geq 1\} \) be a bounded sequence in \( L^p(\Omega,F,\mu) \) with \( 1 < p \leq \infty \). Then there exists a weakly convergent subsequence \( \{f_{n_k} : k \geq 1\} \).

**Proof.** Essentially, this is the Cantor diagonal argument. The conjugate of the exponent \( p \) is \( q \), \( 1/p + 1/q = 1 \), with \( 1 \leq q < \infty \). Thus, let \( \{g_i : i \geq 1\} \) be a dense sequence in \( L^q(\Omega,F,\mu) \). Since the numerical sequence \( \{(f_n,g_i) : n \geq 1\} \) is bounded for each \( i \), by means of Cantor diagonal procedure, we construct a
subsequence \( \{ f_{n_k} : k \geq 1 \} \) such that the numerical sequence \( \{ \langle f_{n_k}, g_i \rangle : n \geq 1 \} \) is convergent for every \( i \geq 1 \).

Since \( \{ g_i : i \geq 1 \} \) is dense, for every \( g \) in \( L^q(\Omega, \mathcal{F}, \mu) \) and for any \( \varepsilon > 0 \) there exists \( g_i \) such that \( \| g - g_i \| < \varepsilon \). Hence the inequality

\[
|\langle f_{n_k} - f_{n_h}, g \rangle| \leq |\langle f_{n_k} - f_{n_h}, g_i \rangle| + \| g - g_i \| \sup_n \{ \| f_n \| \}
\]

shows that the numerical sequence \( \{ \langle f_{n_k}, g \rangle : n \geq 1 \} \) converges for every \( g \) and defines a linear functional on \( L^p(\Omega, \mathcal{F}, \mu) \). By Riesz representation Theorem B.63, there exists \( f \) in \( L^p(\Omega, \mathcal{F}, \mu) \) such that \( f_{n_k} \rightharpoonup f \) weakly.

**Remark 2.37.** Referring to Vitali-Hahn-Saks Theorem 2.30, we can prove that a sequence \( \{ f_n : n \geq 1 \} \) in \( L^1(\Omega, \mathcal{F}, \mu) \) is weakly compact if and only if it is bounded and the integrals

\[
I_n(A) = \int_A f_n d\mu, \quad n \geq 1,
\]

are uniformly \( \sigma \)-additive, e.g., see Dunford and Schwartz[39, Theorem IV.8.9, pp. 292–296]. Certainly, if \( \{ f_n : n \geq 1 \} \) is \( \mu \)-uniformly integrable (see Definition 1.28) then \( \{ I_n : n \geq 1 \} \) is uniformly \( \sigma \)-additive, see Proposition 1.29. Hence, form Corollary 2.31 we deduce that sequentially weakly compact in \( L^1 \) is equivalent to \( \mu \)-uniformly integrable, i.e., the Dunford-Pettis criterium.

### 2.4.2 Totally Bounded Sets

Recall that a subset \( \{ f_i : i \in I \} \) of a metric space \( (X, d) \) is totally bounded if for every \( \varepsilon > 0 \) there exists a finite subset of indexes \( J \subset I \) such that for every \( i \) in \( I \) there exists \( j \) in \( J \) satisfying \( d(f_i, f_j) < \varepsilon \), i.e., any element in \( \{ f_i : i \in I \} \) is within a distance \( \varepsilon \) from the finite set \( \{ f_j : j \in J \} \). Sometimes \( \{ f_j : j \in J \} \) is called an \( \varepsilon \)-net relative to \( \{ f_i : i \in I \} \). It is clear that a Cauchy sequence is a totally bounded set, and conversely, any totally bounded set contains a Cauchy sequence. Indeed, based on the existence of \( \varepsilon \)-nets, we can construct (by induction) a sequence \( \{ f_n : n \geq 1 \} \) (of the given totally bounded set) such that \( d(f_{n-1}, f_n) < 2^{-n} \), for any \( n \geq 2 \), which is a Cauchy sequence. In a metric space, compactness is equivalent to sequentially compactness, and then, a totally bounded sets is equivalent to pre-compact (i.e., closure compact) set on a complete metric space, in particular, this also applied to the \( F \)-spaces \( L^p(\Omega, \mathcal{F}, \mu) \) with \( 0 < p < 1 \) and the distance \( d(f,g) = \| f - g \|_p \) and to the Banach spaces \( L^p(\Omega, \mathcal{F}, \mu) \) with \( 1 \leq p \leq \infty \).

The following characterization of pre-compact (or totally bounded) sets in \( L^p(\Omega) \) is sometime referred to as Fréchet-Kolmogorov Theorem, e.g., Yosida [135, Section X.1, pp. 274–277] and DiBenedetto[31, Section V.22, pp. 260–262]. This applies to \( L^p(\Omega) = L^p(\Omega, \mathcal{F}, \mu) \), where \( \Omega \) an open subset of \( \mathbb{R}^d \) and \( \mu \) is
the Lebesgue measure. We use the notation
\[ \tau_h f = f(\cdot + h), \quad \forall h \in \mathbb{R}^d, \quad h\text{-translations}, \]
\[ \|f\|_{p,A} = \left( \int_A |f(x)|^p \, dx \right)^{1/p}, \quad \forall A \subset \mathbb{R}^d, \text{ measurable,} \]
\[ \Omega_\delta = \{ x \in \Omega : d(x, \partial \Omega) > \delta, \, |x| < 1/\delta \}, \quad \forall \delta > 0, \]
where \( d(x, \partial \Omega) = \inf\{|x - y| : y \in \partial \Omega|\} \) is the distance from the point \( x \) to the boundary \( \partial \Omega \) of \( \Omega \). Also \( \overline{\Omega}_\delta \) denotes the closure of the open set \( \Omega_\delta \).

**Theorem 2.38.** A uniformly bounded \( \{ f_i : i \in I \} \) subset of functions in \( L^p(\Omega) \), with \( 1 \leq p < \infty \) and \( \Omega \) an open subset of \( \mathbb{R}^d \), is pre-compact or totally bounded if and only if for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that
\[ \|\tau_h f_i - f_i\|_{p, \Omega_\delta} \leq \varepsilon \quad \text{and} \quad \| f_i \|_{p, \Omega \setminus \Omega_\delta} \leq \varepsilon, \quad \forall i \in I, \quad (2.3) \]
and for every translation \( \tau_h \) with \( |h| \leq \delta \).

**Proof.** First note that \( \Omega = \bigcup_n \Omega_{1/n} \) and that \( \Omega_\delta \) is an open set with a finite (Lebesgue) measure. Also, the function \( x \mapsto (\tau_h f)(x) \) is almost everywhere defined for \( x \in \Omega_\delta \) and for any \( |h| \leq \delta \). Certainly, the condition that \( \{ f_i : i \in I \} \) is uniformly bounded in \( L^p(\Omega) \) means that there exists a constant \( C > 0 \) such that \( \| f_i \|_p \leq C \), for every \( i \) in \( I \).

Let \( k \) be a smooth kernel, i.e., \( k \) is a smooth nonnegative function with support inside the closed unit ball and integral 1, so that \( k_\eta(x) = \eta^{-d}k(x/\eta) \), \( \eta > 0 \), is called an \( \eta \)-mollifying kernel. By means of the convolution, for any \( \eta \leq \delta \) and almost every \( x \) in \( \Omega_\delta \) we have
\[ |(f * k_\eta)(x) - f(x)| \leq \int_{|y| \leq \eta} |(\tau_y f)(x) - f(x)| \, dy, \]
which yields the estimate
\[ \| f * k_\eta - f \|_{p, \Omega_\delta} \leq \sup_{|h| \leq \eta} \| \tau_h f - f \|_{p, \Omega_\delta}, \quad \forall \eta \leq \delta, \quad (2.4) \]
and similarly,
\[ |(f * k_\eta)(x)| \leq \| k_\eta \|_q \| f \|_{p, \Omega_\delta}, \quad \forall x \in \overline{\Omega}_{2\delta}, \quad \forall \eta < \delta, \quad (2.5) \]
for any \( 1 \leq p \leq \infty \), with \( 1/p + 1/q = 1 \). Note that \( \| k_\eta \|_q \) is unbounded as \( \eta \to 0 \), for any \( q > 1 \) or equivalently, \( 1 \leq p < \infty \).

Also, recall the continuity of the translations in \( L^1(\mathbb{R}^d) \), i.e., the translation operator \( \tau_a f = f(\cdot - a) \) is continuous in \( \mathcal{L}^1 \), which is easily extended to \( L^p(\mathbb{R}^d) \). Therefore, an extension by zero of functions in \( L^p(\Omega) \) shows that
\[ \lim_{h \to 0} \| \tau_h f - f \|_{p, \Omega_\delta} = 0, \quad \forall f \in L^p(\Omega) \quad (2.6) \]
and \( 1 \leq p < \infty \).
Suppose \( \{f_i : i \in I\} \) is a uniformly bounded set in \( L^p \) satisfying (2.3). Then for any \( \varepsilon' > 0 \) we will construct an \( \varepsilon' \)-net, proving that \( \{f_i : i \in I\} \) is totally bounded. Indeed, for a fixed small \( \eta > 0 \), consider the family of functions \( \{f_i \star k_\eta : i \in I\} \), as defined on \( \overline{\Omega}_{2\eta} \), and apply estimate (2.5) to \( f_i, f_\tau, f_i - f_\tau \) to obtain, for every \( x \in \overline{\Omega}_{2\eta} \) the inequalities
\[
| (f_i \star k_\eta)(x) | \leq \|k_\eta\|_q \|f_i\|_{p, \Omega_\eta},
\]
\[
| (f_\tau f_i \star k_\eta)(x) - (f_i \star k_\eta)(x) | \leq \|k_\eta\|_q \|f_\tau f_i - f_i\|_{p, \Omega_\eta} \quad \forall |h| < \eta.
\]
Together with condition (2.3), this shows that \( \{f_i \star k_\eta : i \in I\} \) is uniformly bounded and equi-continuous set of continuous functions on \( \overline{\Omega}_{2\eta} \). Hence, Ascoli Theorem 2.9 implies that \( \{f_i \star k_\eta : i \in I\} \) is pre-compact, and therefore, there exists an \( \varepsilon'' \)-net of continuous functions defined on \( \overline{\Omega}_{2\eta} \), which is denoted by \( \{g_j : j \in J\} \), with \( g_j = f_j \star k_\eta \) and \( J \) a finite subset of indexes. Thus, estimate 2.4 yields the inequality
\[
\|f_i - f_j\|_p \leq \|f_i - f_j\|_{p, \Omega \setminus \Omega_\delta} + \|f_i - g_j\|_{p, \Omega_\delta} + \|f_j \star k_\eta - f_j\|_{p, \Omega_\delta} \leq 2 \sup_i \|f_i\|_{p, \Omega \setminus \Omega_\delta} + \|f_i - g_j\|_{p, \Omega_{2\eta}} + \sup_{|h| \leq \eta} \|f_\tau f_j - f_j\|_{p, \Omega_\delta},
\]
with \( 0 < \eta \leq \delta/2 \). Hence, by means of condition (2.3), we deduce that \( \{f_j : j \in J\} \) is an \( \varepsilon' \)-net for \( \{f_i : i \in I\} \) with
\[
\min_{j \in J} \|f_i - f_j\|_{p, \Omega} \leq \varepsilon', \quad \forall i \in I.
\]
This shows that \( \{f_i : i \in I\} \) is uniformly bounded and that for every \( i \in I \) there exists \( j \in J \) such that
\[
\|f_i\|_{p, \Omega \setminus \Omega_\delta} \leq \|f_j\|_{p, \Omega \setminus \Omega_\delta} + \varepsilon', \quad \forall \delta > 0
\]
and
\[
\|f_i - f_j\|_{p, \Omega_\delta} \leq \|f_i - f_j\|_{p, \Omega_\delta} + \|f_i - f_j\|_{p, \Omega_\delta} + \|f_\tau f_j - f_j\|_{p, \Omega_\delta} + \|f_j - f_i\|_{p, \Omega_\delta} \leq 2\varepsilon' + \|f_\tau f_j - f_j\|_{p, \Omega_\delta}, \quad \forall \delta > 0.
\]
Because \( J \) is finite and the translations are continuous, see property (2.6), there is \( \delta > 0 \) such that
\[
\max_{j \in J} \|f_j\|_{p, \Omega \setminus \Omega_\delta} \leq \varepsilon' \quad \text{and} \quad \max_{j \in J} \|f_\tau f_j - f_j\|_{p, \Omega_\delta} \leq \varepsilon', \quad \forall |h| \leq \delta.
\]
Hence \( \|f_i\|_{p, \Omega \setminus \Omega_\delta} \leq 2\varepsilon' \) and \( \|f_\tau f_i - f_i\|_{p, \Omega_\delta} \leq 3\varepsilon' \), for every \( i \in I \). This shows condition (2.3) with \( \varepsilon = 3\varepsilon' \). \( \square \)

Since any function in \( L^p(\Omega) \) can be extended by zero to a function in \( L^p(\mathbb{R}^d) \), we can rephrase the previous results in \( L^p(\mathbb{R}^d) \) as follows:
Proposition 2.39. A family \( \{ f_i : i \in I \} \) of functions in \( L^p(\mathbb{R}^d) \), with \( 1 \leq p < \infty \), is totally bounded or pre-compact if and only if (1) there exists a constant \( C > 0 \) such that \( \| f_i \|_p \leq C \), for every \( i \in I \); (2) for every \( \varepsilon > 0 \) there exists \( n \) such that \( \| 1_{|x| \geq r} f_i(x) \|_p \leq \varepsilon \) for every \( r > n \) and for every \( i \in I \); and (3) for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( \| \tau_h f_i - f_i \|_p < \delta \) for every \( |h| < \delta \) and for every \( i \in I \).

Proof. Since a totally bounded set is necessarily bounded, this result is a consequence of the previous Theorem 2.38 for \( \Omega = \mathbb{R}^d \). However, it is worthwhile to remark some arguments used.

For instance, to verify (2) we get first an \( \varepsilon/2 \)-net \( \{ f_j : j \in J \} \) with \( J \) a finite subset of indexes of \( I \). Because \( |f_j|^p \) is integrable, the dominate convergence Lebesgue theorem shows that \( 1_{|x| \leq r} f_j(x) \to f_j(x) \) in \( L^p \) for each \( j \) and so, for every \( \varepsilon > 0 \) there exists \( r > n \) such that \( \| 1_{|x| \geq r} f_j(x) \|_p < \varepsilon/2 \) for every \( j \) in \( J \). However, for each \( i \in I \) there exists \( j \) such that \( \| f_i - f_j \|_p < \varepsilon/2 \), and we conclude.

To verify (3), we can compute \( \| \tau_h f - f \|_p \) to show that \( \| \tau_h f - f \|_p \to 0 \) as \( h \to 0 \) for every \( f = 1_A \) where \( A \) is a \( d \)-interval in \( \mathbb{R}^d \). Next, by linearity, this remains true for any finite valued function \( f \) and finally, but density, this holds true for any function \( f \) in \( L^p \).

\[ \blacksquare \]

- Remark 2.40. It is clear that (2.3) of Theorem 2.38 or Proposition 2.39 can be restated as follows: for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that

\[
\| \tau_{r,k} f_i - f_i \|_{p,\Omega_\delta} \leq \varepsilon \quad \text{and} \quad \| f_i \|_{p,\Omega_\delta \setminus \Omega_\delta} \leq \varepsilon, \quad \forall i \in I, \tag{2.7}
\]

and for every one-dimensional translation \( \tau_{r,k} \) with \( |r| \leq \delta \), \( k = 1, \ldots, d \), i.e., with

\[
(\tau_{r,k} f)(x_1, \ldots, x_k, \ldots, x_d) = f(x_1, \ldots, x_k + r, \ldots, x_d).
\]

Indeed, it suffices to note that for any translation \( \tau_h \) with \( h = (h_1, \ldots, h_d) \) we have

\[
\tau_h f - f = (\tau_{h_1,1} f_1 - f_1) + (\tau_{h_2,2} f_2 - f_2) + \cdots + (\tau_{h_d,d} f_d - f_d),
\]

with \( f_1 = f \) and \( f_k = \tau_{h_k,k} f_{k-1} \), for \( k = 2, \ldots, d \).

\[ \blacksquare \]

The interested reader may check the textbook by Stein and Shakarchi [116] for further topics.

Exercise 2.5. If \( A \) is a totally bounded set of a normed space \( (X, \| \cdot \|) \) then prove that the convex hull (or convex envelope) \( \text{co}(A) \) of \( A \) (i.e., the smallest convex set containing \( A \)) is also totally bounded. In particular, the closed convex hull of a compact set of a Banach space is also compact. Hint: Use the following argument (1) if \( F \subset X \) is a finite set then the convex hull \( \text{co}(F) \) of \( F \) is a totally bounded set. Next, let \( A \) be a totally bounded subset of \( X \) and let \( B_1 \) be an open balls containing the origin. By using the previous result, (2) find a finite set \( F \) such that \( A \subset F + B_1 \) and deduce that \( \text{co}(A) \) lies inside \( K + B_1 \) for some

[Preliminary]  
Menaldi  
November 11, 2016
totally bounded set $K$. Now, take any two open balls $B_1$ and $B$ containing the origin and satisfying $B_1 + B_1 \subset B$. Finally, because $K$ is totally bounded, (3) find another finite $E$ such that \( \text{co}(A) \subset (E + B_1) + B_1 \subset E + B \), and deduce that \( \text{co}(A) \) is indeed totally bounded.

\[ \square \]

**Exercise 2.6.** Banach-Saks Theorem states that if \( \{f_n\} \) is a weakly convergence sequence to $f$ in $L^p(\Omega, \mathcal{F}, \mu)$, $1 \leq p < \infty$ then there exists a subsequence \( \{f_{n_k}\} \) such that the arithmetic means \( g_k = (f_{n_1} + \cdots + f_{n_k})/k \) strongly converges to $f$, i.e., \( \|g_k - f\|_p \to 0 \). Prove this result for a Hilbert space $H$ with scalar product $(\cdot, \cdot)$ and norm $\| \cdot \|$, in particular for $p = 2$. \textbf{Hint:} First reduce the problem to the case where $f = 0$, and $\|f_n\| \leq 1$ for every $n \geq 1$. Next, construct a subsequence satisfying $|(f_{n_1}, f_{n_k})| \leq 1/k$, for every $i = 1, \ldots, k$, and deduce that $\|g_k\|^2 \leq 3/k$, see Riesz and Nagy [107, Section 38, pp. 80–81].

\[ \square \]

## 2.5 Basic Interpolation Questions

Interpolation is a useful technique as detailed in comprehensive books, e.g., Bergh and L"ofstr"om [18] and Triebel [127, 129, 130]. A very introductory presentation along the lines in Grafakos [57] is given below.

As it was mentioned early, the $L^p$-spaces on a measure space $(X, \mathcal{X}, \mu)$ with $1 \leq p \leq \infty$ are of great importance, and depending on the context, a common notation is $L^p$ or $L^p(\mu)$ or $L^p(X, \mu)$ or $L^p(X, \mathcal{X}, \mu)$ for real-valued (or complex-valued) function. Another common spaces are $L^p \cap L^q$ (or $L^p(\mu) \cap L^q(\nu)$ if $\mu \neq \nu$) and $L^p + L^q$ (or $L^p(\mu) + L^q(\nu)$ if $\mu \neq \nu$) with suitable norms. The expression

\[
(f, g) = \int_X f(x)g(x)\mu(dx)
\]

is the inner or scalar product in the Hilbert space $L^2$, while

\[
\|f\|_p = \left( \int_X |f(x)|^p \mu(dx) \right)^{1/p}
\]

is the norm in the Banach space $L^p$, for any $1 \leq p \leq \infty$. These norms have a dual representation as

\[
\|f\|_p = \sup_{\|g\|_{p'} = 1} \left| \int_X f(x)g(x)\mu(dx) \right|, \quad 1/p + 1/p' = 1,
\]

and the dual space of $L^p$ is $L^{p'}$ for $1 \leq p < \infty$. Actually, it is clear that $g$ with $\|g\|_p = 1$ in a dense subset of $L^{p'}$ (e.g., the simple functions) in the supremum are sufficient to reproduce the

The case $0 < p < 1$ can also be studied as topological vector spaces in view of the inequality

\[
\|f + g\|_p \leq 2^{(1-p)/p}(\|f\|_p + \|g\|_p), \quad 0 < p < 1,
\]

which make $\| \cdot \|_p$ a quasi-norm.

As remarked early, instead of a direct integration of a measurable function $f : X \to [-\infty, +\infty]$, its \textit{distribution function}

$$
\mu(f, \cdot) : [0, \infty) \to [0, \infty], \quad \mu(f, \lambda) = \mu(\{x \in X : |f(x)| > \lambda\}) \tag{2.9}
$$
is used, namely,

$$
\int_X |f|^p d\mu = p \int_0^\infty \lambda^{p-1} \mu(f, \lambda) d\lambda, \quad 0 < p < \infty. \tag{2.10}
$$

This equality follows by writing

$$
\mu(f, \lambda) = \int_X 1_{\{x \in X : |f(x)| > \lambda\}} \mu(dx)
$$

and exchanging the order of integration. 

Suggested by the inequality

$$
\lambda^p \mu(f, \lambda) \leq \int_{\{x \in X : |f(x)| > \lambda\}} |f(x)|^p d\mu \leq \|f\|_p^p, \quad \forall \lambda > 0, \ f \in L^p,
$$

weak-version of $L^p$-spaces is defined based on the quasi-norm

$$
\|f\|_p = \sup_{\lambda > 0} \{\lambda(\mu(f, \lambda))^{1/p}\}, \quad 0 < p < \infty. \tag{2.11}
$$

Thus, a measurable function belongs to weak-$L^p$ if and only if $\|f\|_p < \infty$. Typically, the function $x \mapsto |x|^{-d/p}$ belongs to weak-$L^p(\mathbb{R}^d, \ell)$, with the Lebesgue measure $\ell$, but it does not belong to the space $L^p(\mathbb{R}^d, \ell)$, i.e., the inclusion of weak-$L^p$ into $L^p$ is usually strict. In general, quasi-norm $\| \cdot \|_p$ in the spaces weak-$L^p$ is equivalent to

$$
\|f\|_p = \sup_{0 < \mu(A) < \infty} (\mu(A))^{1/p-1/r} \left( \int_A |f(x)|^r \mu(dx) \right)^{1/r},
$$

with any $0 < r < p < \infty$. Indeed, use the inequality

$$
\mu(\{x \in A : |f(x)| > \lambda\}) \leq \min \{\mu(A), \lambda^{-p}\|f\|_p^p\}
$$

to deduce

$$
\int_A |f(x)|^r \mu(dx) \leq \frac{p}{p-r} \left( \mu(A) \right)^{1-r/p} \|f\|_p^p,
$$

which yields

$$
(1 - r/p)^{1/r} \|f\|_p \leq \|f\|_p \leq \|f\|_p, \quad \forall f \text{ in weak-}L^p. \tag{2.12}
$$

Hence, if $1 < p < \infty$ then taking $r = 1$ the expression $\|f\|_p$ is a norm, i.e., weak-$L^p$ is also a Banach space for $1 < p < \infty$, while it is complete metric topological vector space for $0 < p \leq 1$. 

No weak types are defined for the limiting cases \( L^0 = L^0(X, \mu) \) (equivalence classes of finite-valued \( \mu \)-almost everywhere functions, under the \( \mu \)-almost everywhere equality) and \( L^\infty = L^\infty(X, \mu) \) (essentially bounded elements in \( L^0 \)). Also recall that if the measure is not \( \sigma \)-finite then another condition is added to these spaces, namely, the support is \( \sigma \)-finite. To make this consistent with the weak-space defined above, this \( \sigma \)-finite support condition is replaced with the assumption that its distribution function is finite, i.e., \( \mu(f, \lambda) < \infty \) for every \( \lambda > 0 \). Thus, the vector space of all simple functions (i.e., equivalence classed under the \( \mu \)-almost everywhere equality of integrable functions with only a finite number of values) is dense in any \( L^p \) with \( 0 < p \leq \infty \), since any nonnegative measurable function with a \( \sigma \)-finite support is an increasing pointwise limit of a sequence of nonnegative integrable simple functions (where the convergence is uniformly if \( f \) is bounded).

Therefore, by convenience, denote by \( S(X, \mu) \) the vector space of all integrable simple (equivalence classes of) functions to obtain the dense inclusion \( S(X, \mu) \subset L^p(X, \mu) \) with \( 0 \leq p \leq \infty \), provided \( L^0 \) is endowed with the convergence in measure over every set of finite measure (which makes \( L^0 \) a topological vector space, even if \( L^0 \) is mainly used as a vector space). All these vector spaces may be considered on the complex or the real field, with specific distinction made when necessary. Note that a convenient form of expressing a complex-valued simple integrable function is the following \( f = \sum_{k=1}^{n} r_k e^{i\theta} \mathbb{1}_{A_k} \), where \( \{A_k\} \) is a sequence of disjoint measurable sets of finite \( \mu \)-measure and any complex value is represented in its polar form with \( r > 0 \) and \( 0 \leq \theta < 2\pi \).

If the measure \( \mu \) on \( X \) is infinite then it may be convenient to define local versions of \( L^p \) (and weak-\( L^p \)) by assuming local integrability (or \( \sigma \)-integrability, in the sense of integrable on every set of finite measure). For instance, the Lebesgue space \( L^p(\mathbb{R}^d, \ell) \) get localized to \( L^p_{\text{loc}}(\mathbb{R}^d, \ell) \), i.e., the space of locally \( p \)-integrable function

\[
\int_K |f(x)|^p \ell(dx) < \infty, \quad \forall \text{ compact } K \subset \mathbb{R}^d.
\]

It is clear that

\[
L^p \subset L^p_{\text{loc}} \subset L^q_{\text{loc}}, \quad p > q
\]

and that \( L^p_{\text{loc}} \) is a metric topological vector space. In any way, our main interest is the case \( 1 \leq p \leq \infty \), moreover, typically, \( 1 < p < \infty \).

There are also the so-called Lorentz spaces, which are defined by means of the decreasing rearrangement \( f^* \) of a measurable function \( f \), namely

\[
f^*(t) = \inf\{\lambda > 0 : \mu(f, \lambda) \leq t\},
\]

with the convention that \( f^*(t) = +\infty \) if \( \mu(f, \lambda) < t \), for all \( \lambda > 0 \). This function \( f^* \) is decreasing with support in \([0, \mu(X)]\). For instance, the interested reader may check the book by Grafakos [57, Section 1.4, pp. 44–63].

The notation (2.8) and (2.11) is used in most of this section.
2.5.1 Preliminary Interpolation

Begin with the following result

**Theorem 2.41.** If $0 < p < q \leq \infty$ then

$$
\|f\|_r \leq \|f\|_p^{1-\theta} \|f\|_q^\theta \quad \text{and} \quad \|f\|_r \leq \|f\|_p^{1-\theta} \|f\|_q^\theta,
$$

and moreover,

$$
\|f\|_r \leq \left( \frac{r}{r-p} + \frac{r}{q-r} \right)^{1/r} \|f\|_p^{1-\theta} \|f\|_q^\theta.
$$

for any choice $p < r < q$, with any some $\theta$ in $(0, 1)$ and $1/r = (1 - \theta)/p + \theta/q$. If $q = \infty$ then $1/r = (1 - \theta)/p$, $r/(q - r) = 0$ and $\|f\|_\infty = \|f\|_\infty$.

**Proof.** First use Hölder inequality with exponents $\bar{p} = p/[r(1 - \theta)]$ and $\bar{q} = q/(r\theta)$ to get

$$
\|f\|_r^r = \int_X |f(x)|^{r(1-\theta)}|f(x)|^r \theta \mu(dx) \leq \left( \int_X |f(x)|^p \mu(dx) \right)^{1/\bar{p}} \left( \int_X |f(x)|^q \mu(dx) \right)^{1/\bar{q}} = \|f\|_p^{r(1-\theta)} \|f\|_q^r,
$$

and write the equality

$$
\lambda(\mu(f, \lambda))^{1/r} = \lambda(\mu(f, \lambda)^{(1-\theta)/p} \lambda(\mu(f, \lambda))^{\theta/q} \leq \|f\|_p^{1-\theta} \|f\|_q^\theta.
$$

Hence, the first interpolation inequalities follow.

Next, by means of the density inequalities, first note that for $q < \infty$,

$$
\mu(f, \lambda) \leq \min \{ \lambda^{-p} \|f\|_p^p, \lambda^{-q} \|f\|_q^q \}, \quad \forall \lambda > 0,
$$

and thus,

$$
\|f\|_r^r = r \int_0^\infty \lambda^{r-1} \mu(f, \lambda) d\lambda \leq r \int_0^\infty \lambda^{r-1} \min \{ \lambda^{-p} \|f\|_p^p, \lambda^{-q} \|f\|_q^q \} d\lambda,
$$

If $a = (\|f\|_q^q/\|f\|_p^p)^{1/(q-p)}$ then split the integral into the intervals $(0, a]$ and $[a, \infty)$ to deduce

$$
\|f\|_r^r \leq r \int_0^a \lambda^{r-1-p} \|f\|_p^p d\lambda + r \int_a^\infty \lambda^{r-1-q} \|f\|_q^q d\lambda = \frac{r}{r-p} \|f\|_p^p a^{r-p} + \frac{r}{q-r} \|f\|_q^q a^{r-q} = \left( \frac{r}{r-p} + \frac{r}{q-r} \right) \|f\|_p^p(q-r)/(q-p)(\|f\|_q^q)^{(r-p)/(q-p)},
$$

and since

$$
\frac{1}{p} - \frac{1}{r} = \theta \left( \frac{1}{q} - \frac{1}{p} \right), \quad \frac{p(q-r)}{r(q-p)} = 1 - \theta, \quad \frac{q(r-p)}{r(q-p)} = \theta,
$$
we deduce the second interpolation inequality for $q < \infty$.

Finally, if $q = \infty$ then $\mu(f, \lambda) = 0$ for $\lambda > \|f\|_{\infty}$ and

$$
\mu(f, \lambda) \leq \min \{ \lambda^{-p} \|f\|_{p}^{p} \}, \quad \forall \lambda \in (0, \|f\|_{\infty}].
$$

Hence, estimate the integral as above to obtain

$$
\|f\|_{r} \leq \frac{r}{r - p} \|f\|_{p}^{1 - \theta} \|f\|_{\infty}^{\theta}
$$
as desired. \hfill \Box

As a simple application, note that if $T$ is a linear operator from $L^{p}$ into $L^{\bar{p}}$ and also from $L^{q}$ into $L^{\bar{q}}$ with $1 < p < q \leq \infty$ and $1 < \bar{p} < \bar{q} \leq \infty$ then apply Theorem 2.41 to obtain the inequality

$$
\|Tf\|_{\bar{r}} \leq \|Tf\|_{\bar{p}}^{1 - \theta} \|Tf\|_{\bar{q}}^{\theta} \leq \left( \|T\|_{p, \bar{p}} \|f\|_{p} \right)^{1 - \theta} \left( \|T\|_{q, \bar{q}} \|f\|_{q} \right)^{\theta} \leq \|T\|_{p, \bar{p}} \|T\|_{q, \bar{q}} \max\{\|f\|_{p}, \|f\|_{q}\},
$$

where $\|T\|_{p, \bar{p}}$ and $\|T\|_{q, \bar{q}}$ are the operator norms, e.g.,

$$
\|T\|_{p, \bar{p}} = \sup \{\|Tf\|_{p} : \|f\|_{p} \leq 1\}.
$$

This proves that if $T$ is also bounded (in the corresponding spaces) then $T$ is bounded from $L^{p} \cap L^{q}$ with the norm $\max\{\|f\|_{p}, \|f\|_{q}\}$ into $L^{\bar{r}}$ with $p < \bar{r} < q$. Certainly, a similar results holds for the weak-$L^{p}$ spaces. However, more interesting is the results presented below.

### 2.5.2 Marcinkiewicz Interpolation Theorem

This argument uses the so-called real method. First some terminology. Recall that $S(X, \mu)$ denotes the subspace of all (equivalence classes of) simple $\mu$-integrable functions defined on $X$ and that $L^{0}(Y, \nu)$ denotes the space of all (equivalence classes of) $\nu$-measurable (real-valued, for simplicity) $\sigma$-finite supported functions defined on $Y$.

**Definition 2.42 (type $(p, q)$).** An operator $T$ initially defined in a dense subset of $L^{p}(X, \mu)$ (e.g., $S(X, \mu)$, the subspace of all simple functions) and with values in $L^{0}(Y, \nu)$ (i.e., the space of measurable functions) is called of strong-type $(p, \bar{p})$ if there is a constant $C_{p, \bar{p}}$ such that

$$
\|Tf\|_{\bar{p}} \leq C_{p, \bar{p}} \|f\|_{p}, \quad \forall f.
$$

Similarly, $T$ is called of weak-type $(p, \bar{p})$ if the norm $\|Tf\|_{\bar{p}}$ is replaced by the quasi-norm $(\|Tf\|_{\bar{p}})$ given by (2.11) (or the equivalent norm $(\|f\|_{\bar{p}})$, when $0 < \bar{p} < \infty$). If $\bar{p} = \infty$ then weak-type means strong-type. \hfill \Box

In view of the density of the domain, it is clear that an operator of strong or weak $(p, \bar{p})$ can be considered defined in the whole space $L^{p}(X, \mu)$ or weak-$L^{p}(X, \mu)$.
**Theorem 2.43.** Let $T$ be a linear operator defined on $S(X, \mu)$ with values in $L^0(Y, \nu)$ of weak-type $(p, p)$ and weak-type $(q, q)$, with $0 < p < q \leq \infty$, i.e.,

$$\|Tf\|_p \leq C_p \|f\|_p \quad \text{and} \quad \|Tf\|_q \leq C_q \|f\|_q,$$

with some constants $C_p$ and $C_q$ independent of $f$. Then for any $p < r < q$,

$$\|Tf\|_r \leq C_r \|f\|_r, \quad \forall f \in S(X, \mu), \quad C_r = 2 \left( \frac{r}{r-p} + \frac{r}{q-r} \right)^{1/r} C_p^{1-\theta} C_q^\theta,$$

with $\theta$ in $(0, 1)$ and $1/r = (1-\theta)/p + \theta/q$. If $q = \infty$ then $1/r = (1-\theta)/p$, $r/(q-r) = 0$ and $\|Tf\|_\infty = \|Tf\|_\infty$.

**Proof.** Take a number $a > 0$ to be determined later, and choose a simple function or element $f$ in $S(X, \mu)$ to consider the functions $g = f 1_{\{|f| > a\lambda\}}$ and $h = f 1_{\{|f| \leq a\lambda\}}$. Both, $g$ and $h$ belong to $S(X, \mu)$ and $f = g + h$. The assumption of the weak-types of $T$ implies

$$\|Tg\|_p^p \leq C_p^p \|g\|_p^p = \int_{\{x : |f(x)| > a\lambda\}} |f(x)|^p \mu(dx),$$

$$\|Th\|_q^q \leq C_q^q \|h\|_q^q = \int_{\{x : |f(x)| \leq a\lambda\}} |f(x)|^q \mu(dx).$$

While, the linearity of $T$ yields $Tf = Tg + Th$, which implies

$$\{x : |Tf(x)| > \lambda\} \subset \{x : |Tg(x)| > \lambda/2\} \bigcup \{x : |Th(x)| > \lambda/2\},$$

and therefore

$$\mu(Tf, \lambda) \leq \mu(Tg, \lambda/2) + \mu(Th, \lambda/2) \leq (\lambda/2)^{-p}\|Tg\|_p^p + (\lambda/2)^{-q}\|Th\|_q^q.$$

Now, combine these inequalities to deduce

$$\mu(Tf, \lambda) \leq C_p^p (\lambda/2)^{-p} \int_{\{x : |f(x)| > a\lambda\}} |f(x)|^p \mu(dx) +$$

$$+ C_q^q (\lambda/2)^{-q} \int_{\{x : |f(x)| \leq a\lambda\}} |f(x)|^q \mu(dx),$$

for any $\lambda > 0$.

Hence, integrate the density function

$$\|Tf\|_r^r = r \int_0^\infty \lambda^{r-1} \mu(Tf, \lambda) d\lambda \leq$$

$$\leq r(2C_p)^p \int_0^\infty \lambda^{r-p-1} d\lambda \int_{\{x : |f(x)| > a\lambda\}} |f(x)|^p \mu(dx) +$$

$$+ r(2C_q)^q \int_0^\infty \lambda^{r-q-1} d\lambda \int_{\{x : |f(x)| \leq a\lambda\}} |f(x)|^q \mu(dx),$$

and exchange the order of integration to obtain

\[
\|Tf\|_r^r \leq r(2C_p)^p \int_X |f(x)|^p \mu(dx) \int_0^{\lambda^{r-p}} \lambda^{r-p-1} d\lambda + \\
+ r(2C_q)^q \int_X |f(x)|^q \mu(dx) \int_\lambda^{\infty} \lambda^{r-q-1} d\lambda,
\]

which yields

\[
\|Tf\|_r^r \leq \frac{r(2C_p)^p}{a^{r-p}(r-p)} \int_X |f(x)|^{p+r-p} \mu(dx) + \\
+ \frac{r(2C_q)^q a^{q-r}}{(q-r)} \int_X |f(x)|^{q+r-q} \mu(dx),
\]

i.e.,

\[
\|Tf\|_r^r \leq r \left( \frac{(2C_p)^p}{a^{r-p}(r-p)} + \frac{(2C_q)^q a^{q-r}}{(q-r)} \right) \|f\|_r^r.
\]

Choose the number \(a\) such that

\[
\frac{(2C_p)^p}{a^{r-p}(r-p)} = \frac{(2C_q)^q a^{q-r}}{(q-r)}
\]

and the desired estimate follows when \(q < \infty\). Note that \(1/r = (1 - \theta)/p + \theta/q\) is equivalent to either \(\theta = [q(q-r)]/[r(q-p)]\) or \(1 - \theta = [p(q-r)]/[r(q-p)]\).

If \(q = \infty\) then take \(a = 1/(2C_\infty)\) to check that

\[
\|Th\|_\infty \leq C_\infty \|h\|_\infty \leq C_\infty a \lambda = \lambda/2,
\]

which yields

\[
\mu(Tf, \lambda) \leq \mu(Tg, \lambda/2) \leq (\lambda/2)^{-p} \|Tg\|^p_p.
\]

As early, combine this inequality with the assumption that the operator \(T\) is of weak-type \((p,p)\) to obtain

\[
\mu(Tf, \lambda) \leq C_p^p (\lambda/2)^{-p} \|g\|^p_p = C_p^p (\lambda/2)^{-p} \int_{\{x:|f(x)| > \lambda/(2C_\infty)\}} |f(x)|^p \mu(dx),
\]

Hence, as in the previous calculation, integrate the density function and exchange the order of integration to deduce

\[
\|Tf\|_r^r \leq r(2C_p)^p (2C_\infty)^{r-p} \|f\|_r^r,
\]

and again the interpolation estimate follows.
Revising the above proof, it is clear that if \( T \) is only quasi-linear, i.e., there exists a constant \( C > 0 \) such that
\[
|T(f + g)(x)| \leq C(|T(f)(x)| + |T(g)(x)|), \quad \forall f, g \in S(X, \mu),
\]
then Theorem 2.43 remains valid with the constant \( C_r \) multiplied by \( C \). Moreover, Marcinkiewicz’ interpolation theorem can be extended as follows: If \( T \) be a quasi-linear operator defined on \( S(X, \mu) \) with values in \( L^0(Y, \nu) \) of weak-type \( (p, \bar{p}) \) and weak-type \( (q, \bar{q}) \) with \( 0 < \bar{p} < \bar{q} \leq \infty \) then \( T \) is of strong type \( (r, \bar{r}) \), with \( \bar{p} < \bar{r} < \bar{q}, \theta \) in \( (0, 1) \), \( 1/r = (1 - \theta)/p + \theta/q \) and \( 1/\bar{r} = (1 - \theta)/\bar{p} + \theta/\bar{q} \).

Again, by convention, if \( q = \infty \) or \( \bar{q} = \infty \) then \( 1/q = 0 \) or \( 1/\bar{q} = 0 \) in the above expressions.

**Remark 2.44.** Recall the Hardy-Littlewood maximal function, i.e.,
\[
f \mapsto f^*(x) = \sup_{r>0} F(x, r), \quad F(x, r) = \frac{1}{|B(x, r)|} \int_{B(x, r)} |f(y)| \, dy,
\]
where \(| \cdot |\) denotes the Lebesgue measure and \( B(x, r) \). This shows that the quasi-linear operator \( Tf = f^* \) is of weak-type \((1,1)\). Since it is also of strong-type \((\infty, \infty)\), we deduce that \( T \) is of strong type \((p,p)\), for any \( 1 < p \leq \infty \). \( \square \)

For instance, the interested reader is referred to Adams and Fournier [3, Chapter 2, pp. 52–58] or Bergh and L"ofstr"om [18, Section 1.3, pp. 6–11] or Grafakos [57, Section 1.4.4, pp. 55–63].

### 2.5.3 Riesz-Thorin Interpolation Theorem

This technique is called the complex method. It is based on Hadamard’s three lines lemma. The proof of this result uses the maximum principle for analytic functions, namely, if \( f \) is a bounded (complex) analytic function on a bounded domain \( D \) (e.g., on a complex rectangle \( D = \{ z \in \mathbb{C} : z = x + iy, a < x < b, c < y < d \} \)) with constants \( a < b \) and \( c < d \) then the maximum value of the function \( z \mapsto |f(z)| \) is attained at the boundary \( \partial D \), i.e., \( \max_D |f| = \max_{\partial D} |f| \).

**Lemma 2.45.** If \( f \) is an analytic function on the open unit strip \( S = \{ z \in \mathbb{C} : z = x + iy, 0 < x < 1, y \in \mathbb{R} \} \), continuous and bounded on its closure \( \overline{S} \), and such that \( |f(\bar{y})| \leq A \) and \( |f(1 + \bar{y})| \leq B \), for every \( y \in \mathbb{R} \), and some positive constants \( A \) and \( B \), then \( |f(x + iy)| \leq A^{1-x}B^y \), for every \( z = x + iy \) in \( S \).

**Proof.** Indeed, consider the analytic functions
\[
F(z) = \frac{f(z)}{A^{1-z}B^z} \quad \text{and} \quad F_n(z) = F(z)e^{(z^2-1)/n}, \quad n \geq 1.
\]

On the closed unit strip, use that fact that \( z \mapsto |f(z)| \) is bounded from above and \( z \mapsto |A^{1-z}B^z| \) is bounded from below to deduce that \( z \mapsto |F| \) is bounded by some constant \( C \) on the closed unit strip. The proof is completed by checking that \( C \leq 1 \).
To this effect, note that the functions $z \mapsto |F(z)|$ and $z \mapsto |F_n(z)|$ are also bounded by 1 on the boundary $\partial S = \{ z \in \mathbb{C} : z = x + iy, x = 0 \text{ or } x = 1, y \in \mathbb{R} \}$, and also that

$$|F_n(x + iy)| \leq Me^{-y^2/n}e^{(x^2-1)/n} \leq Me^{-y^2/n}, \forall z = x_1y \in S.$$ 

Thus $F_n(z) \to 0$ as $|y| \to \infty$, uniformly on $x$ in $[0, 1]$, and therefore, for each $n$ there exists a constant $r_n > 0$ such that $|y| \geq r_n$ implies $|F_n(x + iy)| \leq 1$ for every $x$ in $[0, 1]$. The maximum principle applied to the analytic function $F_n(z)$ on the rectangle $R = [0, 1] \times \mathbb{R}[-r_n, r_n]$ implies that $|F_n(z)| \leq 1$, for any $z \in R$. Hence $|F_n(z)| \leq 1$ on the whole closed strip. Since $F_n(z) \to F(z)$ as $n \to \infty$, this yields $|F(z)| \leq 1$ on the closed strip, i.e., $C \leq 1$ as desired.

Recall that $S(X, \mu)$ denotes the subspace of all (equivalence classes of) simple $\mu$-integrable functions defined on $X$ and that $L^0(Y, \nu)$ denotes the space of all (equivalence classes of) $\nu$-measurable $\sigma$-finite supported (and finite-valued $\nu$-almost everywhere) functions defined on $Y$. Note that in both cases, complex-valued functions are considered.

**Theorem 2.46.** If $T$ be a linear operator defined on $S(X, \mu)$ with values in $L^0(Y, \nu)$ of strong type $(p, \bar{p})$ and strong type $(q, \bar{q})$ with $1 \leq p, q, \bar{p}, \bar{q} \leq \infty$ then $T$ is of strong type $(r, \bar{r})$, i.e.,

$$\|Tf\|_{\bar{p}} \leq \|T\|_{\bar{p},p}\|f\|_p \quad \text{and} \quad \|Tf\|_{\bar{q}} \leq \|T\|_{\bar{q},q}\|f\|_q, \forall f \in S(X, \mu),$$

with some constants $\|T\|_{\bar{p},p}$ and $\|T\|_{\bar{q},q}$ independent of $f$, then

$$\|Tf\|_{\bar{r}} \leq \|T\|_{\bar{p},p}^{1-\theta}\|T\|_{\bar{q},q}^\theta\|f\|_{\bar{r}}, \forall f \in S(X, \mu),$$

for any $\theta$ in $(0, 1)$, with $1/r = (1-\theta)/p + \theta/q$ and $1/\bar{r} = (1-\theta)/\bar{p} + \theta/\bar{q}$. Again, by convention, if $q = \infty$ or $\bar{q} = \infty$ then $1/q = 0$ or $1/\bar{q} = 0$.

**Proof.** Consider simple functions defined on $X$ and on $Y$, i.e.,

$$f = \sum_{k=1}^n a_k e^{i\alpha_k} 1_{A_k}, \quad g = \sum_{k=1}^m b_k e^{i\beta_k} 1_{B_k},$$

where $a_k, b_k > 0$, $\alpha_k, \beta_k$ are real, and $\{A_k\}$ and $\{B_k\}$ are sequences of disjoint measurable sets of finite $\mu$-measure and $\nu$-measure, respectively.

Write the $L^{\bar{r}}$-norm in its dual form as

$$\|Tf\|_{\bar{r}} = \sup \left| \int_Y Tg(y) \nu(dy) \right|,$$

where the supremum is taken over all simple functions $g$ with norm $\|g\|_{\bar{r}'} \leq 1$, $1/\bar{r} + 1/\bar{r}' = 1$. For any complex $z$ in the closed unit strip $\mathbb{S} = \{ z \in \mathbb{C} : 0 \leq \Re(z) \leq 1 \}$ first define

$$R(z) = \frac{r}{p} (1 - z) + \frac{r}{q} z, \quad \bar{R}'(z) = \frac{\bar{r}'}{\bar{p}'} (1 - \bar{z}) + \frac{\bar{r}'}{\bar{q}'} \bar{z},$$

$$[\text{Preliminary}] \quad \text{Menaldi} \quad \text{November 11, 2016}$$
with $1/p + 1/p' = 1$ and $1/q + 1/q' = 1$, and next define the function

$$F(z) = \int_Y T f_z(y) g_z(y) \, \nu(dy)$$

with

$$f_z = \sum_{k=1}^{n} a_k^{R(z)} e^{i\alpha_k} \mathbb{1}_{A_k}, \quad g_z = \sum_{k=1}^{m} b_k^{R(z)} e^{i\beta_k} \mathbb{1}_{B_k}.$$ 

Note that by linearity of the operator $T$ and because $a_k, b_k > 0$, the function $F(z)$ is analytic on the open unit strip.

On the boundary $\Re(z) = 0$, the equalities

$$|a_k^{R(z)}| = a_k^{r/p} \quad \text{and} \quad |b_k^{R(z)}| = b_k^{r/\bar{p}}$$

and the fact that each sequence $\{A_k\}$ and $\{B_k\}$ have disjoint sets, imply

$$\|f_z\|_p = \|f\|_{\bar{r}} \quad \text{and} \quad \|g_z\|_{\bar{p}'} = \|g\|_{\bar{r}'}.$$ 

Similarly, the equalities

$$\|f_z\|_q = \|f\|_{r'} \quad \text{and} \quad \|g_z\|_{\bar{q}'} = \|f\|_{\bar{r}'},$$

holds true on the boundary $\Re(z) = 1$.

Apply Hölder inequality and use the fact that $T$ is of strong type $(p, \bar{p})$ to obtain, first on boundary $\Re(z) = 0$

$$|F(z)| \leq \|T f_z\|_{\bar{p}} \|g_z\|_{\bar{p}'} \leq \|T\|_{\bar{p}, p} \|f_z\|_{p} \|g_z\|_{p'} = \|T\|_{\bar{p}, p} \|f\|_{r/p} \|g\|_{r'/\bar{p}'},$$

and next, on boundary $\Re(z) = 1$

$$|F(z)| \leq \|T f_z\|_{\bar{q}} \|g_z\|_{\bar{q}'} \leq \|T\|_{\bar{q}, q} \|f_z\|_{q} \|g_z\|_{q'} = \|T\|_{\bar{q}, q} \|f\|_{r'/q} \|g\|_{r'/\bar{q}'}.$$ 

At this point, Hadamard’s three lines Lemma 2.45 yields

$$|F(z)| \leq \left(\|T\|_{\bar{p}, p} \|f\|_{r/p} \|g\|_{r'/\bar{p}'}\right)^{1-\theta} \left(\|T\|_{\bar{q}, q} \|f\|_{r'/q} \|g\|_{r'/\bar{q}'}\right)^{\theta} = \|T\|_{\bar{p}, p}^{1-\theta} \|T\|_{\bar{q}, q}^{\theta} \|f\|_{r/(1-\theta)/p+\theta/q} \|g\|_{r'/r'(1-\theta)/\bar{p}'+\theta/\bar{q}'} = \|T\|_{\bar{p}, p}^{1-\theta} \|T\|_{\bar{q}, q}^{\theta} \|f\|_{r} \|g\|_{r'}$$

when $\Re(z) = \theta$.

Note that $R(\theta) = 1$ and $\bar{R}(\theta) = 1$, and thus

$$F(\theta) = \int_Y T f(y) g(y) \, \nu(dy).$$
This implies
\[ \left| \int_Y T f(y) g(y) \nu(dy) \right| \leq \|T\|_{1-p}^{1-\theta} \|T\|_{q}^{\theta} \|f\|_r \|g\|_{r'}, \]
and the desired estimate follows from the duality expression of the \( r \)-norm. \( \square \)

Another way of reading Theorem 2.46 is as follows: A continuous and linear operator from \( L^p \) into \( L^p \) and from \( L^q \) into \( L^q \) with \( 1 \leq p, p, q, \bar{q} \leq \infty \) is also continuous from \( L^r \) into \( L^r \) with \( 1/r = (1-\theta)/p + \theta/q \) and \( 1/\bar{r} = (1-\theta)/\bar{p} + \theta/\bar{q} \), for any \( \theta \) in \( (0,1) \). For instance, the interested reader may check Grafakos [57, Section 1.3.3, pp. 37–39] for interpolation of analytic families of operators.

### 2.5.4 Intermediate Spaces

If the vector space \( L^0 = L^0(X, \mu) \) (also a topological vector space with the convergence in measure over every set of finite measure) is taken as the big reference space with respect to two Banach space \( B_i \) with norm \( \| \cdot \|_i \), with \( i = 0, 1 \) (which is not necessarily the \( L^1 \)-norm). The algebraic sum of an element in \( B_0 \) and an element in \( B_1 \) makes sense as an element in the reference vector space \( L^0 \), i.e. \( B_0 + B_1 \subset L^0 \). If the Banach spaces \( B_i \) is continuous and linear operator from \( L^p \) into \( L^p \) and from \( L^q \) into \( L^q \) with \( 1 \leq p, p, q, \bar{q} \leq \infty \) is also continuous from \( L^r \) into \( L^r \) with \( 1/r = (1-\theta)/p + \theta/q \) and \( 1/\bar{r} = (1-\theta)/\bar{p} + \theta/\bar{q} \), for any \( \theta \) in \( (0,1) \). For instance, the interested reader may check Grafakos [57, Section 1.3.3, pp. 37–39] for interpolation of analytic families of operators.

2.5.4 Intermediate Spaces

If the vector space \( L^0 = L^0(X, \mu) \) (also a topological vector space with the convergence in measure over every set of finite measure) is taken as the big reference space with respect to two Banach space \( B_i \) with norm \( \| \cdot \|_i \), with \( i = 0, 1 \) (which is not necessarily the \( L^1 \)-norm). The algebraic sum of an element in \( B_0 \) and an element in \( B_1 \) makes sense as an element in the reference vector space \( L^0 \), i.e. \( B_0 + B_1 \subset L^0 \). If the Banach spaces \( B_i \) is continuous and linear operator from \( L^p \) into \( L^p \) and from \( L^q \) into \( L^q \) with \( 1 \leq p, p, q, \bar{q} \leq \infty \) is also continuous from \( L^r \) into \( L^r \) with \( 1/r = (1-\theta)/p + \theta/q \) and \( 1/\bar{r} = (1-\theta)/\bar{p} + \theta/\bar{q} \), for any \( \theta \) in \( (0,1) \). For instance, the interested reader may check Grafakos [57, Section 1.3.3, pp. 37–39] for interpolation of analytic families of operators.

2.5.4 Intermediate Spaces

If the vector space \( L^0 = L^0(X, \mu) \) (also a topological vector space with the convergence in measure over every set of finite measure) is taken as the big reference space with respect to two Banach space \( B_i \) with norm \( \| \cdot \|_i \), with \( i = 0, 1 \) (which is not necessarily the \( L^1 \)-norm). The algebraic sum of an element in \( B_0 \) and an element in \( B_1 \) makes sense as an element in the reference vector space \( L^0 \), i.e. \( B_0 + B_1 \subset L^0 \). If the Banach spaces \( B_i \) is continuous and linear operator from \( L^p \) into \( L^p \) and from \( L^q \) into \( L^q \) with \( 1 \leq p, p, q, \bar{q} \leq \infty \) is also continuous from \( L^r \) into \( L^r \) with \( 1/r = (1-\theta)/p + \theta/q \) and \( 1/\bar{r} = (1-\theta)/\bar{p} + \theta/\bar{q} \), for any \( \theta \) in \( (0,1) \). For instance, the interested reader may check Grafakos [57, Section 1.3.3, pp. 37–39] for interpolation of analytic families of operators.

2.5.4 Intermediate Spaces

If the vector space \( L^0 = L^0(X, \mu) \) (also a topological vector space with the convergence in measure over every set of finite measure) is taken as the big reference space with respect to two Banach space \( B_i \) with norm \( \| \cdot \|_i \), with \( i = 0, 1 \) (which is not necessarily the \( L^1 \)-norm). The algebraic sum of an element in \( B_0 \) and an element in \( B_1 \) makes sense as an element in the reference vector space \( L^0 \), i.e. \( B_0 + B_1 \subset L^0 \). If the Banach spaces \( B_i \) is continuous and linear operator from \( L^p \) into \( L^p \) and from \( L^q \) into \( L^q \) with \( 1 \leq p, p, q, \bar{q} \leq \infty \) is also continuous from \( L^r \) into \( L^r \) with \( 1/r = (1-\theta)/p + \theta/q \) and \( 1/\bar{r} = (1-\theta)/\bar{p} + \theta/\bar{q} \), for any \( \theta \) in \( (0,1) \). For instance, the interested reader may check Grafakos [57, Section 1.3.3, pp. 37–39] for interpolation of analytic families of operators.
Now, the nonlinear functionals
\[ J(t, f) = \max \{ \|f\|_0, t\|f\|_1 \}, \]
\[ K(t, f) = \inf \{ \|f_0\|_0 + t\|f_1\|_1 : f = f_0 + f_1, f_i \in B_i, i = 0, 1 \}. \]

It is clear that
\[ \min\{1, t\} \|f\|_{\text{max}} \leq J(t, f) \leq \max\{1, t\} \|f\|_{\text{max}}, \]
\[ \min\{1, t\} \|f\|_{\text{inf}} \leq K(t, f) \leq \max\{1, t\} \|f\|_{\text{inf}}, \]

and that both, as function of \( t \), are continuous and increasing function from \((0, \infty)\) into \([0, \infty] \), and \( J \) is convex and \( K \) is concave. Moreover,
\[ K(t, f) \leq (\min\{1, s^{-1}t\})J(s, f), \quad \forall s, t > 0, f \in B_0 \cap B_1. \]

Recall the Banach space \( \ell^q \), with \( 1 \leq q \leq \infty \) of all \( q \)-bounded sequences, i.e.,
\[ a = \{a_i : k = 0, \pm 1, \pm 2, \ldots\} \]
belongs to \( \ell^q \) iff
\[ \|a\|_\infty = \sup_i |a_i| \quad \text{and} \quad \|a\|_q = \left( \sum_i |a_i|^q \right)^{1/q}, \quad \text{if } 1 \leq q < \infty. \]

**Definition 2.47.** Two interpolation spaces are defined for either \( 1 \leq q < \infty \) and \( 0 < \theta < 1 \) or \( q = \infty \) and \( 0 \leq \theta \leq 1 \) as follows:
(a) An element \( f \) in \( B_0 + B_1 \) belongs to \((B_0, B_1)_{\theta,q,J}\) iff \( f = \sum_i f_i \) with \( \sum_i \|f_i\|_{\text{inf}} < \infty \) and the sequence \( f_\theta = \{2^{-i\theta}K(2^{i\theta}, f) : i = 0, \pm 1, \pm 2, \ldots\} \)
becomes to \( \ell^q \), and \( \|f\|_{\theta,q,J} = \inf \{\|f_\theta\|_q\} \), where the infimum is taken over all possible representation of \( f = \sum_i f_i \).
(b) An element \( f \) in \( B_0 + B_1 \) belongs to \((B_0, B_1)_{\theta,q,K}\) iff the sequence \( f_\theta = \{2^{-i\theta}K(2^{i\theta}, f) : i = 0, \pm 1, \pm 2, \ldots\} \)
becomes to \( \ell^q \), and \( \|f\|_{\theta,q,K} = \|f_\theta\|_q. \)

The norm \( \|f\|_{\theta,q,K} \) can be equivalently defined as
\[ \|f\|_{\theta,q,K} = \int_0^\infty (t^{-\theta}K(t, f))^q \frac{dt}{t}, \quad 1 \leq q < \infty, \ 0 < \theta < 1, \]
\[ \|f\|_{\theta,\infty,K} = \sup_{t > 0} \{t^{-\theta}K(t, f)\}, \quad q = \infty, \ 0 \leq \theta \leq 1, \]
while the norm \( \|f\|_{\theta,q,J} \) can be equivalently defined as
\[ \|f\|_{\theta,q,J} = \inf \left\{ \left( \int_0^\infty \left[ t^{-\theta}J(t, u(t)) \right]^q \frac{dt}{t} \right)^{1/q} \right\}, \quad 1 \leq q < \infty, \ 0 < \theta < 1, \]
\[ \|f\|_{\theta,\infty,J} = \inf \left\{ \text{ess-sup} \left\{ t^{-\theta}J(t, u(t)) \right\} \right\}, \quad q = \infty, \ 0 \leq \theta \leq 1, \]
where the infimum is taken over all possible representation as a \( B_0 + B_1 \)-valued Bochner integral
\[ f = \int_0^\infty u(t) \frac{dt}{t}, \]
with a $B_0 \cap B_1$-valued measurable function $t \mapsto u(t)$. Moreover, it can be proved that $(B_0, B_1)_{\theta,q,J}$ and $(B_0, B_1)_{\theta,q,K}$ are intermediate Banach spaces between $B_0 \cap B_1$ and $B_0 + B_1$, and that

$$\begin{align*}
\min\{1, t\} \|f\|_{\max} &\leq C_{\theta,q,J} \|f\|_{\theta,q,J} \leq \max\{1, t\} \|f\|_{\max}, \\
\min\{1, t\} \|f\|_{\inf} &\leq C_{\theta,q,K} \|f\|_{\theta,q,K} \leq \max\{1, t\} \|f\|_{\inf},
\end{align*}$$

for some suitable constants $C_{\theta,q,J}$ and $C_{\theta,q,K}$. There are many more results concerning reiteration of interpolation spaces, the so-called exact interpolation, and the class where both, the $K$ and the $J$ interpolation coincide. For instance, the interested reader is referred to Adams and Fournier [3, Chapter 7, pp. 205–260] and Bennett and Sharpley [16, Chapter 3, pp. 96–1005], as well as Tartar [121, Lectures 21–26, pp. 103–129] and others texts.
Chapter 3

Elements of Distributions Theory

In all this section $\Omega$ denotes an open domain of $\mathbb{R}^d$, i.e., a connected open set satisfying $\overline{\Omega} = \Omega$, the interior of the closure reproduces the initial open set. Sometimes, only open set suffices, in any way, $\Omega$ would be a typical simple example of a locally compact space.

As discussed later, test functions are elements of the space $C_0^\infty(\Omega)$. The dual of this space (under a suitable topology), denoted by $\mathcal{D}'(\Omega)$, will be the focus of this section, and elements there will be called distributions in $\Omega$. The key idea is to realize that practically any usable linear operation (specially differentiation) can be done inside the space $C_0^\infty(\Omega)$, and using duality, it can also be considered in $\mathcal{D}'(\Omega)$. This space $\mathcal{D}'(\Omega)$ is so big that essentially any usable element is in there, e.g., locally integrable functions or Radon measures in $\Omega$ can be thought as being distributions. Certainly, a comprehensive treatment on the theory of distributions and locally convex topological vector spaces can be found in several places, e.g., Schwartz [112] or Tréves [125], among many other books. The whole book by Hörmander [68] is an excellent source for details of most of what follows. Some recent books may also be beneficial for the reader, e.g., Duistermaat and Kolk [37], Grubb [61], Narici and Beckenstein [92], Strichartz [117], among others. Again in this chapter, for instance, the comprehensive guide (to infinity dimensional analysis) Aliprantis and Border [6] is valuable to some readers.

3.1 Locally Convex Spaces

A vector space structure is necessary to discuss convexity properties, and so, convex sets are well defined in a topological vector spaces. As briefly developed below, the interaction of convexity with topology give rise to the so-called locally convex topological vector spaces.

A subset $A$ of a vector space $V$ is (1) convex if $tu + (1 - t)v \in A$ for every
u, v ∈ A and t ∈ (0, 1); (2) absorbing (or radial) if for every v ∈ V there exists s > 0 such that tv ∈ A for every t ≥ s; (3) balanced (or circled) if tv ∈ A for every |t| ≤ 1 and v ∈ A. A seminorm p on a vector space V is a (nonnegative) function p : V → [0, ∞) such that (a-homogeneous) p(λv) = |λ|p(v), for every scalar λ and any vector v, (b-subadditive) p(u + v) ≤ p(u) + p(v), for every vectors u and v. It is not hard to show that for any seminorm p, the set A = \{ v ∈ V, p(v) < 1 \} is convex, absorbing and balanced, and that p = p_A, where

\[ p_A(v) = \inf \{ t > 0 : t^{-1}v \in A \}, \quad \forall v \in V. \]

is called a Minkowski functional. Moreover, for any convex, absorbing and balanced set A, the functional p_A is a seminorm. In general, the infimum p_A could be defined for any set A, and \{ v : p_A(v) < 1 \} ⊂ A ⊂ \{ v : p_A(v) ≤ 1 \}.

Thus a seminorm is almost a norm, except that we may have p(v) = 0 for some v ≠ 0, and then, to obtain a Hausdorff topology, we need a family \{ p_i : i ∈ I \} of seminorms defined on a vector space V, which separate points (called a separating family of seminorms), i.e., such that for any vector v there exists some p_i such that p_i(v) > 0. The topology in V is generated by open semi-balls \( O = \{ v ∈ V : p_i(v - u) < a \} \), for any a > 0, any point u in V and any index i in I, i.e., a basis for this topology is the family

\[ \{ v ∈ V : p_i(v - u) < a, \ ∀ i ∈ J \}, \quad a > 0, \ u ∈ V, \ J \text{ a finite subset of } I. \]

It is clear that in this topology, every seminorms p_i is a continuous function. Moreover, the addition of vectors and the scalar multiplication are continuous operations, i.e., V becomes a topological vector space, and so, the topology is invariant under translations and (non-zero) scalar multiplications. In particular, only a basis of open sets containing 0 suffices to characterize the topology, i.e., the family of finite intersections of set of the form \{ v ∈ V : p_i(v) < 1/n \} for i ∈ I and n ≥ 1 is a local base of convex, absorbing and balanced open sets.

The vector space V with a topology given by a separating family of seminorms is called a locally convex topological vector space, in short lctvs. Because the seminorm p satisfies

\[ |p(u) - p(v)| ≤ p(u - v), \quad \forall u, v ∈ V, \]

the continuity of p at any point reduces to the continuity only at the origin.

It should be clear that a normed space is also a lctvs, and if the separating family of seminorms is countable, say \{ p_i : i = 1, 2, \ldots \} (which can be ordered satisfying p_i ≤ p_j if i ≤ j), then the lctvs is (usually) called a quasi-normed space or seminormed space. As described later on text, in the same way that a complete normed space is known as a Banach space, a complete quasi-normed space is referred as a Fréchet space. However, there interesting (and very general) lctvs which are not quasi-normed spaces.

• Remark 3.1. If F is a closed, convex, symmetric and absorbing set in a lctvs of second category then F is a neighborhood of zero, i.e., there exist an open set O containing zero such that O ⊂ F. Indeed, absorbing means that V = \bigcup_k kF,
and because $F$ is second category, $F$ cannot be nowhere dense, i.e., there exist $x$ in $F$ and an open balanced set $O$ containing zero such that $x + O \subset F$. The relation $-x + O = -x - O \subset -F = F$ follows form the symmetry of $F$ and $O$, and then, the convexity implies that $y = [(x + y) + (-x + y)]/2$ belongs to $F$ for every $y$ in $O$, i.e., $O \subset F$.

- **Remark 3.2.** Without entering in details, when dealing with sequential topology recall that a set $F$ is (sequential) closed if and only if the limit of any converging sequence (of points in $F$) belongs to $F$. Equivalently, a set $O$ is (sequential) open if and only if every sequence converging to some point in $O$ is eventually in $O$. This means that a sequence $\{x_n\}$ is converging to $x$ if and only if for any subsequence of $\{x_n\}$ it is possible to extract a further subsequence convergent to $x$. This is usually called a space of type $L^*$ or a sequential convergence of type $L^*$ or Fréchet-Urysohn space. Note that a topological space with a countable local base or subbase (i.e., satisfying the so-called first axiom of countability) is necessarily a sequential topology.

- **Remark 3.3.** If we use a separating family pseudo-metric (instead of seminorms) to define the topology then we obtain a completely regular topological space, i.e., the topology is given by a family of pseudo-metric (or semi-metric) which is Hausdorff separated, i.e., a family $\{d_i : i \in I\}$ of pseudo-metric such that for every two points $x \neq y$ there exists an index $i$ satisfying $d_i(x, y) > 0$. These are more general topological spaces (next to the Polish spaces), where a vector structure is not necessarily assumed. A completely regular topological space is a space of type $L^*$. Indeed, if a sequence $\{x_n\}$ does not converge to $x$ then there exists a pseudo-metric $d_i$, $\varepsilon > 0$ and a subsequence $\{x_{n_k}\}$ such that $d_i(x_{n_k}, x) \geq \varepsilon$, for every $k$. Hence, the subsequence $\{x_{n_k}\}$ does not admit a further subsequence convergent to $x$. Therefore, in a completely regular topological space, the convergence of sequences completely defines its topology, i.e., a set $F$ is closed if and only if the limit of any converging sequence (of point in $F$) belongs to $F$.

We say that a subset $B$ of a locally convex topological vector space $V$ is bounded if for every open subset $O$ containing 0 there exists $t > 0$ such that $B \subset tO = \{tv : v \in O\}$.

**Lemma 3.4.** Let $V$ be a vector space with a locally convex topology given by a separating family of seminorms $\{p_i : i \in I\}$. Then (1) a seminorm $p$ is continuous if and only if there exists a finite set of indices $J \subset I$ and a constant $C > 0$ such that $p \leq C \max_{i \in J} p_i$; (2) a set $B$ is bounded if and only if any continuous seminorm $p$ is bounded on $B$.

**Proof.** To check (1), we note that only continuity at zero matters, and if $p$ is continuous then for every $\varepsilon > 0$ there exists $\delta > 0$ and $J$ such that $\max_{i \in J} p_i(v) < \delta$ implies $p(v) < \varepsilon$. Hence, for any $v$ in $V$ the vector $u = \delta v / [r + \max_{i \in J} p_i(v)]$, with $r > 0$, satisfies $p_i(u) = \delta p_i(v) / [r + \max_{i \in J} p_i(v)] < \delta$. This implies that

$$p(u) = \frac{\delta p(v)}{r + \max_{i \in J} p_i(v)} < \varepsilon,$$

and, as \( r \to 0 \) we deduce \( p(v) \leq C \max_{i \in J} p_i(v) \), for any \( v \in V \), with \( C = \varepsilon/\delta \). The converse follows from the fact that \( \max_{i \in J} p_i \) is a continuous seminorm.

To prove (2), if \( B \) is an bounded set and \( p \) is a continuous seminorm then \( O = \{ v \in V : p(v) < 1 \} \) is an open set, and there exist \( t > 0 \) such that \( A \subset tO \), i.e., \( p(v) < t \) for every \( v \in B \). Conversely, given an open set \( O \) containing the origin, there exist \( a > 0 \) and a finite number of seminorms \( p_i \), with \( i \in J \), finite set, such that \( O \supset \{ v \in V : p_i(v) < a, \forall i \in J \} \). Since \( \max_{i \in J} p_i \) is a continuous seminorm, there exists \( C > 0 \) such that \( \max_{i \in J} p_i(v) < C \), for every \( v \in B \), i.e., \( B \subset tO \), for \( t = C/a \).

• **Remark 3.5.** Since the topology is invariant under translations, it is clear that a linear functional is continuous if and only if it is continuous at the origin. Moreover, if \( V \) is a vector space with a locally convex topology given by a separating family of seminorms \( \{ p_i : i \in I \} \), then the arguments of Lemma 3.4 shows that a linear functional \( f : V \to \mathbb{C} \) is continuous if and only if there exists a continuous seminorm \( p \) such that \( |\langle f, v \rangle| \leq p(v) \), for every \( v \in V \). Similarly, by means of the equivalence (a) and (d) in Remark 2.14, a linear functional \( f : V \to \mathbb{C} \) is continuous if and only if there exists a continuous seminorm \( p \) such that \( \sup_{p(v) \leq 1} |\langle f, v \rangle| < \infty \). Similarly, a linear operator \( T \) from a lctvs \( V \) into another lctvs \( W \) is continuous if and only if for any continuous seminorm \( q \) in \( W \) there exists a continuous seminorm \( p \) on \( V \) such that \( q(Tv) \leq p(v) \), for every \( v \in V \), or equivalently, for any continuous seminorm \( q \) in \( W \) there exists a continuous seminorm \( p \) on \( V \) such that \( \sup_{p(v) \leq 1} q(Tv) < \infty \). On the other hand, a linear operator is called bounded if the image of any bounded set is a bounded set. Therefore, the arguments in the previous Lemma 3.4 proves that a linear operator \( T \) is continuous if and only if \( T \) is bounded on some neighborhood.

• **Remark 3.6.** Note that if a linear operator \( T \) between two lctvs \( X \) and \( Y \) is bounded on some neighborhood then \( T \) maps bounded sets into bounded sets. Indeed, for any continuous seminorm \( q \) in \( Y \) there exists a continuous seminorm \( p \) on \( X \) such that \( q(Tx) \leq p(x) \), for every \( x \in X \). Hence, if \( B \) is a bounded set then \( p(B) \) is a bounded set in \( \mathbb{R} \) for any continuous seminorm \( p \), which implies that \( q(TB) \) is also bounded set in \( \mathbb{R} \) for any continuous seminorm \( q \), i.e., the image \( TB \) is a bounded set in \( Y \). Sometimes, a linear operator \( T \) is called “bounded” if \( T \) maps bounded sets in \( X \) into bounded sets in \( Y \), which is the same as being bounded in a neighborhood of zero if the space \( X \) is metrizable. Indeed, in view of the previous Remark 3.5, it suffice to show that if linear operator \( T \) preserves bounded sets then \( T \) is continuous at zero. By contradiction, suppose that there exists a balanced neighborhood \( V \) of zero in \( Y \) such that the pre-image \( T^{-1}V \) does not contains any neighborhood of zero in \( X \). Because \( X \) is metrizable, there exists a countable basis \( \{ U_n \} \) of open set containing zero and points \( x_n \) in \( (1/n)U_n \setminus T^{-1}V \), for every \( n = 1, 2, ... \), which means that \( nx_n \to 0 \) in \( X \) and \( x_n \) does not belongs to \( T^{-1}V \). Because the sequence \( \{ nx_n \} \) is bounded, the image \( \{ nTx_n \} \) is also bounded, i.e., for this neighborhood \( V \) of zero there exists a number \( r > 0 \) such that \( \{ nTx_n \} \subset rV \).
Hence, recall that \( V \) is balance to deduce \( \{ Tx_n \} \subset (r/n)V \subset V \), for every \( n > r \), which contradict the condition \( x_n \not\in T^{-1}V \).

- **Remark 3.7.** Based on Remark 3.5, the uniformly boundedness principle or Banach-Steinhaus Theorem 2.17 on lcvs can be rephrased as follows: If \( \{ T_i : i \in I \} \) is a family of continuous linear operators between two lcvs \( X \) and \( Y \) such that (1) \( X \) is a space of second category and (2) for every \( x \) and any continuous seminorm \( q \) in \( Y \), the set of numbers \( \{ q(T_i x) : i \in I \} \) is bounded then the family \( \{ T_i : i \in I \} \) is equi-continuous, i.e., for every continuous seminorm \( q \) in \( Y \) there exist a continuous seminorm \( p \) in \( X \) such that \( q(T_i x) \leq p(x) \), for every \( i \in I \) and every \( x \) in \( X \).

If a family of convex, absorbing and balanced family \( \{ U_i : i \in I \} \) of sets in a vector space \( V \) satisfying \( \bigcap_{i \in I} U_i = \{ 0 \} \) then the family of Minkowski seminorms corresponding to the sets \( U_i \) is a separating family of seminorms.

A locally convex basis \( \{ U_i : i \in I \} \) in a lcvs is a family of convex, absorbing and balanced open sets such that (1) for any open set \( O \) containing the origin there exists an index \( i \) and \( t > 0 \) satisfying \( O \supset tU_i \) and (2) for any \( v \neq 0 \) there exists an index \( i \) satisfying \( p_i(v) > 0 \). In this case, the corresponding Minkowski seminorms are continuous and the lcvs generated by these seminorms is the same as the initial lcvs. This is to say that convex, absorbing and balanced open sets, i.e., open semi-balls corresponding to a continuous seminorm \( B = \{ v \in V : p(v) < 1 \} \) yield the locally convex topology.

A sequence \( \{ v_n \} \) in a vector space \( V \), with a locally convex topology defined by a separating family of seminorms \( \{ p_i : i \in I \} \), is called a Cauchy sequence if \( p_i(v_n - v_m) \to 0 \) as \( n, m \to \infty \), for any seminorm \( p_i \). The lcvs \( V \) is called complete if any Cauchy sequence has a limit.

- **Remark 3.8.** In view of Lemma 3.4, \( \{ v_n \} \) is a Cauchy sequence if and only if \( p(v_n - v_m) \to 0 \) as \( n, m \to \infty \), for any continuous seminorm \( p \). Also, a Cauchy sequence is necessarily bounded. Indeed, for any continuous seminorm there exists \( N \) such that \( p(v_n - v_m) \leq 1 \) for every \( n, m \geq N \), which implies

\[
p(v_n) \leq p(v_m) + p(v_n - v_m) \leq \max_{m \leq N} p(v_m) + 1,
\]

proving that \( \{ v_n \} \) is a bounded set in \( V \).

It is clear that if the family of seminorm \( \{ p_i : i \in I \} \) defining the locally convex topology in \( V \) is countable then \( V \) is metrizable, for instance by taking

\[
d(u, v) = \sum_{i=1}^\infty \frac{2^{-i} p_i(u - v)}{1 + p_i(u - v)}, \quad \forall u, v \in V.
\]

Moreover, if there is a metric \( d \) which complete then \( V \) is complete. A complete metrizable lcvs is called a Fréchet space, i.e., a complete lcvs with a countable basis (or family of seminorms). Because a \( F \)-space (i.e., a complete, metrizable topological vector space) is not necessarily a locally convex space, it is clear that a Fréchet space is a \( F \)-space, but the converse does not hold.
Remark 3.9. In a lctvs $X$, a barrel (or tonneau) is a convex, balanced and absorbing closed subset of $X$, see Remark 3.1. Every locally convex topological vector space has a neighbourhood basis consisting of barrel sets. Nevertheless, a space $X$ is called a barrel (or barreled) space if any barrel in $X$ is a neighborhood of zero. It can be proved (e.g., see Yosida [135, Appendix to Chapter V.2, pp. 138–139]) that a lctvs of second category is a barrel space, in particular, it is clear that Fréchet space (complete and metrizable lctvs) is a barrel space. Note that, a priori, a complete lctvs may not be of second category. A lctvs is a barrel space if and only if any seminorm which is semi-continuous from below is continuous. A related concept is bornologic (or bornological) spaces, where any balanced and convex set that absorb any bounded set is a neighborhood of zero. A lctvs is a bornological space if and only if any seminorm, which is bounded on any bounded set, is continuous, see Proposition 2.28. For instance, the interested reader may check the book Schaefer [110, Sections II.7-8, pp. 63].

Remark 3.10. The uniformly boundedness principle or Banach-Steinhaus Theorem 2.17 holds for barrel spaces, i.e., if $\{T_i : i \in I\}$ is family of continuous linear operators $T_i : X \to Y$ between two lctvs with $X$ a barrel, and for every $x$ in $X$ the set $\{T_i(x) : i \in I\}$ is a bounded in $Y$ then the family $\{T_i : i \in I\}$ is equi-continuous, i.e., the intersection $\bigcap_{i \in I} T_i^{-1}(U)$ is a neighborhood $U$ of zero in $X$ for every neighborhood $U$ of zero in $Y$. Indeed, because $Y$ is a lctvs, any neighborhood of zero $U$ contains another neighborhood of zero which is convex, balanced, absorbing and closed (i.e., a barrel) $V \subset U$. The continuity and linearity of $T_i$ ensures that $T_i^{-1}(V)$ is convex, balanced and closed, and so is the intersection $\bigcap_{i \in I} T_i^{-1}(V)$. Now, because $X$ is a barrel lctvs, to check that this intersection is a neighborhood of zero, we need to verify only that $\bigcap_{i \in I} T_i^{-1}(V)$ is absorbing. To this effect, for any point $x$ in $X$, because the set $\{T_i(x) : i \in I\}$ is a bounded in $Y$ there exists a number $r > 0$ such that $\{T_i(x) : i \in I\} \subset rV$, i.e., $x$ belongs to $rT_i^{-1}(V)$, which means that $\bigcap_{i \in I} T_i^{-1}(V)$ is absorbing. For instance, the interested reader may check the book Tréves [125, Chapter 33, pp. 347–50].

Exercise 3.1. Use the argument in Exercise 2.5 to show that the closed convex hull of a totally bounded subset $A$ in a Fréchet space is a compact set.

Exercise 3.2. Following Remark 2.16, let $N$ be a closed (vector) subspace of a locally convex topological vector space $X$ with a separating family of seminorms $\{p_i : i \in I\}$. The quotient space $X/N$ is the space of cosets $\bar{x} = x + N$. Verify that $X/N$ is a vector space and that $\{\bar{p}_i : i \in I\}$ with

$$\bar{p}_i(\bar{x}) = \inf_{x \in \bar{x}} p_i(x), \quad \forall \bar{x} \in X/S$$

is a separating family of seminorms for $X/N$, i.e., $X/N$ becomes a lctvs. Next show that if $X$ is complete, metrizable or separable then so is $X/N$. 

Exercise 3.3. Let $A$ and $B$ be two closed subsets of a topological vector space $X$. Give an example where $A + B = \{a + b : a \in A, b \in B\}$ is not necessarily
closed. Next show (1) if $A$ or $B$ is (sequentially) compact then $A + B$ is closed and (2) if $A$ and $B$ are independent closed vector subspaces, i.e., $A \cap B = \{0\}$, and $X$ is $F$-space (complete and metrizable) then $A + B$ is closed. Finally, (3) deduce that if $A$ and $B$ are closed vector subspaces and $A$ or $B$ is finite dimensional and $X$ is $F$-space then $A + B$ is also closed. What about the general case? Hint: for (2) note that the mapping $(a, b) \mapsto a + b$ is a one-to-one application from $A \times B$ onto $A + B$, and use the open mapping theorem as in Remark 2.22 to deduce that any Cauchy sequence of the form $\{a_n + b_n\}$ is pre-mapped from Cauchy sequences $\{a_n\}$ and $\{b_n\}$; for (3) use Remark 2.15 to know that any finite dimensional subspace of a topological vector space is necessarily closed. 

**Exercise 3.4.** Prove that a locally convex (Hausdorff space) is normable (i.e., there exists a norm yielding the same topology) if and only if its zero vector has a bounded neighborhood. For instance, the reader may consult the book Al-Gwaiz [4, Theorem 1.6, p.15], among others. 

### 3.1.1 Dual Spaces

The dual space of a locally convex topological space $V$ is the vector space of all continuous functional, denoted by $V'$, and endowed with the weak* topology, i.e., the locally convex topology induces by the (usually uncountable) family of seminorms $p_v(\cdot) = |\langle \cdot, v \rangle|$ with $v$ in $V$. Therefore, a sequence $\{f_n\}$ in $V'$ converges to $f$ if and only if $\langle f_n, v \rangle \to \langle f, v \rangle$, for every $v$ in $V$. However, the strong topology on $V'$ is induced by the family of seminorms $p_B(\cdot) = \sup_{v \in B} |\langle \cdot, v \rangle|$, with $B$ any bounded subset of $V$. Thus, a sequence $\{f_n\}$ in $V'$ converges to $f$ if and only if $\langle f_n, v \rangle \to \langle f, v \rangle$, uniformly on any $v$ of $B$, for every bounded subset $B$ of $V$. Briefly, for the continuous linear functional (i.e., the elements of $V'$), weak* convergence means pointwise, while strong convergence means uniformly on bounded sets.

Thus, the dual space $V'$ becomes a lctvs with either of those topologies, say, the (strong) dual space and the weakly* dual space. Moreover, on the initial vector space $V$, we may consider the same (usually uncountable) family of seminorms $p_f(\cdot) = |\langle f, \cdot \rangle|$ with $f$ in $V'$ to produce the weak topology, which makes $V$ also a lctvs. Summing up, beginning with a lctvs $V$ the dual space $V'$ is defined, but there two topologies at our disposal in either one, (1) the strong (or initial) and the weak topologies on $V$, and (2) the strong (or default) and weak* topologies on $V'$. Moreover, this process can be repeated to obtain the bidual (or double dual) space $V'' = (V')'$, which requires to specify which topology is used on $V'$ (the strong convergence is the default topology) and makes appear the weak topology on the dual space $V'$. Therefore, $V' = V'_s \oplus V'_{w^*}$, and $V_w$, using the sub-index $w$ or $s$ to identify either the weak or the strong topology. When making the bidual spaces, we limit ourselves to the default topology, i.e., $V''_s = (V'_s)'_s$, $V''_{w^*} = (V'_s)_{w^*}$, and $V''_w = (V'_s)'_w$.

The evaluation operator $\langle \cdot, \cdot \rangle$ between $V$ and its dual $V'$ (also called the pairing duality) establishes the inclusion of $V$ into its bidual $V''$, i.e., for every
If \( v \) in \( V \) the mapping \( f \mapsto \langle f, v \rangle \) can be regarded as a continuous linear functional on \( V' \) (an element of \( V'' \)). If this inclusion is surjective (onto) then the initial space \( V \) is called semi-reflexive, and reflexive when the topologies agrees, i.e., \( V'' = V \), via the pairing duality mapping from \( V \) onto \( V'' \), \( v \mapsto \langle \cdot, v \rangle \).

A couple of classic results are stated without any proof: (1) the dual of a Banach space is a Banach space, (2) any Hilbert space is reflexive, (3) if a normed space is semi-reflexive then it is also reflexive (i.e., when dealing with normed spaces, semi-reflexive is of no use), (4) the (strong) dual of a reflexive space is reflexive. Usually, all these results (and much more) are discussed in courses on functional analysis.

**Remark 3.11.** A nice simplification occurs when the weak and the strong topologies are the same, this is the case of the Montel space, which are defined as barrel lctvs where every closed and bounded set is compact, i.e., a barrel space satisfying the Heine-Borel property. Moreover, (a) (strong) duals of Montel spaces are Montel spaces, (b) Montel spaces are reflexive (actually, a lctvs is reflexive if and only if every bounded set is relatively compact), e.g., see Bourbaki [21, Section IV.19.5], or Schaefer [110, Sections IV.5.6-7].

Another key point is the *separation of convex sets* by seminorms (nonnegative, homogeneous and subadditive functional) in a topological vector space. Thus the Hahn-Banach Theorem 2.26 can be restated as

**Theorem 3.12.** Let \( V_0 \) be a subspace of a locally convex vector space \( V \) and \( f_0 \) be a continuous linear functional on \( V_0 \), i.e., \( f_0 : V_0 \to \mathbb{R} \) such that there exists a continuous seminorm \( p \) on \( V \) such that \( |\langle f_0, v \rangle| \leq p(v) \), for every \( v \) in \( V_0 \). Then \( f_0 \) can be extended to an element of \( V' \), namely, there exists a continuous linear functional \( f : V \to \mathbb{R} \) such that (a) \( \langle f, v \rangle = \langle f_0, v \rangle \) for every \( v \) in \( V_0 \) and (b) \( |\langle f, v \rangle| \leq p(v) \) for every \( v \) in \( V \), for the same continuous seminorm \( p \).

A direct application of Hahn-Banach Theorem 3.12 shows that the family \( \{ p_v(\cdot) = |\langle \cdot, v \rangle| : v \in V \} \) of seminorms defining the locally convex topology in the dual space \( V' \) separates points, so that \( V' \) is indeed a (Hausdorff) locally convex topological space. For instance, if \( p \) is a continuous seminorm in \( V \) and \( v_0 \neq 0 \) then the linear functional \( \langle f_0, v \rangle = \lambda v_0 \) defined on \( V_0 = \{ v = \lambda v_0 : \lambda \text{ scalar} \} \), can be extended to a continuous functional \( f \) on \( V \) such that \( \langle f, v_0 \rangle = p(v_0) \) and \( |\langle f, v \rangle| \leq p(v) \) for every \( v \) in \( V \). Moreover, if \( A \) and \( B \) are two non-empty, convex and disjoint sets in a lctvs \( V \) then (1) if \( A \) is open then there exists \( f \) in \( V' \) and \( r \) in \( \mathbb{R} \) such that \( \langle f, a \rangle < r \leq \langle f, b \rangle \), for every \( a \) in \( A \) and \( b \) in \( B \); (2) if \( A \) is compact and \( B \) closed then there exists \( f \) in \( V' \) and \( r, s \) in \( \mathbb{R} \) such that \( \langle f, a \rangle < r < s < \langle f, b \rangle \), for every \( a \) in \( A \) and \( b \) in \( B \). The interested reader may check, for instance, the book by Rudin [109, Chapters 3, pp. 55–86] for a detailed discussion on convexity.

If \( V \) is a complete lctvs of second category then its dual space (also called its conjugate space) \( V' \) is a complete (Hausdorff) locally convex topological vector space, with either the weak* topology or the strong topology. Indeed, if \( \{ f_n \} \) is a Cauchy sequence then \( \{ f_n(x) \} \) is a numerical Cauchy sequence with limit denoted by \( f(x) \). The linearity of \( f \) follows immediately, and the continuity
follows from uniformly boundedness principle, Theorem 2.17. It is also clear that the equality
\[ |f_n(x) - f(x)| \leq |f_m(x) - f(x)| + |f_n(x) - f_m(x)| \]
yields the convergence \( f_n \to f \) in the appropriate topology.

**Remark 3.13.** If \((V, \| \cdot \|)\) is a normed (Banach) space then its dual space \(V'\) is a normed (Banach) space with the dual norm \(\|v'\|' = \sup_{\|v\| \leq 1} |\langle v', v \rangle|\). Moreover, Hahn-Banach Theorem 3.12 implies that for every nonzero vector \(b\) in \(V\) there exists an element \(f_b\) in \(V'\) such that \(\langle f_b, b \rangle = \|b\|\) and \(\|f_b\|' = 1\). Furthermore,
\[
\|v\| = \sup_{\|v'\| = 1} |\langle v', v \rangle| = \sup_{\|v'\| \leq 1} |\langle v', v \rangle| = \sup_{\|v'\| \neq 0} \frac{|\langle v', v \rangle|}{\|v'\|'}, \tag{3.1}
\]
for every \(v\) in \(V\). Indeed, temporarily denote by \(\|v\|\) the right hand side of (3.1), and note that the definition of the dual norm implies that \(|\langle v', v \rangle| \leq \|v'\|' \|v\|\), proving that \(\|v\| \leq \|v\|\). The converse inequality follows by choosing \(v' = f_b\) with \(b = v\).

In a normed space \((V, \| \cdot \|)\), a seminorm \(p\) is continuous if and only if there exists a constant \(c > 0\) such that \(p(v) \leq c\|v\|\), for every \(v\) in \(V\); and a set \(B\) is bounded if and only if there exists a constant \(r > 0\) such that \(\|v\| \leq r\), for every \(v\) in \(B\). This implies that a linear functional \(f\) on \(V\) is continuous if and only if a constant \(c > 0\) such that \(|\langle f, v \rangle| \leq c\|v\|\), for every \(v\) in \(V\).

As mentioned in Remark 3.5 a linear functional \(f\) on \(V\) is continuous if and only if there exists a continuous seminorm \(p\) on \(V\) such that \(|\langle f, v \rangle| \leq p(v)\), for every \(v\) in \(V\), or equivalently, there exist a constant \(C > 0\) and a finite number of seminorms \(p_i, i \in J\), of the family of seminorms defining the locally convex topology in \(V\), such that \(|\langle f, v \rangle| \leq C \max_{i \in J} p_i(v)\), for every \(v\) in \(V\). On the other hand, as a particular case of Remark 3.6, if \(f\) is a linear functional in a metrizable lctvs \(V\), which maps bounded sets into bounded sets of scalars, then \(f\) is continuous.

On the dual space \(V'\) of a lctvs \(V\), there are bounded sets in the strong sense or in the weak* sense. However, if \(V\) is a barrel lctvs (or second category) then the uniformly boundedness principle Banach-Steinhaus Theorem 2.17 can be applied to deduce that any weakly* bounded set \(B'\) in \(V'\) is equi-continuous, in particular, \(B'\) is also a strongly bounded set. This is to say that if \(\sup_{b' \in B'} |\langle b', v \rangle| < \infty\), for every \(v\) in \(V\) then there exists a continuous seminorm \(p\) on \(V\) such that \(|\langle b', v \rangle| \leq p(v)\), for every \(b'\) in \(B'\).

### 3.1.2 Inductive Limits

Consider a vector space \(V\) with a locally convex topology given by a separating family of seminorms \(\{p_i : i \in I\}\), which turn out to be non complete. The possibility of completing the space \(V\) is ruled out, because, for some reason, the vector space \(V\) should be keep intact. Thus a stronger topology should be given to \(V\), so that \(V\) becomes a complete locally convex topological space.
Suppose that there is a strictly monotone sequence of vector spaces, $V_k \subset V_{k+1}$, $V_k \neq V_{k+1}$, such that (a) $V = \bigcup_{k=1}^{\infty} V_k$ and (b) the same separating family of seminorms $\{p_i : i \in I\}$ yields a complete locally convex topology on each $V_k$.

Actually, this is referred to either as strictly inductive topology, or as the LF-spaces (inductive limit of an increasing sequence of Fréchet spaces) or as LB-spaces (inductive limit of an increasing sequence of Banach spaces), e.g., see Bourbaki [21, Chapter II], Schaefer [110, Section II.6, pp. 54–60], Trèves [125, Chapter 13, pp. 126–135], or as a countably normed spaces, e.g., see Friedman [45, Chapter 1, pp. 1–24].

Moreover, assume that (c) there exists a sequence $\{q_k : k \geq 1\}$ of seminorms on $V$ satisfying for every $n > k \geq 1$,

$$q_k(v) = 0, \quad \forall v \in V_k \quad \text{and} \quad 0 < q_k(v) \leq C_n \max_{i \in J_n} p_i(v), \quad \forall v \in V_n \setminus V_k,$$

where $J_n$ is a finite set of indices of $I$. This means that $V_k$ is exactly the null space of seminorm $q_k$ on $V$, which is dominated (within each $V_n$) by the initial family of seminorms. For instance, if there is a norm $p_0$ on $V$, which is dominated as above (i.e., continuous on each $V_k$) then the sequence of quotients seminorms defined by $q_k(v) = \inf\{p_0(v + v_k) : v_k \in V_k\}$, would have the required property.

The so-called topology of inductive limits (also called final topology) uses the notion of completeness and convergence in each lcpts $V_k$ and passes them to the whole space $V$, but the complete lcpts $V$ cannot be a Fréchet space, namely, to possess a countable subbase of seminorms.

Define the uncountable family of seminorms

$$p_{i,r}(v) = \max\{p_i(v), \sup_{k} r_k q_k(v)\}, \quad \forall i \in I, \ r = \{r_k\},$$

where the sequences $r = \{r_k : k \geq 1\}$ are unbounded monotone increasing, i.e., $r_{k+1} \geq r_k > 0$ and $r_k \to \infty$. In view of (3.2), the seminorm $q_k$ vanishes on $V_n$, for $k \geq n$, and so $\sup_k r_k q_k(v) \leq r_n \max_{k \leq n} q_k(v)$, for every $v$ in $V_n$. Thus, note that $q_k$ is a continuous seminorms on $V_n$, for every $n \geq 1$, to conclude that the families $\{p_{i,r} : i, r\}$ and $\{p_i : i\}$ define the same locally convex topology on $V_n$, for every $n$.

**Proposition 3.14.** Under the previous notation, consider $V$ with the locally convex topology induced by the family of seminorms $\{P_{i,r} : i, r\}$ define above. Then any bounded set $B$ in $V$ is a bounded set in some $V_k$, and consequently, $V$ is a complete locally convex topological space. Moreover, a sequence $\{v_n\} \subset V$ converges to $v$ (i.e., $p_{i,r}(v_n - v) \to 0$ as $n \to \infty$, for every $i, r$) if and only if (1) there exists a $k$ such that $\{v_n\} \subset V_k$, and (2) $v \in V_k$ and $v_n \to v$ in $V_k$.

**Proof.** We use Lemma 3.4. Indeed, if $B$ is a bounded set in $V$ then $p_{i,r}$ is bounded over $B$, i.e., for every $i$ and $r$ there exists a constant $C = C_{i,r}$ such that $p_{i,r}(v) \leq C_{i,r}, \ \forall v \in B$. By contradiction, suppose that there exists a sequence $\{v_n : n \geq 1\}$ in $B$ such that $v_n \in V \setminus V_n$. Then $q_n(v_n) > 0$ and we may use the sequence $r = \{r_k\}$ with $r_k = k/\min\{q_n(v_n) : n \leq k\}$ to deduce
\[ p_i, r(v_n) \geq r_n q_n(v_n) \geq n, \] which implies that \( \{v_n\} \) is unbounded. Therefore, \( B \) is contained in some \( V_k \), and because \( p_{i,r}(v) \geq p_i(v) \), the set \( B \) results bounded in some \( V_k \).

Next, any Cauchy sequence in \( V \) is a bounded set and therefore, is contained in some \( V_k \). Because each \( V_k \) is a complete locally convex topological space, any Cauchy sequence is convergence, i.e., \( V \) is complete.

Finally, if a sequence \( \{v_n\} \subset V \) converges to \( v \) in \( V \) then \( \{v_n\} \) is a bounded set in \( V \) and so, it must be contained in some \( V_k \), and therefore, \( v_n \rightarrow v \) in \( V_k \).

Conversely, if a sequence \( \{v_n\} \subset V \) satisfies (1) and (2) then, in view of estimate (3.2), the restriction of each seminorm \( p_{i,r} \) to \( V_k \) is dominated by a finite number of seminorms of the initial family \( \{p_i : i \in I\} \). Hence, \( p_{i,r}(v_n - v) \rightarrow 0 \), i.e., \( v_n \rightarrow v \) in \( V \).

**Remark 3.15.** The assumption (3.2) give a convenient way of introducing the inductive limit topology. For instance, for each \( k \geq 1 \) choose a subset of indexes \( J_k \subset I \) with the property that for every \( v \) in \( V \setminus V_n \) there exist \( m \geq n \) and \( j \in J_m \) such \( p_j(v) \neq 0 \) (e.g., take \( J_k = I \)).

Next, for any \( i \) in \( J_k \) first define \( q_{i,k}(v) = \inf\{p_i(v + v_k) : v_k \in V_k\} \) and then define the family of seminorms

\[
 p_{i,j,r}(v) = \max\{p_i(v), \sup_k r_k q_{j,k}(v)\}, \quad \forall i \in I, \quad r = \{r_k\} \quad j = \{j_k\},
\]

where \( r = \{r_k : k \geq 1\} \) is any unbounded monotone increasing sequence and each \( j_k \) belongs \( J_k \). Thus, following the arguments of Proposition 3.14, the same inductive limit topology can be established. As mentioned early, if there is a norm \( p_0 \) defined on \( V \) and continuous on every \( V_k \) then take \( J_k = \{0\} \).

Moreover, the family of subspaces \( \{V_k\} \) may be uncountable, say \( V = \bigcup_{k \in K} V_k \) as long as the initial family of seminorms \( \{p_i : i \in I\} \) yields a complete lctvs on each \( V_k \), for every \( k \) in \( K \).

In other words, inductive topology is the finest (or strongest) topology that make each inclusion \( V_k \subset V \) a continuous function, i.e., a subset \( O \) of \( V \) is open if and only if \( O \cap V_k \) is open in \( V_k \) for every \( k \). Certainly, there are other more general ways of introducing the inductive limit topology, e.g. Schaefer [110, Section II.6, pp. 54–60], Narici and Beckenstein [92, Chapter 12, pp. 286–298], among others.

Note that the initial \( p_i \) and \( q_r(v) = \sup_k r_k q_k(v) : k \geq 1 \}, \) defined for unbounded increasing sequences \( r = \{r_k\} \), are continuous seminorms on \( V \). Thus, beside the convex, absorbing and balanced open sets \( U_i = \{v \in V : p_i(v) < 1\} \), with \( i \in I \), we have \( U_r = \{v \in V : q_r(v) < 1\} = \{v \in V : r_k q_k(v) < 1, \forall k\} \), for every unbounded increasing sequence \( r = \{r_k\} \).

Usually, the set of indices \( I \) is countable so that \( V_k \) is a Fréchet space (and in particular, the topology is sequential), but \( V \) is not metrizable, by construction the family of seminorms is uncountable. However, the locally convex topology in \( V \) is sequential (see Remarks 3.3 and 3.2) if \( V_k \) are so, i.e., a subset \( C \) of \( V \) is closed if and only if any convergence sequence of elements in \( C \) converges to a point in \( C \). Indeed, we have to show that a sequence \( \{v_n\} \) is convergence to \( v \) if any subsequence \( \{v_{n'}\} \) contains another subsequence \( \{v_{n''}\} \) convergent...
to \( v \). Now, if a sequence \( \{v_n\} \) possesses the previous property, then it is necessarily bounded (because any convergence sequence is bounded) and so, \( \{v_n\} \) is contained in some \( V_k \), where this previous property holds.

Therefore if \( V_k \) are Fréchet spaces then \( V \) is a complete locally convex topological space and the topology is sequential. It is clear that if each \( V_k \) is separable then \( V \) is also separable.

A sequence \( \{v_n\} \) converges to a point \( v \) in \( V \) if and only if two conditions are met, namely (1) there exists a \( k \) such that \( \{v_n\} \subset V_k \), and (2) \( v \in V_k \) and \( v_n \to v \) in \( V_k \). Hence, it is clear that initially, we could define (1) and (2) as the meaning of a convergent sequence, which a posteriori generates the topology, i.e., without explicitly introducing the uncountable family of seminorms \( \{p_i, r : i, r\} \), we could define the convergence of sequences in \( V \) by the conditions (1) and (2).

As mentioned early, the actual objective of the inductive limit is to obtain the convergence (1) and (2).

With the inductive topology, a linear operator \( T \) from \( V \) into a lctvs \( W \) is continuous (or bounded) if and only if the restriction \( T|_{V_k} \) is a continuous linear operator from \( V_k \) into \( W \). Thus, the dual space \( V' \) is the space of all linear functional \( f \) on \( V = \bigcup_k V_k \) such that when considered on \( V_k \), i.e., \( f|_{V_k} : V_k \to \mathbb{R} \) (or \( f|_{V_k} : V_k \to \mathbb{C} \)) is continuous for every \( k \geq 1 \), i.e., \( V' = \bigcap_k V_k' \).

- **Remark 3.16.** The dual space \( V_k' \) of a complete lctvs \( V_k \) is necessarily complete and the expressions \( V = \bigcup_k V_k \) and \( V' = \bigcap_k V_k' \) may be confusing, and may be giving the impression that \( V' \) is a “small space”. Actually, \( V \subset X \), where \( X \) is Fréchet space given as the intersection of Banach (or Hilbert) spaces \( X = \bigcap_k X_k \). Thus \( V' \supset X' = \bigcup_k X_k' \), and the “size” of the “big” space \( V' \) becomes clear. Since bounded sets in \( V \) are sets bounded in some \( V_k \), we deduce that either the weak or the strong topology in \( V' \) is a sequential topology, provided each \( V_k \) is a Fréchet space.

- **Remark 3.17.** Recall that a lctvs is called a barrel space if any convex, balanced and absorbing closed subset of \( X \) is a neighborhood of zero, see Remark 3.1. Thus, the inductive limit preserves barrel spaces, i.e., if each \( V_k \) is a barrel space then so is \( V = \bigcup_k V_k \) with the inductive topology. Indeed, if \( B \) is a barrel in \( V \) then, by the continuity of the identity mapping \( I : x \mapsto x \) from \( V_k \) into \( V \), the inverse image \( I^{-1}(B) = B \cap V_k \) is closed in \( V \), which implies that \( B \cap V_k \) is a barrel in \( V_k \). Thus \( B \cap V_k \) is a neighborhood of zero in \( V_k \), which means that \( B \) must be a neighborhood of zero in \( V \).

**Exercise 3.5.** On a barrel lctvs \( X \) and its dual space \( X' \), (1) show that a weakly* bounded sequence in the dual space \( X' \) is also strongly bounded. Finally, assume that \( X \) satisfies the Heine-Borel property, i.e., every closed and bounded set is compact, and (2) prove that any sequence is strongly convergence in the dual space \( X' \) if and only if it is weakly* convergence.

### 3.1.3 Test Function Spaces

Recall that an open connected subset \( \Omega \) of \( \mathbb{R}^d \) satisfying \( \Omega = \overset{\circ}{\Omega} \) (i.e., it is the interior of its closure) is called an open domain in \( \mathbb{R}^d \). The space \( C(\Omega) \) of
continuous functions on an open domain $\Omega$ of $\mathbb{R}^d$ is a Fréchet space under the locally convex topology induced by the sequence $\{p_k : k \geq 1\}$ of seminorms $p_k(u) = \sup\{|u(x)| : x \in K_k\}$, with $\Omega = \bigcup_k K_k$, and $K_k$ a compact subset of the interior of $K_{k+1}$. More general situations can be considered, $C(X;Y)$ of continuous functions from a locally compact space $X$ into a locally convex and complete space $Y$.

Now, we reconsider $C_0^0(\Omega)$, the space of real (or complex) continuous functions with compact support on an open domain $\Omega$ of $\mathbb{R}^d$, and, for any compact set $K$, the subspace $C_0^0(\Omega)$ of all functions in $C_0^0(\Omega)$ with support in $K$, which is a Banach space isomorphic to $C_0^0(K)$, the space of continuous functions on $K$ vanishing on the boundary $\partial K$ of $K$ (so extended by zero on $\Omega \setminus K$) and with the sup-norm. We write $\Omega = \bigcup_k K_k$, with $K_k$ a compact subset of the interior of $K_{k+1}$, to have a situation as described early, $V = C_0^0(\Omega)$, $V_k = C_0^0(K_k)$, $V = \bigcup_k V_k$, with the seminorm (actually the sup-norm) $p(u) = \sup\{|u(x)| : x \in \Omega\}$ and the seminorms

$$q_k(u) = \sup\{|u(x)| : x \in \Omega \setminus K_k\},$$

which satisfy $q_k(u) = 0$ for every $u \in C_0^0(K_k)$, $q_k(u) > 0$ for every $u \in C_0^0(\Omega) \setminus C_0^0(K_k)$, and $q_k(u) \leq p(u)$ for any $u \in C_0^0(\Omega)$. Thus, we endow $C_0^0(\Omega)$ with sequential locally convex and complete topology satisfying: (1) any bounded set is indeed a bounded set in $C_0^0(K)$ for some compact set $K \subset \Omega$, (2) a sequence $\{u_n\}$ converges to $u$ if and only if there exist a compact $K$ of $\Omega$ such that supp $u_n \subset K$ for every $n$ and $u_n \to u$ in $C_0^0(K)$.

At a lower level of complexity, we have the space $V = \mathbb{R}^\infty$ (or $\mathbb{C}^\infty$) of all real-valued (or complex-valued) sequences $u = \{u_i : i \geq 1\}$ and the (Euclidian) subspaces $V_k = \mathbb{R}^k$ (or $\mathbb{C}^k$) with the sup-norm $p(u) = \sup_i |u_i|$ and the seminorms $p_k(u) = \sup_{i \leq k} |u_i|$ and $q_k(u) = \sup_{i > k} |u_i|$, which satisfy $q_k(u) = 0$ for any $u \in V_k$, $q_k(u) > 0$ for every $u \in V \setminus V_k$, and $p_k(u) + q_k(u) = p(u)$ for any $u \in V$. This space $V$ (sequences of real or complex numbers), sometimes denoted by $\mathbb{R}^n$ or $\mathbb{C}^n$, becomes a Fréchet space with the topology induced by the sequence of seminorms $\{p_k : k \geq 1\}$, which the same as the product topology. Actually, this space can be viewed as the space of polynomials.

The inductive limit topology can also be applied to $\mathbb{R}^\infty$. Indeed, consider the subspace $\mathfrak{d}$ (lower case “euftrak” $d$), which could be denoted by $\mathbb{R}_0^\infty$, with the subindex 0 for “compact support”) of all sequences with only a finite number of non-zero terms, i.e., the space $\mathfrak{d}_k = \mathbb{R}^k$, after extending with zeros. The seminorms (actually sup-norm) $p$ and $q_k$ defined previously, induces a sequential locally convex and complete topology on $\mathfrak{d}$ satisfying: (1) any bounded set $B$ is indeed a bounded set in $\mathfrak{d}_k$ for some $k$, i.e., there exists a constant $C > 0$ such that $|u_i| \leq C$ if $i \leq k$ and $u_i = 0$ if $i > k$, for every $u = \{u_i : i \geq 1\}$ in $B$; (2) a sequence $\{u^{(n)}\}$ converges to $u$ if and only if there exist $k$ such that $u^{(n)}_i = 0$ for any $i > k$ and for every $n$, and $u^{(n)}_i \to u_i$ for every $i$.

One step further is the subspace $\mathfrak{s}$ of all rapidly deceasing sequences, i.e., $u = \{u_i : i \geq 1\}$ satisfying $i^k |u_i| \to 0$ for every $k = 0,1,\ldots$, which is a Fréchet space with the seminorms (weighted sup-norms) $p_k(u) = \sup_i \{i^k |u_i|\}$.
This space is of particular interest since it has a Hilbertian structure, i.e., we can easily check that the scalar or inner products \((u,v) = \sum_i i^{2k} u_i v_i\) yield norms \(\|p_k(u) = \sqrt{(u,u)_{k}}\) and the locally convex topologies defined by either \(\{p_k : k \geq 0\}\) or \(\{\bar{p}_k : k \geq 0\}\) are equivalent. Moreover, if \(h_k\) denotes the subspace of all sequences \(u\) satisfying \(\sqrt{(u,u)_{k}} < \infty\) then \(h_k\) is a Hilbert space and \(s = \bigcap_k h_k\).

To properly treat the spaces \(C^n_0(\Omega)\) or \(C^\infty_0(\Omega)\), of real (or complex) continuously differentially functions, either up to the order \(n\) or of any order, with compact support on an open domain \(\Omega\) of \(\mathbb{R}^d\), we need some notation for the derivatives. For a \(d\)-dimensional multi-index \(\alpha = (\alpha_1, \ldots, \alpha_d)\), \(\alpha_i\) nonnegative integers, of order \(\alpha = \alpha_1 + \cdots + \alpha_d\) and a function \(f\) of \(d\)-variables, we write \(\partial^\alpha f\) or \(D^\alpha f\) for the derivatives, namely

\[
\partial^\alpha f(x) = D^\alpha f(x) = \frac{\partial^{\mid\alpha\mid} f(x)}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}},
\]
e.g., \(\partial^{(2,0,0)} f\) indicates the second derivative in \(x_1\). Thus

\[
C^n_0(\Omega) = \{u \in C^n_0(\Omega) : \partial^\alpha u \in C^n_0(\Omega), \forall \alpha, |\alpha| \leq n\},
\]
and

\[
C^\infty_0(\Omega) = \{u \in C^n_0(\Omega) : \partial^\alpha u \in C^n_0(\Omega), \forall \alpha, \}\}
\]
with the seminorms

\[
p_\alpha(u) = \sup\{|\partial^\alpha u(x)| : x \in \Omega\}
\]
or, if preferred, the norms

\[
\|u\|_n(\alpha) = \sup\{|\partial^\alpha u(x)| : x \in \Omega, |\alpha| \leq n\} = \sup_{|\alpha| \leq n} p_\alpha(u). \quad (3.3)
\]

Similarly, for the spaces \(C^n_K(\Omega)\) or \(C^n_0(K)\), and \(C^\infty_K(\Omega)\) or \(C^\infty_0(K)\), of functions with support in \(K\) or vanishing on the boundary \(\partial K\) and extended by zero. It is clear that \(C^n_K(\Omega)\) is a Banach space with the norm \(\|\cdot\| = p_n(\cdot)\), while \(C^\infty_K(\Omega)\) is a Fréchet space with the countable family of seminorms either \(\{p_\alpha : \alpha\}\) or \(\{|\|u||_k : k = 0, 1, \ldots\}\), with the extra order property \(\|u||_k \leq \|u||_n\) if \(k \leq n\). Therefore, as in the case of \(C^n_0(\Omega)\), we endow \(C^n_K(\Omega)\) with sequential locally convex and complete topology satisfying: (1) any bounded set \(B\) in \(C^n_0(\Omega)\) is indeed a bounded set in \(C^n_0(K)\) for some compact set \(K \subset \Omega\), i.e., if \(\text{supp} v \subset K\) and \(\|v\|_n(\alpha) \leq C\), for every \(v \in B\) and some constant \(C\); (2) a sequence \(\{u_k\}\) converges to \(u\) in \(C^n_0(\Omega)\) if and only if \(\|u_k - u\|_n(\alpha) \to 0\) and there exist a compact \(K\) of \(\Omega\) such that \(\text{supp} u_k \subset K\) for every \(k\).

Analogously, the sequential locally convex and complete topology on \(C^\infty_0(\Omega)\) satisfies: (1) any bounded set \(B\) in \(C^\infty_0(\Omega)\) is indeed a bounded set in \(C^\infty_0(K)\) for some compact set \(K \subset \Omega\), i.e., if \(\text{supp} u_k \subset K\) and for every multi-index \(\alpha\) there exist a constant \(C_\alpha\) such that \(p_\alpha(u) \leq C_\alpha\), for every \(v \in B\); (2) a sequence
\{u_k\} converges to \(u\) in \(C_0^\infty(\Omega)\) if and only if there exist a compact \(K\) of \(\Omega\) such that \(\text{supp } u_k \subseteq K\) for every \(k\) and \(p_\alpha(u_k - u) \to 0\), for every multi-index \(\alpha\). This topological vector space \(C_0^\infty(\Omega)\) is usually denoted by \(D(\Omega)\) and their elements called test functions. Also, \(D_K(\Omega)\) denotes the space \(C_K^\infty(\Omega)\) with the inductive topology, which indeed is a Fréchet space, namely, all functions in \(D(\Omega)\) with support in a fixed compact set \(K\) of \(\Omega\).

In both cases, \(C_0^n(\Omega)\) and \(C_0^\infty(\Omega)\), besides the seminorms \(p_\alpha\), the uncountable family of seminorms obtained from

\[
q_{k,n}(u) = \sup\{|\partial^\alpha u(x)| : x \in \Omega \setminus K_k, |\alpha| \leq n\},
\]

\(\Omega = \bigcup_k K_k\), with \(K_k\), a compact subset of the interior of \(K_{k+1}\), yields the inductive limit topology. For instance, for any (unbounded monotone increasing) sequence \(\{r_k\}\) of positive numbers, the seminorm \(u \mapsto \sup_k r_k q_{k,n}(u)\) is continuous in \(C_0^n(\Omega)\), while \(u \mapsto \sup_k r_k q_{k,k}(u)\) is continuous in \(D(\Omega)\). Certainly, the norm \(|\cdot|_{(k)}\), as defined in (3.3), is continuous in \(C_0^n(\Omega)\) if \(k \leq n\) and in \(D(\Omega)\), for every \(k\).

On the other hand, the space \(C^\infty(\Omega)\) of real-valued (or complex-valued) continuously differentiable functions of any order becomes a Fréchet space with the countable family of seminorms

\[
p_{\alpha,k}(u) = \sup\{|\partial^\alpha u(x)| : x \in K_k\} = p_{\alpha,K_k}(u),
\]

where \(\Omega = \bigcup_k K_k\), with \(K_k\) a compact subset of the interior of \(K_{k+1}\) and \(\alpha\) is a \(d\)-dimensional multi-index. This space is commonly denoted by \(E(\Omega)\) and the locally convex topology satisfies: (1) a subset \(B\) is bounded in \(E(\Omega)\) if and only if for any multi-index \(\alpha\) and any compact subset \(K\) of \(\Omega\) there exists a constant \(C = C(\alpha, K)\) such that \(p_{\alpha,K}(v) \leq C\), for every \(v\) in \(B\), or equivalently, all derivative \(\partial^\alpha v\) are equibounded in any compact set \(K\); (2) a sequence \(\{u_k\}\) converges to \(u\) in \(E(\Omega)\) if and only if \(\partial^\alpha u_i \to \partial^\alpha u\) uniformly on any compact \(K\) of \(\Omega\), for every multi-index \(\alpha\) or equivalently, \(p_{\alpha,K}(u_i - u) \to 0\), for every compact \(K\) of \(\Omega\). The space \(C^n(\Omega)\), where only derivatives up to the order \(n\) are used, behaves very similar. Comparing with \(D(\Omega)\), note that for any multi-index \(\alpha\) and for any sequence \(\{r_k\}\) of positive numbers the expressions either \(u \mapsto \sup_k \{p_{\alpha,K_k}(u) + r_k p_{\alpha,\Omega \setminus K_k}(u)\}\) or \(u \mapsto p_{\alpha,\Omega}(u)\) is an example of a continuous seminorm in \(D(\Omega)\), which is not continuous in \(E(\Omega)\).

- **Remark 3.18.** For each compact \(K\) of \(\Omega\) the space \(D_K(\Omega)\) is a Fréchet space, and so, a barrel space of second category. Hence, the \(D(\Omega)\) is a barrel space of second category, see Remark 3.17. \(\square\)

**Exercise 3.6.** Similar to Exercise 2.1, discuss the spaces \(C_0^0(\Omega)\). \(\square\)

**Exercise 3.7.** Let \(Z_a\) be the space of (complex) entire functions \(f : \mathbb{C} \to \mathbb{C}\) of exponential type \(a > 0\), namely, for each \(k \geq 0\) there exists a constant \(C_p\) such that

\[
(1 + |z|)^k |f(z)| \leq C_k e^{a|y|}, \quad \forall z = x + iy \in \mathbb{C}.
\]
Consider the family of seminorms given by
\[ p_k(f) = \sup \{ e^{-a|y|}(1 + |z|)^k |f(z)| : z = x + iy \in \mathbb{C} \}, \]
and discuss the “inductive limit generated”, see Friedman [45, Section 2.3, pp 33–34].

We have seen the Fréchet space \( \mathcal{E}(\Omega) \) and the (sequentially) locally convex and complete topological vector space \( \mathcal{D}(\Omega) \). Now, we consider \( \mathcal{S} \) or \( \mathcal{S}(\mathbb{R}^d) \), the space of rapidly decreasing smooth functions on \( \mathbb{R}^d \), i.e., real-valued (or complex-valued) continuously differentiable functions \( u \) of any order such that \( |x|^k |\partial^\alpha u(x)| \to 0 \) as \( |x| \to \infty \), for every power \( k \geq 0 \) and any multi-index \( \alpha \). With the countable family of seminorms
\[
p_{n,k}(u) = \sup \{ |x|^k |\partial^\alpha u(x)| : x \in \mathbb{R}^d, |\alpha| \leq n, \quad n, k = 0, 1, \ldots, \quad (3.4)\]
\( \mathcal{S}(\mathbb{R}^d) \) becomes a Fréchet space. Therefore, (1) a subset \( B \) is bounded in \( \mathcal{S}(\mathbb{R}^d) \) if and only if for any integers \( n, k \geq 0 \) there exists a constant \( C = C(n,k) \) such that \( p_{n,k}(u) \leq C \), for every \( u \in B \); (2) a sequence \( \{u_i\} \) converges to \( u \) if and only if \( |x|^k \partial^\alpha u_i \to \partial^\alpha u \) uniformly on \( \mathbb{R}^d \), for every multi-index \( \alpha \) and any integer \( k \geq 0 \).

**Proposition 3.19.** The Hilbertian norms \( \bar{p}_{n,k}(u) = \sqrt{(u,u)_{n,k}} \) obtained from the scalar or inner products
\[
(u,v)_{n,k} = \sum_{|\alpha| \leq n} \int_{\mathbb{R}^d} (1 + |x|^2)^k \partial^\alpha u(x) \partial^\alpha v(x) \, dx, \quad n, k = 0, 1, \ldots
\]
yield the same locally convex topology in \( \mathcal{S} = \mathcal{S}(\mathbb{R}^d) \) as defined by \( \{p_{n,k} : n, k \geq 0\} \), i.e., \( u_i \to 0 \) in \( \mathcal{S} \) if and only if \( \bar{p}_{n,k}(u_i) \to 0 \) for every index \( n, k \).

**Proof.** First, it is clear that
\[
\bar{p}_{n,k}(u) = \left| \int_{\mathbb{R}^d} (1 + |x|^2)^{-m}(1 + |x|^2)^{k+m} |\partial^\alpha u(x)|^2 \, dx \right|^{1/2} \leq C_m p_{n,k+m}(u), \quad \forall u,
\]
for any choice of \( m > d/2 \).

To obtain the converse inequality, note that if \( f \) is a smooth function such that \( f \) all its derivatives vanish as \( |x| \to \infty \) then the equality
\[
f(x_1, \ldots, x_i, \ldots, x_d) = \int_{-\infty}^{x_i} \partial_i f(x_1, \ldots, t_i, \ldots, x_d) \, dt_i = \\
= \int_{-\infty}^{x_1} dt_1 \ldots \int_{-\infty}^{x_i} dt_i \ldots \int_{-\infty}^{x_d} \partial_1 \ldots \partial_d f(t_1, \ldots, t_i, \ldots, t_d) \, dt_d,
\]
yields the estimate
\[
\operatorname{ess-sup}_{x \in \mathbb{R}^d} |f(x)| \leq \int_{\mathbb{R}^d} |\partial^d f(x)| \, dx, \quad \text{with} \quad \partial^d = \partial_1 \ldots \partial_d.
\]
Hence, apply this inequality to the functions \( f(x) = (1 + |x|^2)^k |\partial^\alpha u(x)|^p \) to find a constant \( C_{n,k,d} \), depending only on \( n, k \) and the dimension \( d \), such that \( p_{n,k}(u) \leq C_{n,k,d} \bar{p}_{n+d,k}(u) \), for every \( u \). This completes the proof.

Note that the same argument of Proposition 3.19 can be applied to the family of norms

\[
\hat{p}_{n,k}(u) = \sum_{|\alpha| \leq n} \left[ \int_{\mathbb{R}^d} (1 + |x|^2)^k |\partial^\alpha u(x)|^p \, dx \right]^{1/p}, \quad n, k = 0, 1, \ldots ,
\]

for any given \( 1 \leq p \leq \infty \), i.e., they define the same locally convex topology in \( S(\mathbb{R}^d) \) as the initial family \( \{\hat{p}_{n,k} : n, k \geq 0\} \). Moreover, using the fact that for any compact set \( K \) in \( \Omega \) there exists an element \( \chi \) in \( D(\Omega) \) such that \( \chi = 1 \) on \( K \), and (after Leibniz’s rule) the estimate

\[
\sum_{|\alpha| \leq n} \int_{\mathbb{R}^d} |\partial^\alpha (\chi \varphi)|^p \, dx \leq C_\chi \sum_{|\alpha| \leq n} \int_{\text{supp} \chi} |\partial^\alpha \varphi|^p \, dx, \quad \forall \varphi \in C^\infty,
\]

for some constant \( C_\chi \) depending only on \( \chi, p \) and the dimension \( d \), with a fixed \( 1 \leq p \leq \infty \), we deduce that the uncountable family of seminorms obtained from seminorms of the type

\[
\hat{p}_{\alpha,K}(u) = \left[ \int_K |\partial^\alpha u(x)|^p \, dx \right]^{1/p},
\]

for any multi-index \( \alpha \) and any compact set \( K \) in \( \Omega \) could be used to define the same locally convex topology in either \( D(\Omega) \) or \( E(\Omega) \). This means that (a) \( u_i \rightarrow 0 \) in \( D(\Omega) \) if and only if (1) there exists a compact \( K \) in \( \Omega \) such that all supports of \( u_i \) are contained in \( K \) and (2) \( \hat{p}_{\alpha,K}(u_i) \rightarrow 0 \) for every index \( \alpha \) and compact \( K \) in \( \Omega \), while (b) \( u_i \rightarrow 0 \) in \( E(\Omega) \) if and only if \( \hat{p}_{\alpha,K}(u_i) \rightarrow 0 \) for every index \( \alpha \) and compact \( K \) in \( \Omega \).

Recall Lemma 3.4, a subset \( B \) of lctvs is bounded if and only if \( \{p(\varphi) : \forall \varphi \} \) is bounded in \( \mathbb{R} \), for any continuous seminorm \( p \), in particular, if \( B \) is bounded in \( D(\Omega) \) [or \( E(\Omega) \) or \( S(\mathbb{R}^d) \)] then so is the image \( \partial^\alpha B \). Thus, in view of Arzela-Ascoli Theorem 2.9, we deduce that a bounded subset \( B \) in any of the space \( D(\Omega) \) [or \( E(\Omega) \) or \( S(\mathbb{R}^d) \)] is pre-compact (i.e., has a compact closure). Therefore, the strong and the weak* topologies of the dual space \( D'(\Omega) \) [or \( E'(\Omega) \) or \( S'(\mathbb{R}^d) \)] coincides, i.e., for a sequence of distributions \( \{T_n\} \) such that \( \langle T_n, \varphi \rangle \rightarrow \langle T, \varphi \rangle \) for each \( \varphi \) it follows that

\[
\sup_{\varphi \in B} |\langle T_n, \varphi \rangle - \langle T, \varphi \rangle| \rightarrow 0,
\]

for any bounded subset \( B \) of \( D(\Omega) \) [or \( E(\Omega) \) or \( S(\mathbb{R}^d) \)].
yields the locally convex topology of the dual spaces. Since \( \varphi \mapsto |\langle T, \varphi \rangle| \) is a continuous seminorm, the expression \( |\langle T, \varphi \rangle| \) is bounded (in \( \mathbb{R} \)), for \( \varphi \) in \( B \) and \( T \) in \( B' \) if and only if \( B \) and \( B' \) are both bounded sets. This implies that the space \( D(\Omega) \) [or \( E(\Omega) \) or \( S(\mathbb{R}^d) \)], and its dual space \( D'(\Omega) \) [or \( E'(\Omega) \) or \( S'(\mathbb{R}^d) \)], are reflexive, i.e., each one is the dual of the other, e.g., see Schwartz [112, Theorem XIV, p. 75].

Recall the space \( L^1_{\text{loc}}(\Omega) \) of all Lebesgue locally integrable functions on \( \Omega \) becomes a Fréchet space with the sequence of seminorms
\[
p_k(u) = \int_{K_k} |u(x)| \, dx,
\]
where \( \Omega = \bigcup_k K_k \), with \( K_k \) a compact subset of the interior of \( K_{k+1} \). As seen later, the space \( L^1_{\text{loc}}(\Omega) \) can be considered included in the dual space \( D'(\Omega) \), i.e., any locally integrable function is a distribution. Moreover, from the examples below, it will be clear that any “reasonable” linear expression defined on the test functions \( D(\Omega) \) becomes a distribution, i.e., it is not obvious to construct a simple example of a linear functional on \( D(\Omega) \) which is not continuous, meaning, which does not belong to \( D'(\Omega) \).

### 3.2 Calculus with Distributions

As implicitly mentioned early, a (Schwartz) distribution in \( \Omega \subset \mathbb{R}^d \) is an element of \( D'(\Omega) \), i.e., a linear functional \( T \) on the test functions (continuously differentiable functions of any order with a compact support in \( \Omega \)) satisfying:
\[
\langle T, \varphi_n \rangle \to 0 \quad \text{for every sequence } \{\varphi_n : n \geq 1\} \text{ of test functions } \text{such that (a) } D^\alpha \varphi_n \to 0 \text{ uniformly } \forall \alpha, \text{ and (b) the supports of } \varphi_n \text{ are contained in a compact subset of } \Omega.
\]
Certainly, a natural zero-extension gives sense to the inclusion \( D'(\Omega) \subset D'(\mathbb{R}^d) \), which yields \( D'(\mathbb{R}^d) \subset D'(\Omega) \). On the contrary, a smooth function defined on \( \mathbb{R}^d \) can be considered as defined on \( \Omega \), i.e., \( E(\mathbb{R}^d) \subset E(\Omega) \), and if \( \Omega \) is bounded then \( S(\mathbb{R}^d) \subset E(\Omega) \), and in both cases, the converse inclusions hold for the dual spaces.

Our interest is on the dual space of the (sequentially) locally convex and complete topological vector space \( D(\Omega) \), and on the dual spaces of the Fréchet spaces \( E(\Omega) \) and \( S(\mathbb{R}^d) \). Similarly, the dual spaces of \( C^0_0(\Omega) \), \( C^n(\Omega) \), \( \mathbb{R}^n \), \( \mathcal{D}_k \), \( \mathcal{D} \) and \( \mathfrak{s} \) could be discussed. Note that
\[
D(\Omega) \subset E(\Omega) \quad \text{and} \quad D(\mathbb{R}^d) \subset S(\mathbb{R}^d) \subset E(\mathbb{R}^d),
\]
including the topology, e.g., a Cauchy sequence in \( D \) is also a Cauchy sequence in \( E \). Hence, the reverse inclusions hold for the dual spaces, namely
\[
E'(\Omega) \subset D'(\Omega) \quad \text{and} \quad E'(\mathbb{R}^d) \subset S'(\mathbb{R}^d) \subset D'(\mathbb{R}^d).
\]
A linear functional \( T \) belongs to \( S'(\mathbb{R}^d) \) if and only if there exist a constant \( C > 0 \) and indices \( n, k \) such that
\[
|\langle T, \varphi \rangle| \leq C \sup \{(1 + |x|^k)|\partial^\alpha \varphi(x)| : x \in \mathbb{R}^d, |\alpha| \leq n\}, \quad \forall \varphi \in S(\mathbb{R}^d),
\]
and a linear functional $T$ belongs to $E'(\Omega)$ if and only if there exist a compact set $K$ of $\Omega$, a constant $C > 0$ and an index $n$ such that

$$|\langle T, \varphi \rangle| \leq C \sup \{ |\partial^\alpha \varphi(x)| : x \in K, |\alpha| \leq n \}, \forall \varphi \in E(\Omega), \quad (3.5)$$

while $T$ belongs to $D'(\Omega)$ if and only if for every compact $K$ of $\Omega$ there exist a constant $C > 0$ and an index $n$ such that

$$|\langle T, \varphi \rangle| \leq C \sup \{ |\partial^\alpha \varphi(x)| : x \in K, |\alpha| \leq n \}, \forall \varphi \in D_K(\Omega),$$

where $D_K$ is the subspace of $D(\Omega)$ containing all functions with support inside $K$.

Note that a functional $T$ is continuous in $D(\Omega)$ if and only if the restriction $T|_K$ to $D_K$ is continuous, for every compact $K$ of $\Omega$. Moreover, in view of Remark 3.5, a linear operator from $D(\Omega)$ into a topological vector space is continuous if and only if it is bounded, i.e., it transforms bounded sets into bounded sets.

Therefore, the family of seminorms

$$p_{n,K}(\varphi) = \sup \{ |\partial^\alpha \varphi(x)| : x \in K, |\alpha| \leq n \}, \quad n = 0, 1, \ldots, \quad (3.6)$$

and $K$ a compact of $\Omega$, defines the locally convex topology in $D_K(\Omega)$ and in $E(\Omega)$, but only “determines” the inductive limits convergence in $D(\Omega)$. Anyway, the countable family of seminorms (3.4) defines the Fréchet topology in $S(\mathbb{R}^d)$.

As mentioned early, elements in $D(\Omega)$ are called test functions (or smooth functions with compact support) in $\Omega$, while elements in the dual space $D'(\Omega)$ are called distributions in $\Omega$. Similarly, elements in the dual space $S'(\mathbb{R}^d)$ are called tempered distributions in $\mathbb{R}^d$. In view of Banach-Steinhaus Theorem 2.17, if a sequence $\{T_n : n \geq 1\}$ of (tempered) distributions converges pointwise, i.e., the numerical sequence $\{(T_n, \varphi) : n \geq 1\}$ converges for every test function $\varphi$, then the limit is a (tempered) distribution.

Before going further, let us look at some examples of distributions:

- $L_{\text{loc}}^1(\Omega)$: Any locally integrable function in $\Omega$ can be considered as a distribution. Indeed, given $f$ in $L_{\text{loc}}^1(\Omega)$ define

$$\langle T_f, \varphi \rangle = \int_{\Omega} f(x)\varphi(x)dx, \quad \forall \varphi \in D(\Omega).$$

For any compact $K \subset \Omega$ we have

$$|\langle T_f, \varphi \rangle| \leq \left( \int_K |f(x)|dx \right) \sup \{ |\varphi(x)| : x \in K \}, \quad \forall \varphi \in D_K(\Omega),$$

and then $T_f$ is a continuous functional on $D(\Omega)$, actually, the expression of $T_f$ remains valid for any continuous function $\varphi$ with compact support, i.e., $T_f$ is also a continuous functional on $C_0^0(\Omega)$. On the other hand, if $T_f = T_g$ with $f, g$ in $L_{\text{loc}}^1(\Omega)$ then by taking $\varphi(x) = k_\varepsilon(y - x)$ we check that $\langle T_f, \varphi \rangle = (f * k_\varepsilon)(y)$, and $\varepsilon \to 0$, and the approximation by smooth functions arguments show that $f = g$ almost everywhere. This means that $f \mapsto T_f$ is injective, i.e., $L_{\text{loc}}^1$ can be identify with the subspace of distributions of the form $T_f$ as above.
• $C^0_0(\Omega)$: Any “signed” Radon measure on $\Omega$ is a distribution. Indeed, for any $\mu = \mu^+ - \mu^-$ we define
\[
\langle \mu, \varphi \rangle = \int_\Omega \varphi(x)\mu(dx), \quad \forall \varphi \in D(\Omega).
\]
Actually, this linear expression makes sense for any $\varphi$ in $C^0_0(\Omega)$ and for any compact $K \subset \Omega$ we have
\[
|\langle \mu, \varphi \rangle| \leq (\mu^+(K) + \mu^-(K)) \sup \{|\varphi(x)| : x \in K\}, \quad \forall \varphi \in D_K(\Omega),
\]
proving that $\mu$ is a continuous linear functional on $C^0_0(\Omega)$ and on $D(\Omega)$.

• Any function in $L^1_{\text{loc}}(\mathbb{R}^d)$ which slowly increases at infinity can be considered as a tempered distribution. Indeed, if $f$ slowly increases at infinity (called slowly increasing) then there is a compact set $K$ and a constant $r \geq 0$ such that
\[
|f(x)| \leq C(1 + |x|^r), \quad \forall x \in \mathbb{R}^d \setminus K, \text{ a.e.}
\]
Thus $T_f$ satisfies, with $k > d + r$,
\[
|\langle T_f, \varphi \rangle| \leq \left( \int_K |f(x)|dx \right) \sup \{|\varphi(x)| : x \in K\} + 
+ C \left( \int_{\mathbb{R}^d \setminus K} \frac{1 + |x|^r}{1 + |x|^k} dx \right) \sup \{(1 + |x|^k)|\varphi(x)| : x \in K\},
\]
for every $\varphi$ in $\mathcal{S}(\mathbb{R}^d)$, i.e., $T_f$ belongs to $\mathcal{S}'(\mathbb{R}^d)$. Also any derivative of a slowly increasing function is a tempered distribution, even if it corresponds to a not necessarily slowly increasing function, e.g., the function $f(x) = e^x \sin e^x$ is not slowly increasing, but it is the derivative of $F(x) = -\cos e^x$, and it can be considered a tempered distribution.

Even if the next two Exercises (i.e., 3.8 and 3.9) are practically discussed in the next section, it could be interesting to allow the reader to tackle these problems early.

Exercise 3.8. Verify that for $x$ in $\mathbb{R}$, the expression
\[
\langle \text{p.v.}(1/x), \varphi \rangle = \lim_{\varepsilon \to 0} \int_{|x| > \varepsilon} \frac{\varphi(x)}{x} dx = \int_0^\infty \frac{\varphi(x) - \varphi(-x)}{x} dx,
\]
defines a distribution in $\mathbb{R}$. Moreover, let $f$ be a continuous function $f$ in $\mathbb{R}^d \setminus \{0\}$ which is positively homogeneous of degree $-d$ and has mean zero on the unit sphere $\{x : |x| = 1\}$, i.e.,
\[
f(\lambda x) = \lambda^{-d} f(x), \quad \forall x \in \mathbb{R}^d, \lambda > 0 \quad \text{and} \quad \int_{|x|=1} f(x') dx' = 0,
\]
where \( dx' \) denotes the surface area measure on the unit sphere. Show that the expression

\[
\langle p.v.(f), \varphi \rangle = \lim_{\varepsilon \to 0} \int_{|x| > \varepsilon} f(x) \varphi(x) \, dx,
\]

defines a distribution in \( \mathbb{R}^d \). \( \square \)

**Exercise 3.9.** Consider the function \( x \mapsto \ln |x| \) as a distribution in \( \mathbb{R}^d \) and calculate its first order derivatives.

The differentiation \( \partial^\alpha \) can be considered as an operator from \( C^n_0(\Omega) \) into \( C^{n-|\alpha|}_0(\Omega) \), \( n \geq |\alpha| \) or from \( D(\Omega) \) into itself. Now, with the above notation, if \( f \) belongs to \( C^1(\Omega) \subset L^1_{loc}(\Omega) \) then an integration by parts shows that

\[
\int_{\Omega} [\partial^\alpha f(x)] \varphi(x) \, dx = (-1)^{|\alpha|} \int_{\Omega} f(x) [\partial^\alpha \varphi(x)] \, dx, \quad \forall \varphi \in D(\Omega).
\]

Hence, by transposition, we can consider the differential operator \( \partial^\alpha \) acting on distribution, i.e., if \( T \) is a distribution then

\[
\langle \partial^\alpha T, \varphi \rangle = (-1)^{|\alpha|} \langle T, \partial^\alpha \varphi \rangle, \quad \forall \varphi \in D(\Omega).
\]

Thus \( \partial^\alpha \) is a continuous linear operator from \( \mathcal{D}'(\Omega) \) into itself, and any derivative of the previous examples is also a distribution.

Similarly, if \( f \) belongs to \( C^\infty(\Omega) \) then for any distribution \( T \) we can define the distribution \( fT \) by the expression

\[
\langle fT, \varphi \rangle = \langle T, f\varphi \rangle, \quad \forall \varphi \in D(\Omega),
\]

i.e., the multiplication by a \( C^\infty \)-function is a continuous operation on \( D(\Omega) \).

Actually most of the operations defined on \( D(\Omega) \) can be extended to \( \mathcal{D}'(\Omega) \). For instance, if we want to extend a linear operation on real-valued (or complex-valued) functions defined on \( \Omega \) then (1) we cast the operation as a linear mapping \( Q \) defined on \( D(\Omega) \) with values in \( L^1_{loc}(\Omega_1) \) and (2) we transpose \( Q \) (via some integration-by-parts formula) regarded as a distribution, i.e., we look for a linear mapping \( Q^* \) from \( D(\Omega_1) \) into \( L^1_{loc}(\Omega) \) such that

\[
\int_{\Omega} \varphi(x) (Q^* \psi)(x) \, dx = \int_{\Omega_1} \psi(y) (Q \varphi)(y) \, dy, \quad \forall \psi \in D(\Omega_1), \varphi \in D(\Omega),
\]

to check if \( Q^* \psi \) is indeed an element of \( D(\Omega) \) (or at least, of some \( C^n_0(\Omega) \) with \( n \geq 0 \)). If so, (3) we can extend the definition of \( Q \) to another operator \( Q \) defined on \( \mathcal{D}'(\Omega) \) (or on the dual of \( C^n_0(\Omega) \) with the same \( n \geq 0 \)) as

\[
\langle QT, \psi \rangle = \langle T, Q^* \psi \rangle, \quad \forall T \in \mathcal{D}'(\Omega), \psi \in D(\Omega_1).
\]

with values in \( \mathcal{D}'(\Omega_1) \). Clearly, the point is that if \( \varphi \) is interpreted as the distribution \( T_\varphi \) then we have \( T_{Q \varphi} = QT_\varphi \). Certainly, we have to check the
continuity of the linear operator $Q^*$ to make sure that $Q$ is well defined. Hence, if the linear mapping $Q^*$ is continuous, as an operator from $\mathcal{D}(\Omega_1)$ into either $\mathcal{D}(\Omega)$ or $C^n(\Omega)$, then the linear operation $Q$ admits an extension $\overline{Q}$ as a linear continuous operator from either $\mathcal{D}'(\Omega)$ or the dual space of $C^n(\Omega)$ into $\mathcal{D}'(\Omega_1)$. Similar statements hold if the space $C^n(\Omega)$ is replaced by either $C^n_0(\Omega)$ or $S(\Omega)$.

For instance, in the case of the derivative, we took $Q = \partial^\alpha$, and $\Omega_1 = \Omega$ to check that $Q^* = (-1)^{|\alpha|} \partial^{\alpha}$, which defines $Q$ as desired. Another typical example is a diffeomorphism $\vartheta : \Omega_1 \rightarrow \Omega$ and $Q\varphi(y) = (\varphi \circ \vartheta)(y) = \varphi(\vartheta(y))$. We calculate “transposed” expression

$$
\int_{\Omega_1} \psi(y)\varphi(\vartheta(y)) \, dy = \int_{\Omega} \psi(\vartheta^{-1}(x))\varphi(x) J_\vartheta(x) \, dx,
$$

where $J_\vartheta(x) = |\det(\partial \vartheta^{-1}(x)/\partial x)|$ is the Jacobian of $\vartheta$. Thus $(Q^*\psi)(x) = \psi(\vartheta^{-1}(x))J_\vartheta(x)$ and

$$
\langle QT, \psi \rangle = \langle T, (\varphi \circ \vartheta^{-1})J_\vartheta \rangle, \quad \forall T \in \mathcal{D}'(\Omega), \, \psi \in \mathcal{D}(\Omega_1),
$$

i.e., from the known expression $\varphi \circ \vartheta$ we deduce $QT = T \circ \vartheta$. Certainly, we need some assumptions for this to hold, e.g., the diffeomorphism $\vartheta$ should be of class $C^\infty$ (or at least, of class $C^n$).

**Exercise 3.10.** Discuss (a) the translation operator $\tau_h$ defined as $\tau_h \varphi(x) = \varphi(x + h)$ with $\Omega_h = h + \Omega$ and (b) the reflection operator $\tilde{\varphi}(x) = \varphi(-x)$. □

**Exercise 3.11.** For a unit vector $e$ in $\mathbb{R}^d$, consider the expression $\Lambda_{e,t}\varphi(x) = [\varphi(x + te) - \varphi(x)]/t$, for $t > 0$. Discuss (a) the directional rate operator $\Lambda_{e,t}$ as defined on either $\mathcal{D}(\mathbb{R}^d)$ or $\mathcal{D}(\Omega)$, and (b) extend the definition of $\Lambda_{e,t}$ as a linear continuous operator on the spaces of distributions, i.e., on $\mathcal{D}$, $\mathcal{E}$ and $\mathcal{S}$. Moreover, also discuss (c) the iteration $\Lambda_{e,t}\Lambda_{e,-t}$ written as $\Lambda^2_{e,t}\varphi(x) = \varphi(x + et) + \varphi(x - et) - 2\varphi(x)/t^2$, and then (d) consider the continuity of the directional derivative $\lim_{t \to 0} \Lambda_{e,t}$ and the Hessian $\lim_{t \to 0} \Lambda^2_{e,t}$ as operator acting on distributions. □

### 3.2.1 Positivity, Differentiability and Integrability

A distribution $T$ is called nonnegative (or positive) on an open subset $U$ of $\Omega$ if $T \geq 0$ on $U$, i.e., $\langle T, \varphi \rangle \geq 0$ for every $\varphi \in \mathcal{D}(\Omega)$ with $\text{supp}(T) \subset U$ and $\varphi \geq 0$. To say that a distribution $T \geq 0$ on $\Omega$ we write only $T \geq 0$. Moreover, if $T$ and $S$ are two distributions in $\Omega$ then $T \leq S$ means $S - T \geq 0$ or equivalently $\langle T, \varphi \rangle \leq \langle S, \varphi \rangle$, for every $\varphi \in \mathcal{D}(\Omega)$ satisfying $\varphi \geq 0$.

Therefore, if $f$ and $g$ are functions in $L^1_{\text{loc}}(\Omega)$ and $T$ is a distribution on $\Omega$ then the expression $f \leq T \leq g$ actually means that the distribution $T_f$ and $T_g$ corresponding to $f$ and $g$ satisfy $T_g - T \geq 0$ and $T - T_f \geq 0$, in short,

$$
\int_{\Omega} f(x)\varphi(x) \, dx \leq \langle T, \varphi \rangle \leq \int_{\Omega} g(x)\varphi(x) \, dx,
$$

for every $\varphi \geq 0$ in $\varphi \in \mathcal{D}(\Omega)$.
Proposition 3.20. If $T$ is a nonnegative distribution in $\Omega$ then $T$ is actually a Radon measure. Moreover, if $f \leq T \leq g$ for some locally integrable functions $f$ and $g$ then $T$ is actually a locally integrable function, i.e., $T = T_h$ for some element $h$ in $L^1_{\text{loc}}(\Omega)$.

Proof. First, we need to show that $T$ is continuous on $\mathcal{D}(\Omega)$ with the topology of $C_0^0(\Omega)$, i.e., if a sequence $\{\varphi_k\} \subset \mathcal{D}(\Omega)$ satisfies $\varphi_k \to 0$ uniformly and supp$(\varphi_k) \subset K$ for some compact $K$ of $\Omega$, then $\langle T, \varphi_k \rangle \to 0$. To this purpose, choose $\chi \geq 0$ in $\mathcal{D}(\Omega)$ such that $\chi = 1$ on $K$ and define $\varepsilon_k = \sup \{ |\varphi_k(x)| : x \in \Omega \}$, which satisfies $\varepsilon_k \to 0$. Since $T \geq 0$, from $\varepsilon_k \chi \pm \varphi_k \geq 0$ we obtain $\varepsilon_k \langle T, \chi \rangle \pm \langle T, \varphi_k \rangle \geq 0$, which yields $|\langle T, \varphi_k \rangle| \to 0$.

Next, the distribution $T$ can be extended uniquely to a continuous linear functional on $C_0^0(\Omega)$, which is a lattice and Stone-Daniell Proposition 1.26 shows that $T$ is indeed a Radon measure.

Now, by means of the previous argument, if $f \leq T \leq g$ then $T - T_f$ and $T_g - T$ are measures, and $(T - T_f) + (T_g - T)$ is the measure $T_{g-f}$, which is absolutely continuous with respect to the Lebesgue measure. Therefore, both $T - T_f$ and $T_g - T$ are absolutely continuous, i.e., if $(\cdot)'$ denotes the Radon-Nikodym (see part I) derivative with respect to the Lebesgue measure then $T = T_h$, where either $h = (T - T_f)' + f$ or $h = g - (T_g - T)'$. □

Exercise 3.12. Let $f$ be a real-valued function defined on a convex open set $\Omega$ of $\mathbb{R}^d$. Recall that $f$ is called convex whenever $f(sx + ty) \leq sf(x) + tf(y)$, for every $x, y \in \Omega$ and any $s, t \geq 0, s + t = 1$. Also, $f$ is called concave if $-f$ is convex. Assuming that $f$ is twice continuously differentiable, (a) prove that $f$ is convex if and only if the Hessian matrix $D^2f$ is nonnegative definite, i.e., $(v, D^2 f(x)v) \geq 0$ for every $v \in \mathbb{R}^d$ and any $x \in \Omega$. Now, a function $f$ is called semi-convex (or semi-concave) if there exists a twice continuously differentiable $g$ such that $f + g$ is convex (or concave). Prove that (b) if locally integrable function $f$ is semi-convex and also semi-concave then the Hessian matrix $D^2f$, regarded as a matrix-valued distribution, is actually a locally bounded matrix-valued function. □

An integrable function $f$ can be considered uniquely as a distribution $T_f$ and $\langle T_f, 1 \rangle$ is the integral of $f$ over $\Omega$, so in general, if $T$ is a distribution with compact support, i.e., $T$ in $\mathcal{E}'(\Omega)$, we can consider $\langle T, 1 \rangle$ as the integral of $T$ over $\Omega$.

As mentioned early, elements in $L^1_{\text{loc}}(\Omega)$ are uniquely considered as distributions, and the differential operator $\partial^\alpha$ is a continuous operation on $\mathcal{D}(\Omega)$, for any multi-index $\alpha$. Thus the the derivatives $\partial^\alpha f$ of a locally integrable function makes sense as a distribution. Now, suppose that the distribution $\partial^\alpha f$ is indeed a locally integrable function then how this relates to the usual pointwise derivative? We have some answers, e.g., essentially by definition, if $f(x)$ has a continuous partial derivative $\partial_i f$ then the integration by part formula shows that $\partial_i f$ agree with the derivative in the distribution sense. However, as soon as the continuity of the derivative is relaxed, the two meanings may disagree. For instance, we may have a function with derivative in any point, and yet,
the derivative function may not be locally integrable; moreover, we may have a pointwise derivative locally integrable (defined almost everywhere) which is different from the derivative in the distribution sense.

First we discuss a three typical examples in \( \mathcal{D}(\mathbb{R}) \):

1.- Jumps: Perhaps the simplest jump is given by the Heaviside’s function \( h(x) = 0 \) for \( x < 0 \) and \( h(x) = 1 \) for \( x > 0 \), which is continuous (infinite differentiable) for any \( x \neq 0 \) and both lateral limits exist and are finite at the discontinuous point \( x = 0 \). As a distribution,

\[
\langle h, \varphi \rangle = \int_0^\infty \varphi(x) \, dx, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}),
\]

is of order 0, i.e., an element of the dual space of \( C^0_0(\mathbb{R}) \). Its first derivative,

\[
\langle h', \varphi \rangle = -\int_0^\infty \varphi'(x) \, dx = \varphi(0), \quad \forall \varphi \in \mathcal{D}(\mathbb{R}),
\]

is an element of \( C^0(\mathbb{R}) \), i.e., a distribution of order 0 with compact support, actually, its support is \( \{0\} \) and \( h' \) is the Dirac measure.

2.- Finite Parts: The function \( h(x) = 0 \) for \( x < 0 \) and \( h(x) = x^{-1/2} \) for \( x > 0 \) is also infinite differentiable for any \( x \neq 0 \), but its pointwise derivative is not locally integrable at 0. Its first derivative in the distribution sense satisfies

\[
\langle h', \varphi \rangle = -\int_0^\infty \varphi'(x) x^{-1/2} \, dx = -\lim_{\varepsilon \to 0} \int_{\varepsilon}^\infty \varphi'(x) x^{-1/2} \, dx = \lim_{\varepsilon \to 0} \left\{ \int_{\varepsilon}^\infty \varphi(x) \left( -\frac{1}{2} x^{-3/2} \right) \, dx + \varphi(0) \varepsilon^{-1/2} \right\}, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}),
\]

after using an integration by parts and the asymptotic \( \varphi(\varepsilon) = \varphi(0) + O(\varepsilon) \). The expression of the derivative is commonly known as the Hadamard finite part.

3.- Singular Integrals: The function \( h : x \mapsto x^{-1} \) is not locally integrable at 0, but the expression

\[
\langle h, \varphi \rangle = \lim_{\varepsilon \to 0} \int_{\{|x| \geq \varepsilon\}} \varphi(x) x^{-1} \, dx = \int_{\mathbb{R}} \frac{\varphi(x) - \varphi(0)}{x} \, dx, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}),
\]

is a distribution of order 1, since \( (1/|x|) \, dx \) is not a Radon measure. However, the above expression is finite if \( \varphi \) is merely Hölder continuous at the point 0. Similarly, it may be instructive to verify that the function \( k(t, x) = \mathbb{1}_{\{t > 0\}} e^{-x^2/2t} / \sqrt{2t\pi} \) satisfies the partial differential equation \( \partial_t - \partial_x^2 k = \delta \) in the sense of distributions, where \( \delta \) is the Dirac measure in \( \mathbb{R}^2 \), i.e., \( \langle \delta, \varphi \rangle = \varphi(0, 0) \), for every \( \varphi = \varphi(t, x) \).

By definition, the derivative of a distribution is another distribution, but the converse needs some discussion. Given an element \( S \) in \( \mathcal{D}(\mathbb{R}) \) we are interested
in finding antiderivatives of $S$, i.e., distributions $T$ such that $T' = S$. If $D_0(\mathbb{R})$ denotes the elements in $D(\mathbb{R})$ with zero mean, i.e.,

$$\chi \in D_0(\mathbb{R}) \quad \text{if and only if} \quad \chi \in D(\mathbb{R}) \quad \text{and} \quad \int_{-\infty}^{+\infty} \chi(x) \, dx = 0,$$

then any element in $D_0(\mathbb{R})$ has an anti-derivative in $D(\mathbb{R})$, namely,

$$\phi(x) = \int_{-\infty}^{x} \chi(x) \, dx.$$

Therefore, if we select a test function $\varphi_0$ such that

$$\int_{-\infty}^{+\infty} \varphi_0(x) \, dx = 1,$$

then any distribution $\varphi$ can be (uniquely) written as $\varphi = \lambda \varphi_0 + \chi$, where $\varphi_0$ belongs to $D(\mathbb{R})$, and $\chi = \phi'$ belongs to $D_0(\mathbb{R})$, and the constant $\lambda$ satisfies

$$\lambda = \langle 1, \varphi \rangle = \int_{-\infty}^{+\infty} \varphi(x) \, dx.$$

Hence, the equation $T' = S$ implies

$$\langle T, \varphi \rangle = \langle T, \lambda \varphi_0 \rangle + \langle T, \chi \rangle = \lambda \langle T, \varphi_0 \rangle - \langle T', \phi \rangle,$$

i.e.,

$$\langle T, \varphi \rangle = \langle T, \varphi_0 \rangle \langle 1, \varphi \rangle + \langle S, \int_{-\infty}^{x} [\langle 1, \varphi \rangle \varphi_0(x) - \varphi(x)] \, dx \rangle,$$

which can be used to define $T$ as an element of $D(\mathbb{R})$ when $T' = S$ is given. On the other hand, if $T_1$ and $T_2$ are two distributions satisfying $T_1' = T_2' = S$ then

$$\langle T_1 - T_2, \varphi \rangle = \langle T_1 - T_2, \varphi_0 \rangle \langle 1, \varphi \rangle, \quad \forall \varphi \in D(\mathbb{R}),$$

i.e., $T_1 - T_2$ is the distribution associated with the constant function $C = \langle T_1 - T_2, \varphi_0 \rangle$, recall that $\varphi_0$ is a fixed test function.

As a consequence of the about result in $D(\mathbb{R})$ we have:

\begin{itemize}
  \item[a.-] Any distribution admits infinity many antiderivatives of order $n$, and any two of them differ in a polynomial of degree at most $n - 1$.
  \item[b.-] The derivative of a distribution is a locally finite signed measure if and only if the distribution is indeed a function of bounded variation on any bounded interval. Indeed, if a distribution $T = f$ a function of bounded variation on any bounded interval then the integration by parts shows that

$$\langle T', \varphi \rangle = -\int_{-\infty}^{+\infty} f(x) \varphi'(x) \, dx = \int_{-\infty}^{+\infty} \varphi(x) \, df(x),$$

\end{itemize}
i.e., $T'$ is the measure induced by $f$. On the other hand, let $T$ be such that $T' = \mu$ is a measure. For a fixed point $a$, a suitable extension of the function $f(x) = \mu([a, x])$ is cad-lag (namely, right-continuous having left-hand limits) and has bounded variation in any bounded interval interval, and in view of the above $f$ is an antiderivative of $\mu$. Therefore, $T - f$ is a constant, i.e., $T$ is a function of bounded variation on any bounded interval.

c.- Based on the previous assertion, we have: (1) a distribution has a nonnegative derivative if and only if the distribution is indeed an increasing function; (2) a distribution has a nonnegative second derivative if and only if the distribution is indeed a convex function; (3) a function is the difference of two convex functions if and only if its second derivative (as a distribution) is a locally finite signed measure.

d.- Let $g$ be a locally integrable function and for some $a$, define
$$f(x) = f(a) + \int_a^x g(x) \, dx, \quad \forall x \in \mathbb{R}.$$ Then by means of the Lebesgue theory we show that $f$ is a continuous function with derivative $g$ almost everywhere in the usual sense. The above arguments on antiderivative show that $f$ has the above representation if and only if the derivative in the distribution sense of function $f$ is actually $g$, i.e., the distribution identified by the locally integrable function $g$.

e.- If the derivative of order $n$ of a distribution is locally finite signed measures then the distribution is indeed a function infinitely differentiable. Indeed, if the derivative of order $n + 2$ of a distribution is a locally finite signed measure then the derivative of order $n$ is a continuous function.

**Exercise 3.13.** Give more detail on assertion (e) above, namely, use Proposition 3.20 to show that for any open interval $I$ in $\mathbb{R}$ and any element $T$ in $\mathcal{D}'(I)$ we have (1) if $T' \geq 0$ then $T = T_f$ is the distribution associated to some increasing function $f$; (2) if $T'' \geq 0$ then $T = T_f$ is the distribution associated to some convex function $f$. Moreover, (3) if $T'$ is a signed Radon measure on any compact sub-interval of $I$ then $T = T_f$ is the distribution associated to some function $f$ with bounded variation on every compact sub-interval of $I$ (i.e., $f$ has locally bounded variation on the open interval $I$); and finally (4) if $T''$ is a signed Radon measure on any compact sub-interval of $I$ then $T = T_f$ is the distribution associated to some function $f$ which is a difference of two convex functions.

A mapping $z \mapsto T_z$ from some open set $Z$ of the Complex plane $\mathbb{C}$ into (complex) $\mathcal{D}'(\Omega)$ is called analytic if for every $\varphi$ in $\mathcal{D}'(\Omega)$ the (possible complex valued) function $z \mapsto \langle T_z, \varphi \rangle$ is analytic in $Z$. Now, the functions $z \mapsto |x|^z$, $z \mapsto (x^+)^z$ and $z \mapsto (x^-)^z$ define analytic distribution for $\Re(z) > -1$. Indeed, it suffices to remark that the expressions $|x|^z = \exp(z \ln |x|)$, $(x^+)^z = \exp(z \ln(\max\{x, 0\}))$ and $(x^-)^z = \exp(z \ln(-\min\{x, 0\}))$ are locally integrable (in $x$) when $\Re(z) > -1$. Therefore, we can extend the definition of $|x|^z$, $(x^+)^z$, $(x^-)^z$
and \((x^-)^z\) as a distribution beyond \(\mathcal{R}(z) > -1\) by continuing the functions \(\langle |x|^z, \varphi \rangle, \langle (x^+)^z, \varphi \rangle\) and \(\langle (x^+)^z, \varphi \rangle\) analytically, for every \(\varphi\) in \(\mathcal{D}(\mathbb{R})\), to a larger connected subset of the complex \(z\)-plane. For instance, the interested reader may take a look at Grafakos \[57, \text{Section } 2.4.3, \text{pp. } 127–133\].

**Exercise 3.14.** For the powers distributions \(\langle |x|^z, \varphi \rangle, \langle (x^+)^z, \varphi \rangle, \langle (x^+)^z, \varphi \rangle\) obtained from a locally integrable function. A distribution "independent of \(x\)" is understood as a distribution beyond \(\mathcal{D}'(\mathbb{R})\) then the translation in \(h\) is defined first on locally integrable functions as \(\tau_h \varphi(x) = \varphi(x + h)\) and extended by duality to \(\mathcal{D}'(\mathbb{R}^d)\) by the formula \(\langle T, \varphi \rangle = \langle T, 

\tau_{-h} \varphi \rangle\), so that the expression agree on distributions obtained from a locally integrable function. A distribution “independent of \(x_i\)” is understood as a distribution “invariant under translation in \(x_i\),” e.g., \(T\) is independent of \(x_1\) if and only if \(\tau_h T = T\), for any \(h = (h_1, 0, \ldots, 0)\), which is equivalent to: a distribution \(T\) is independent of \(x_i\) if and only if \(\partial_i T = 0\).

**Proposition 3.21.** For any given distribution \(S_1\) in \(\mathcal{D}'(\mathbb{R}^d)\), the partial differential equation \(\partial_1 T = S_1\) admits infinity many solutions and any two of them differ in a distribution independent of \(x_1\).

**Proof.** Essentially, we repeat the arguments on antiderivatives for one variable, but now, the subspace \(\mathcal{D}_1(\mathbb{R}^d)\) of all test functions of the form \(\partial_1 \phi\) for some \(\phi\) in \(\mathcal{D}(\mathbb{R}^d)\) is not an hyperplane, since this requires

\[
\int_{-\infty}^{+\infty} \varphi(x_1, x_2, \ldots, x_d) \, dx_1 = 0, \quad \forall x_2, \ldots, x_d.
\]

Nevertheless, if \(\varphi_0(x_1)\) is a test function in the variable \(x_1\) with integral equal to 1 then any element in \(\varphi\) of \(\mathcal{D}(\mathbb{R}^d)\) can be written uniquely as

\[
\varphi(x_1, x_2, \ldots, x_d) = \varphi_0(x_1) I_\varphi(x_2, \ldots, x_d) + \partial_1 J_\varphi(x_1, x_2, \ldots, x_d),
\]

\[
I_\varphi(x_2, \ldots, x_d) = \int_{-\infty}^{+\infty} \varphi(x_1, x_2, \ldots, x_d) \, dx_1, \quad \text{and}
\]

\[
J(x_1, x_2, \ldots, x_d) = \int_{-\infty}^{x_1} \left[ \varphi(r, x_2, \ldots, x_d) - \varphi_0(r) I_\varphi(x_2, \ldots, x_d) \right] \, dr,
\]

where \(J_\varphi\) belongs to \(\mathcal{D}(\mathbb{R}^d)\) and therefore \(\partial_1 J_\varphi\) belongs to \(\mathcal{D}_1(\mathbb{R}^d)\). Thus, any distribution satisfying \(\partial_1 T = S_1\) can be written as

\[
\langle T, \varphi \rangle = \langle T, \varphi_0 I_\varphi \rangle - \langle S_1, J_\varphi \rangle, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d),
\]

Since the operation \(\varphi \mapsto I_\varphi\) is linear, continuous and surjective from \(\mathcal{D}(\mathbb{R}^d)\) onto \(\mathcal{D}(\mathbb{R}^{d-1})\), it is clear that \(R : \varphi \mapsto \langle T, \varphi_0 I_\varphi \rangle\) is an element in \(\mathcal{D}'(\mathbb{R}^d)\), which is invariant under translation of the form \(h = (h_1, 0, \ldots, 0)\), i.e., \(\partial_1 S = 0\). Similarly, the mapping \(\tilde{S}_1 : \varphi \mapsto \langle S_1, J_\varphi \rangle\) is an element in \(\mathcal{D}'(\mathbb{R}^d)\). Hence, \(T = R - \tilde{S}_1\) provides a solution to the desired PDE, and any two of them differ in a distribution independent of \(x_1\). \(\square\)
More analysis is needed to the following

**Proposition 3.22.** For any given $s S_i$, $i = 1, \ldots, k \leq d$, in $D(\mathbb{R}^d)$, satisfying the compatibility conditions $\partial_j S_i = \partial_i S_j$, the partial differential system of equations $\partial_i T = S_i$, $i = 1, \ldots, k$, admits infinity many solutions and any two of them differ in a distribution independent of $x_i$, $i = 1, 2, \ldots, k$. As a consequence, if every derivative of order $n$ of a distribution $T$ are zero then the distribution is a polynomial of degree at most $n - 1$.

**Proof.** The arguments are similar to those used in Proposition 3.21, so that only the part concerning the compatibility conditions is discussed.

Indeed, to see where these compatibility conditions intervene, note that Proposition 3.21 means that any two solutions to the first equation $\partial_1 T = S_1$ differ in a distribution in last $d - 1$ variables, i.e., in $D'(\mathbb{R}^{d-1})$. Therefore, if $T_1$ is a particular solution to the system of equations then any other solution $T$ must be of the form $T = T_1 + R_1$, where $R_1$ is a distribution in $d - 1$ last variables, i.e., in $D'(\mathbb{R}^{d-1})$. Hence, the second equation is written as $S_2 = \partial_2 T_1 + \partial_2 R_1$. The compatibility conditions for $S_1$ and $S_2$ imply that

$$\partial_1 (S_2 - \partial_2 T_1) = \partial_2 (S_1 - \partial_1 T_1) = 0,$$

which means that $S_2 - \partial_2 T_1$ is independent of the variable $x_1$. Applied Proposition 3.21, there exists a distribution $T_{1,2}$ in the last $d - 1$ variables $x_2, \ldots, x_d$ such that $S_2 - \partial_2 T_1 = T_{1,2}$. Hence, the unknown element $R_1$ in $D'(\mathbb{R}^{d-1})$ must solve the equation $\partial_2 R_1 = T_{1,2}$.

Therefore, the $k$ equations have effectively reduced to $k - 1$ equations, and at each step, Proposition 3.21 can be used. Hence, if $T_1$ is chosen to be a constant then this argument provides a solution to the system of equations, and the fact that any two solutions differ in the last $d - k$ variables, or in a constant if $k = d$.

An iterating the assertion that the constant distributions are the only elements in $D'(\mathbb{R}^d)$ that are either invariant under any translation $\tau_h$, with $h$ in $\mathbb{R}^d$, or have all its first order derivatives equal to zero, shows that if every derivative of order $n$ of a distribution $T$ are zero then the distribution is a polynomial of degree at most $n - 1$. $\square$

• **Remark 3.23.** If in Proposition 3.22 with $k = d$ the given distributions $S_i$ are indeed continuous functions $g_1, \ldots, g_d$ then any distribution solution $T$ can be identified with a continuously differentiable function $f$ satisfying

$$f(x) = f(0) + \int_0^x [g_1(y)dy_1 + \cdots + g_d(y)dy_d],$$

where the integration is over any rectifiable curve joining the points 0 and $x$. If the data $g_1, \ldots, g_d$ are not necessarily continuous then this formula may still be valid, but in any case, the compatibility conditions $\partial_j g_i = \partial_i g_i$, i.e.,

$$\int_{\mathbb{R}^d} \left( g_i(x)\partial_j \varphi(x) - g_j(x)\partial_i \varphi(x) \right)dx = 0, \quad \forall \varphi \in D(\mathbb{R}^d).$$
3.2. Calculus with Distributions

are always enforced in the distribution sense. However, if \( g_1, \ldots, g_d \) are locally integrable functions satisfying the compatibility conditions in the distribution sense then Proposition 3.22 does not affirms that there is a distribution solution \( T \) which can be identified with a locally integrable function.

- **Remark 3.24.** A locally integrable function \( f \) is called absolutely continuous in the variable \( x_1 \) on almost every parallel to the \( x_1 \)-axis if the relation

\[
 f(x_1, x_2, \ldots, x_d) - f(y_1, x_2, \ldots, x_d) = \int_{y_1}^{x_1} g_1(r_1, x_2, \ldots, x_d) \, dr_1,
\]

holds true for almost every point \( x = (x_1, x_2, \ldots, x_d) \) and \( y = (y_1, x_2, \ldots, x_d) \) in \( \mathbb{R}^d \), and for some locally integrable function \( g \), which necessarily is equal to the usual \( x_1 \)-partial derivative of \( f \). The point is that a locally integrable function \( f \) is absolutely continuous in the variable \( x_1 \) on almost every parallel to the \( x_1 \)-axis if and only if \( g_1 = \partial_1 f \) in the sense of distributions.

In general, the concept of absolutely continuity in several variables could be expressed by require that a function either \( f \) maps set of Lebesgue measure zero into sets of Lebeague measure zero (so that \( f \) has a Radon-Nikodym derivative, ) or that its first partial derivatives (in the sense of distributions) be locally integrable functions, and clearly, these to concepts are not equivalent if the dimension \( d \geq 2 \). Similar arguments could discussed for bounded variation functions of several variables, but in general, a bounded variation function is understood as a function which has signed measures as its first partial derivatives in the sense of distributions.

We conclude this section with the following

**Definition 3.25 (Weak Derivative).** A locally integrable function \( f \) in \( \Omega \subset \mathbb{R}^d \) has an element \( \partial^\alpha f \) in \( L^1_{\text{loc}}(\Omega) \) as its *weak derivative* if

\[
 \int_{\Omega} \partial^\alpha f(x) \varphi(x) \, dx = (-1)^{|\alpha|} \int_{\Omega} f(x) \partial^\alpha \varphi(x) \, dx, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d).
\]

Naturally, this is also referred to as the derivative in the *distribution sense*.

Similarly, a function \( f \) in \( L^p(\Omega), 1 \leq p \leq \infty \), has a weak derivative in \( L^q(\Omega), 1 \leq q \leq \infty \), if \( \partial^\alpha f \) belongs to \( L^q(\mathbb{R}^d) \).

We have seen above that two distributions with a common derivative differ in a constant, so in particular, if the weak derivatives \( \partial^\alpha f = 0 \) for every multi-index of order \( |\alpha| = 1 \) then the locally integrable function \( f \) is actually a constant. This is a fundamental contrast with the almost everywhere (pointwise) derivative. Nevertheless, continuously (pointwise) partial derivatives agrees with weak (or distribution sense) derivatives.

- **Remark 3.26.** Using the density of \( \mathcal{D}(\Omega) \) in \( L^p(\Omega) \), it is a good exercise to verify that the weak (directional) derivative \( \partial e f \) of a locally integrable function \( f \) (or in \( L^p \)) exists if and only the function \( \delta^e_t f : x \mapsto [f(x+te) - f(x)]/t \) converges to some locally integrable function (or in \( L^p \)) as \( t \) vanishes, in \( L^1_{\text{loc}}(\Omega) \), i.e., in \( L^1 \) for every compact subset of \( \Omega \).
Recall the convolution for functions with support in \([0, \infty)\), namely, if \(f\) belongs to \(L^p([0, \infty[)\) and \(g\) belongs to \(L^q([0, \infty[)\) then
\[
(f \ast g)(t) = \int_0^t f(t-s)g(s)ds = \int_0^t f(s)g(t-s)ds
\]
define a function belonging to \(L^r([0, \infty[)\), \(\|f \ast g\|_r \leq \|f\|_p \|g\|_q\), with \(1/p + 1/q - 1/r = 1\), see Young inequality B.65. Now, if \(\Gamma\) denotes the Gamma function, i.e.,
\[
\Gamma(\alpha) = \int_0^\infty s^{\alpha-1}e^{-s}ds, \quad \forall \alpha > 0, \tag{3.7}
\]
then consider the function \(\Phi_\nu(t) = t^{\nu-1}/\Gamma(\nu)\) for any \(t \geq 0\) and \(\nu > 0\). Define the integral of order \(\nu\) of a function \(f\) with support in \([0, \infty)\) by the expression
\[
I_\nu^t f = (f \ast \Phi_\nu)(t), \quad \text{i.e.,}
\]
\[
I_\nu^t f = \frac{1}{\Gamma(\nu)} \int_0^t (t-s)^{\nu-1}f(s)ds,
\]
as long as \(f\) belongs to \(L^1_{\text{loc}}([0, \infty[)\) and the integral exists, for almost every \(t > 0\).

**Exercise 3.15.** With the previous notation on fractional integrals, verify that \(\Phi_\nu \ast \Phi_\mu = \Phi_{\nu+\mu}\) and deduce that \(I_\nu^t I_\mu^t = I_{\nu+\mu}^t\). Moreover, if \(p, q\) belong to \([0, \infty)\) and \(0 < \nu < 1\) then \(I_\nu^t\) is a bounded operator from \(L^p\) into \(L^q\) if \(1 < p < 1/\nu\) and \(q = p/(1-p\nu)\), i.e., such that
\[
\int_0^\infty dt \left| \int_0^t (t-s)^{\nu-1}f(s)ds \right|^q \leq C \|f\|_p^q,
\]
for some a constant \(C = C_{p,q,\nu}\).

Certainly, once the fractional integral is defined, one can consider the fractional derivative as the inverse operator. For instance, the fractional derivative of order \(0 < r < 1\) of a locally integrable function \(f\) could be defined as \(\partial^r f(t) = I_t^{1-r}f'\) if the first derivative \(f'\) is locally integrable (i.e., \(f\) is locally absolutely continuous). An alternative definition is \(\partial^r f(t) = (I_t^{1-r}f)'\) if the derivative of function \(t \mapsto I_t^{1-r}f\) makes sense, and both ways may not be equivalent. This is known as the Riemann-Liouville (right-sided) fractional integral, e.g., see Oldham and Spanier [96], Kilbas et al. [72], Miller and Ross [91], among others.

### 3.2.2 Support and Finite Order

The spaces \(C^n_0(\Omega)\) and \(C^n(\Omega)\) have also a locally convex topology given in the same manner, i.e., \(\varphi_k \to \varphi\) in \(C^n_0(\Omega)\) if and only if (a) \(\partial^\alpha \varphi_k \to \partial^\alpha \varphi\) locally uniform in \(\Omega\) (i.e., in the sup-norm on any compact subset of \(\Omega\)) for every \(|\alpha| \leq n\); and (b) the supports of \(\varphi_k\) are all contained in a fixed compact subset.
of \( \Omega \). However, \( \varphi_k \to \varphi \) in \( C^n(\Omega) \) if and only if \( \partial^\alpha \varphi_k \to \partial^\alpha \varphi \) locally uniform in \( \Omega \) for every \( |\alpha| \leq n \). The convolution arguments prove that \( \mathcal{D}(\Omega) \) is a dense subspace of the lctv spaces \( C^n_0(\Omega) \), \( C^n(\Omega) \) and \( \mathcal{E}(\Omega) \). Thus, any continuous functional on \( C^n_0(\Omega) \), \( C^n(\Omega) \) or \( \mathcal{E}(\Omega) \) is uniquely determinate by its restriction to \( \mathcal{D}(\Omega) \).

It is then clear that \( \bigcap_n C^n(\Omega) = \mathcal{E}(\Omega) \) and \( \bigcap_n C^n_0(\Omega) = \mathcal{D}(\Omega) \). Hence, based on (3.5), the dual space \( \bigcup_n (C^n(\Omega))' = \mathcal{E}'(\Omega) \), while \( \bigcup_n (C^n_0(\Omega))' \subset \mathcal{D}'(\Omega) \), the equality of dual spaces is deduced from the condition (3.5), and the inclusion is actually strict, for instance, the distribution

\[
\langle T, \varphi \rangle = \sum_n \partial^n \varphi(x_n), \quad \forall \varphi \in \mathcal{D}(\mathbb{R}),
\]

for some unbounded sequence \( x_n \), does not belongs to the union \( \bigcup_n (C^n_0(\Omega))' \). Indeed we have

**Definition 3.27.** A distribution \( T \) on \( \Omega \) is called of finite order if there exists \( n = n(T) \) such that for every compact \( K \) of \( \Omega \) there is a constant \( C = C(K,T) \) satisfying

\[
|\langle T, \varphi \rangle| \leq C \sup \{ |\partial^\alpha \varphi(x)| : x \in K, \ |\alpha| \leq n \}, \quad \forall \varphi \in \mathcal{D}_K(\Omega),
\]

in other words, \( T \) is the restriction to \( \mathcal{D}(\Omega) \) of a continuous functional on \( C^n_0(\Omega) \), and the smallest \( n \) is called the order of \( T \). On the other hand, a distribution \( T \) vanishes near \( x_0 \) if there exists an open neighborhood \( U \subset \Omega \) of \( x_0 \) such that \( \langle T, \varphi \rangle = 0 \) for every \( \varphi \) in \( \mathcal{D}(U) \); and the relative (to \( \Omega \)) closed set of all points where \( T \) vanishes is called the support of \( T \) and denoted by \( \text{supp}(T) \).

In general, we cannot define the value of a distribution at a point \( x_0 \), but we can consider a distribution restricted to an open subset \( U \) of \( \Omega \). Hence, for two distributions \( T \) and \( S \) on \( \Omega \) we say \( T = S \) on \( U \) if and only if \( \langle T, \varphi \rangle = \langle S, \varphi \rangle \), for every \( \varphi \) in \( \mathcal{D}(U) \). Recall that a linear functional \( T \) defined on \( \mathcal{D}(\Omega) \) belongs to \( \mathcal{D}'(\Omega) \) (or equivalently is a distribution on \( \Omega \)), if for every compact \( K \) the restriction of \( T \) to \( \mathcal{D}_K(\Omega) \) is continuous, i.e., for every compact \( K \) of \( \Omega \) there exists an index \( n = n(K,T) \) such that

\[
|\langle T, \varphi \rangle| \leq C \sup \{ |\partial^\alpha \varphi(x)| : x \in K, \ |\alpha| \leq n \}, \quad \forall \varphi \in \mathcal{D}_K(\Omega),
\]

and some constant \( C = C(n,K,T) > 0 \). In contract, \( T \) is a distribution of finite order if the index \( n = n(K,T) \) can be chosen \( n = n(T) \) independent of the compact \( K \). However, the linear functional \( T \) becomes (or can be extended to) an element in \( \mathcal{E}'(\Omega) \) if there exist a compact \( K = K(T) \), an index \( n = n(T) \) and a constant \( C = C(T) > 0 \) such that

\[
|\langle T, \varphi \rangle| \leq C \sup \{ |\partial^\alpha \varphi(x)| : x \in K, \ |\alpha| \leq n \}, \quad \forall \varphi \in \mathcal{D}(\Omega).
\]

Indeed, the above estimate implies that \( \langle T, \varphi \rangle = 0 \) whenever \( \varphi = 0 \) in a neighborhood of the compact \( K \), and therefore, if \( \chi \) is a function in \( \mathcal{D}(\Omega) \) such that
\[ \chi = 1 \text{ on } K \text{ then the expression } \langle T, \varphi \rangle = \langle T, \chi \varphi \rangle \text{ can be used to define } T, \text{ for every } \varphi \text{ in } C^\infty(\Omega). \text{ Note that the support of } T \text{ is not necessarily } K, \text{ but } \text{supp}(T) \subset \text{supp}(\chi). \text{ Actually,}
\]

**Proposition 3.28.** Let \( T \) be a distribution on \( \Omega \) and \( U \) be the open set complementary to its support, i.e., \( U = \Omega \setminus \text{supp}(T) \). Then \( T = 0 \) on \( U \), i.e., \( T \) vanishes outside its support. Moreover, if \( \text{supp}(T) \) is a compact set of \( \Omega \) then \( T \) has a finite order and in fact, there exist a compact \( K = K(T) \), a nonnegative integer \( n = n(T) \) and a constant \( C = C(T) > 0 \) such that (3.8) hold true, and the distribution \( T \) has a unique extension to an element of \( \mathcal{E}'(\Omega) \). Furthermore, any element in \( \mathcal{E}'(\Omega) \) is indeed a distribution with compact support.

**Proof.** First, for any \( x \) in \( U \) there exists an open set \( U(x) \) such that \( T = 0 \) on \( U(x) \). Because the family of open sets \( \{U(x) : x \in U\} \) cover \( U \), by means of Theorem B.88, there exists a partition of the unity \( \{\chi_i : i \geq 1\} \) subordinate to \( \{U(x) : x \in U\} \). For any \( \varphi \) in \( \mathcal{D}(U) \), we write \( \varphi = \sum_i \varphi_i \) with \( \varphi_i = \chi_i \varphi \), where the sum is finite since \( \varphi \) has a compact support. Thus \( \langle T, \varphi \rangle = \sum_i \langle T, \varphi_i \rangle \), and because each \( \varphi_i \) has support in some \( U(x) \), we deduce that \( \langle T, \varphi_i \rangle = 0 \) for every \( i \), i.e., \( T = 0 \) on \( U \).

If \( \chi \) is an element in \( C^\infty(\Omega) \) satisfying \( \chi = 1 \) in an open set containing \( \text{supp}(T) \) then \( \chi T = T \), i.e., \( \langle T, \varphi \rangle = \langle T, \chi \varphi \rangle \), for every \( \varphi \) in \( \mathcal{D}(\Omega) \). Hence, if \( \text{supp}(T) \) is compact then we can choose \( \chi \) as above with support in some compact set \( K \) of \( \Omega \); and because \( T \) is a continuous linear functional, there exists an index \( n = n(T) \) and a constant \( C_1 = C_1(T) \) such that

\[ |\langle T, \chi \varphi \rangle| \leq C_1 \sup \{|\partial^\alpha (\chi \varphi)(x)| : x \in K, |\alpha| \leq n\}, \quad \forall \varphi \in \mathcal{D}_K(\Omega). \]

Since \( \chi \) and \( K \) are fixed, and \( \chi \varphi \) belongs to \( \mathcal{D}_K(\Omega) \) for every \( \varphi \) in \( \mathcal{D}(\Omega) \), for another constant \( C_2 = C_2(\chi) \) we obtain

\[ \sup \{|\partial^\alpha (\chi \varphi)(x)| : x \in K, |\alpha| \leq n\} \leq C_2 \sup \{|\partial^\alpha \varphi(x)| : x \in K, |\alpha| \leq n\}, \quad \forall \varphi \in \mathcal{D}(\Omega), \]

which yields the desired estimate (3.8) with \( C = C_1 C_2 \). Hence, because \( \mathcal{D}(\Omega) \) is dense in \( \mathcal{E}(\Omega) \), the distribution \( T \) with compact support can be extended to an element in \( \mathcal{E}'(\Omega) \).

Finally, if \( T \) is an element of \( \mathcal{E}'(\Omega) \) then estimate (3.8) holds true for some compact \( K \) and some \( n \), which implies that \( T \) belongs to the dual space \( C^n(\Omega)' \), i.e., \( T \) has a compact support and a finite order. \( \square \)

- **Remark 3.29.** The argument about the partition of the unity used in the proof of Proposition 3.28 also shows that for any family \( \{T_\omega\} \) of distributions on \( \omega \), such that \( \{\omega\} \) is an open cover of \( \Omega \) and \( T_\omega = T_\omega' \) on \( \omega \cap \omega' \), there exists a unique distribution \( T \) on \( \Omega \) such that \( T = T_\omega \) on \( \omega \). \( \square \)

- **Remark 3.30.** If \( T \) is a distribution with compact support then estimate (3.8) may not hold with \( K = \text{supp}(T) \). Indeed, consider \( \langle T, \varphi \rangle = \sum_k \langle \varphi(1/k) - \]
\( \varphi(0)/k \). The estimate

\[
|\langle T, \varphi \rangle| \leq \left( \sum_k 1/k^2 \right) \sup_{x \in [0,1]} |\varphi'(x)|,
\]

shows that \( T \) is a distribution in \( \mathbb{R} \), indeed, \( T \) belongs to the dual space \( C^1(\mathbb{R})' \). The set \{0\} \( \cup \{1/k : k = 1, 2, \ldots \} \) is compact and \( \langle T, \varphi \rangle = 0 \) for any test function \( \varphi \) with support in \( \mathbb{R} \setminus K \), i.e., the support of \( T \) is exactly the compact set \( K \). Hence, take a sequence \( \{\varphi_k\} \) of test functions satisfying \( \varphi_k = 1 \) near \([1/k, 1)\) and \( \varphi_k = 0 \) near \([1/(k+1), 0]\) to check that \( \sup \{|\partial^\alpha \varphi_k(x)| : x \in K, |\alpha| \leq n\} = 1 \) and \( \langle T, \varphi_k \rangle = \sum_{i \leq k} 1/i \), for every \( k = 1, 2, \ldots \), which proves that the estimate (3.8) cannot hold true with \( K = \text{supp}(T) \) and some constant \( C = C(T) \).

Thus, elements in \( C^\infty(\Omega) = \mathcal{E}(\Omega) \) are called sometimes smooth functions, and certainly, the elements of its dual space \( \mathcal{E}'(\Omega) \) are called distribution with compact support.

**Exercise 3.16.** Show that the expressions \( \langle T, \varphi \rangle = \sum_k 2^k \varphi(1/k) \) and \( \langle S, \varphi \rangle = \sum_k 2^k \varphi(k)/k \) define two distributions \( (0, 1) \subset \mathbb{R} \), where \( T \) is of order 0 while \( S \) is not of finite order. Check that the support of each of them is not compact. Can you modify the above expressions to produce a distribution which is not of finite order and has a compact support?

**Exercise 3.17.** Let \( \{x_k\} \) be a sequence of points in \( \Omega \) such that the distance from \( x_k \) to the boundary \( \partial \Omega \) goes to zero, or such that \( |x_k| \to \infty \) if \( \Omega = \mathbb{R}^d \). Define \( \langle T, \varphi \rangle = \sum_k \varphi(x_k) \) and \( \langle S, \varphi \rangle = \sum_k \partial_x^\alpha \varphi(x_k) \). Discuss if \( T \) and \( S \) are distributions, and if so, find their order and support.

**Proposition 3.31.** If \( T \) is a distribution on \( \Omega \) of finite order \( n \) and with a compact support \( K \) then \( \langle T, \varphi \rangle = 0 \) for every \( \varphi \) in \( \mathcal{D}(\Omega) \) satisfying \( \partial^\alpha \varphi = 0 \) on \( K \) for every multi-index \( \alpha \) of order \( |\alpha| \leq n \).

**Proof.** Note that because \( T \) has a compact support in \( \Omega \), it is necessary of finite order. For a given \( \delta \) sufficiently small, denote by \( U_\delta \) the open set of all points in \( \Omega \) within a distance \( \delta \) of \( K \). Now, for any \( \varphi \) in \( \mathcal{D}(\Omega) \) satisfying \( \partial^\alpha \varphi = 0 \) on \( K \), for every \( |\alpha| \leq n \), we have

\[
\sup_{x \in U_\delta, |\alpha| \leq n} |\partial^\alpha \varphi(x)| = \delta^{n-|\alpha|} \varepsilon, \quad \text{with } \varepsilon \to 0 \text{ as } \delta \to 0.
\]

Next, by convolution with characteristic function \( \mathbb{1}_{U_{\delta/4}} \), we construct a smooth function \( \chi \) such that \( \chi = 1 \) in \( U_{\delta/4} \), \( \chi = 0 \) in \( \mathbb{R}^d \setminus U_\delta \) and \( |\partial^\alpha \chi(x)| \leq C \chi \delta^{-|\alpha|} \). Hence, \( |\partial^\alpha (\varphi \chi)(x)| \leq C \varepsilon \), for every \( x \), and some constant \( C \) independent of \( \delta \).

Since \( \langle T, \varphi \chi \rangle = \langle T, \varphi \rangle \) we deduce

\[
|\langle T, \varphi \rangle| \leq C \|\varphi\|_{(n)} \leq C \varepsilon,
\]

which implies that \( \langle T, \varphi \rangle = 0 \). \( \square \)
Exercise 3.18. Consider the distribution \(\langle T, \varphi \rangle = \sum_{|\alpha| \leq n} c_\alpha \partial^\alpha \varphi(x_0)\), where \(c_\alpha\) are constants, and \(x_0\) is a point in \(\Omega\). (a) Verify that the support of \(T\) is the point \(x_0\) and that the order is \(n\) if for some \(\alpha\) with \(|\alpha| = n\) we have \(c_\alpha \neq 0\). (b) Prove that the only distributions on \(\Omega\) with support equal to a simple point \(\{x_0\}\) are finite linear combinations of the derivative of the Dirac delta at \(x_0\), i.e., as \(T\) above.

Exercise 3.19. Verify that if \(u\) is a function in \(C^n(\mathbb{R}^d)\) then, for any \(|\alpha| \leq n\), the function defined by \(U_\alpha(x, x) = 0\) and

\[
U_\alpha(x, y) = \bigg| \partial^\alpha u(x) - \sum_{|\beta| \leq n - |\alpha|} \frac{\partial^{\alpha + \beta} u(y)(x - y)^\beta}{\beta!} \bigg| |x - y|^{-n}, \quad \forall x \neq y,
\]

is continuous on \(\mathbb{R}^d \times \mathbb{R}^d\). Actually, the converse of is called Whitney’s Extension Theorem, i.e., given continuous functions \(u_\alpha, |\alpha| \leq n\), on a compact set \(K\) of \(\mathbb{R}^d\), define the functions \(U_\alpha(x, y)\) on \(K \times K\) by means of the above expression replacing \(\partial^\alpha u(x)\) with \(u_\alpha(x)\) and \(\partial^{\alpha + \beta} u(y)\) with \(u_{\alpha + \beta}(y)\). If \(U_\alpha\) are continuous on \(K \times K\) then there exits a function \(u\) in \(C^n(\mathbb{R}^d)\) such that \(\partial^\alpha u = u_\alpha\) and

\[
\sum_{|\alpha| \leq n} \sup_{K} |\partial^\alpha u| \leq C \left[ \sum_{|\alpha| \leq n} \sup_{K \times K} |U_\alpha| + \sum_{|\alpha| \leq n} \sup_{K} |u_\alpha| \right],
\]

for some constant \(C\) depending only on \(K\), e.g., see Hörmander [68, Section 2.3, pp. 44–52].

- Remark 3.32. Complementing Exercise 3.18, let us express points in \(\mathbb{R}^d\) as \(x = (x', x_d)\) and let \(T\) be a distribution in \(\mathbb{R}^d\) of finite order \(n\) and with support in \(\mathbb{R}^{d-1} \times \{0\}\). Now, given any test function \(\varphi\) in \(\mathbb{R}^d\), even if it is not permissible expand \(\varphi\) in a power series at boundary points of it support, we can write

\[
\varphi^*(x) = \varphi(x) - \sum_{k \leq n} (\partial^k_d \varphi(x', 0)) \frac{x_d^k}{k!},
\]

where \(\varphi^*\) satisfies \(\partial^\alpha \varphi^*(x', 0) = 0\) for every multi-index with \(|\alpha| \leq n\). If \(\psi\) is an element of \(\mathcal{D}(\mathbb{R}^{d-1})\) then we can check that for \(k \geq 0\), the expression

\[
\langle T_k, \psi \rangle = \langle T, \overline{\psi}_k \rangle, \quad \text{with} \quad \overline{\psi}_k(x', x_d) = \frac{(-x_d)^k}{k!} \psi(x'),
\]

defines a distribution on \(\mathbb{R}^{d-1}\). Even if \(T\) has not a compact support, the arguments Proposition 3.31 show that \(\langle T, \varphi^* \rangle = 0\), which implies

\[
\langle T, \varphi \rangle = \sum_{k \leq n} \langle T_k, (-1)^k \partial^k_d \varphi(\cdot, 0) \rangle, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d).
\]

This is to say that \(T = \sum_{k \leq n} \partial^k_d T_k\), where \(\overline{T}_k\) is the extension of \(T_k\) to \(\mathbb{R}^d\), namely \(\langle \overline{T}_k, \varphi \rangle = \langle T_k, \varphi(\cdot, 0) \rangle\). Certainly, we can extend these arguments to the case where \(\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}\).
### 3.2.3 Distribution on Manifolds

Recall that if $S \subset \mathbb{R}^d$ is a $C^\infty$ manifold of dimension $m$ with local coordinates $\phi_i : D_i \subset \mathbb{R}^m \to \mathbb{R}^d$ and atlas $\Phi = \{\phi_i : i \geq 1\}$ then the Euclidean (Lebesgue) surface measure $d\sigma$ on $S$ is locally defined by

$$
\int_S f(x)\sigma(dx) = \sum_i \int_{\mathbb{R}^m} \chi_i(\phi_i(y')) f(\phi_i(y')) \sqrt{\det (\nabla \phi_i(y')^* \nabla \phi_i(y'))} \, dy'.
$$

where $\{\chi_i : i \geq 1\}$ is a $C^\infty$ partition of the unity of a neighborhood of $S$ subordinate to the atlas $\Phi$.

Therefore, a locally integrable function $f$ on $S$, i.e., $f$ in $L^1_{\text{loc}}(S)$, can be regarded as a distribution on $S$, i.e.,

$$
\langle f, \varphi \rangle = \int_S f(x)\varphi(x)\sigma(dx),
$$

as long as the factor $\sqrt{\det (\nabla \phi_i(y')^* \nabla \phi_i(y'))}$ is involved.

Essentially, by $C^\infty$ local charts the space of distribution $\mathcal{D}'(\mathbb{R}^m)$ is transported (locally) to the manifold $S$, to form $\mathcal{D}'(S)$, as well as the test functions $\mathcal{D}(S)$. However, the pairing $\langle \cdot, \cdot \rangle$ contain a density, with respect to the Lebesgue measure. If the manifold $S$ is only of class $C^k$ with $k \geq 1$ then only distribution of order $k$ can be considered on the manifold $S$. A special situation is the case of Lipschitz manifold, where the density is only defined almost everywhere.

On the other hand, by means of the co-area formula with the surface Lebesgue (or the Hausdorff) measure $\ell_{d-1}$ in $\mathbb{R}^d$ (e.g., see our previous book Menaldi [89]), i.e.,

$$
\int_\Omega f(x)|\nabla g(x)|\,dx = \int_{\mathbb{R}} ds \int_{g^{-1}(s)} f(x)\ell_{d-1}(dx),
$$

in the case where $g$ is continuously differentiable and the gradient is nowhere zero on the manifold $S = \{x \in \mathbb{R}^d : g(x) = 0\}$ of dimension $(d-1)$, the Dirac delta function (measure or distribution) can be defined by

$$
\langle \delta_g^{-1(0)}, \varphi \rangle = \int_{\mathbb{R}^d} \delta_0(g(x))\varphi(x)\,dx = \int_S \varphi(x)|\nabla g(x)|^{-1}\sigma(dx),
$$

i.e., $\delta_0(g(x))$ is a distribution (or measure) in $\mathbb{R}^d$ with support in $S$.

More general, the co-area formula (e.g., see our previous book), i.e.,

$$
\int_\Omega f(x)\sqrt{\nabla g^* \nabla g} \,dx = \int_{\mathbb{R}^m} ds \int_{g^{-1}(s)} f(x)\ell_{d-m}(dx),
$$

suggests that if $g = (g_1, \ldots, g_m)$ is a continuously differentiable and the Jacobian $\sqrt{\nabla g^* \nabla g}$ is nowhere zero on the manifold $S$ of dimension $(d-m)$ in $\mathbb{R}^d$ given by $S = \{x \in \mathbb{R}^d : g_i(x) = 0, i = 1, \ldots, m\}$, then the distribution

$$
\langle \delta_S, \varphi \rangle = \int_{\mathbb{R}^d} \delta_0(g(x))\varphi(x)\,dx = \int_S \varphi(x)(\nabla g^* \nabla g)^{-1/2}\sigma(dx)
$$
represents the equivalent of the Dirac delta function on the surface $S$, as associated by the simple layer potentials on $S$. For instance, on a subspace $S = \{x_1 = \ldots = x_m = 0\}$ the Dirac delta function takes the form

$$\langle \delta_S, \varphi \rangle = \int_{\mathbb{R}^{d-m}} \varphi(0, \ldots, 0, x_{m+1}, \ldots, x_d) \, dx_{m+1} \ldots dx_d,$$

which can be extended to any affine manifold. Certainly, these arguments can be localized, and therefore, extended first to $C^1$ manifolds and then to Lipschitz manifolds.

In general the situation is far from trivial and a lot of details need consideration. Depending on the reader interest, for instance the comprehensive books Gelfand and Shilov [53] and Gelfand and Vilenkin [52] can be consulted. Partial differential equations are a good source of challenging problems, e.g., see Chazarain and Piriou [26] and references therein.

### 3.2.4 Avoiding Inductive Limits

Actually, it not completely necessary to discuss inductive limit to develop a calculus of distributions. Indeed, beginning with the vector space $C_0^\infty(\Omega)$, we may introduce the concept of distribution on an open set $\Omega \subset \mathbb{R}^d$ as a linear functional $T$ on $C_0^\infty(\Omega)$ such that $T(\varphi_n) \to 0$ each time that the sequence $\{\varphi_n\}$ satisfies (a) for every multi-index $\alpha$, $\partial^\alpha \varphi_n \to 0$ uniformly, and (b) the supports of $\varphi_n$ are contained in a compact subset of $\Omega$. From this point forward, we can construct a useful set of properties for the distributions, but we are losing the functional analysis aspect of the matter, e.g., see the book Richards and Youn [105].

As implicitly mentioned, there is not problem in considering distribution with complex values, i.e., the space $D(\Omega)$ could be defined as complex-valued test functions. More general, we may have test function with valued in some Banach (or Hilbert) space $B$ in a strong way (i.e., taking derivative in the Fréchet sense) or in the weak sense (i.e., assuming $x \mapsto \langle b', \varphi(x) \rangle$ in $C_0^\infty(\Omega)$ for every $b'$ in the dual space of $B$). Clearly, depending on the properties of the space $B$, we may have serious difficulties, but many arguments can be extended.

The topology of the space $S(\mathbb{R}^d)$ and its dual $S'(\mathbb{R}^d)$ are easier, they are Fréchet (metrizable complete locally convex vector) spaces, and a linear functional $T$ on $S(\mathbb{R}^d)$ belongs to $S'(\mathbb{R}^d)$ (which are called tempered distributions) if $T(\varphi_n) \to 0$ each time that the sequence $\{\varphi_n\}$ satisfies $\sum (1 + |x|^m) \partial^\alpha \varphi_n(x) \to 0$ uniformly in $x$ belonging to $\mathbb{R}^d$, for every multi-index $\alpha$ and any positive integer $m$. As seen later, this space plays a key role in the theory of the Fourier transform.

The space $\mathcal{E}(\Omega)$ and its dual $\mathcal{E}'(\Omega)$ are also Fréchet spaces, and a linear functional $T$ on $\mathcal{E}(\Omega)$ belongs to $\mathcal{E}'(\Omega)$ (i.e., a distribution with compact support) if $T(\varphi_n) \to 0$ each time that the sequence $\{\varphi_n\}$ satisfies $\partial^\alpha \varphi_n \to 0$ locally uniformly for every multi-index $\alpha$.

These Fréchet spaces can be regarded as countable normed spaces (e.g., see Friedman [45, Section 1.3, pp. 6–8]), where the sequence of seminorms (or
norms) \( p_1 \leq p_2 \leq \cdots \) have been sorted.

Clearly, the pattern is to replace or completely eliminate the condition “compact support” in the initial space \( C_0^\infty \). Namely, by using a condition of the type “\( p \)-integrable” or “essentially bounded”, we may define \( \mathcal{D}_{L^p}(\Omega) \) as the space \( \{ \varphi \in C^\infty(\Omega) : \partial^\alpha \varphi \in L^p(\Omega), \ \forall \alpha \} \), with \( 1 \leq p \leq \infty \). Certainly, these spaces contain the test functions, but new difficulties appear when \( \mathcal{D}(\Omega) \) is not necessarily a dense subspace, which is essentially the case when \( \Omega \neq \mathbb{R}^d \) or \( p = \infty \). An immediate remedy is to essentially require density, i.e., adding a condition \( \mathcal{D}(\Omega) \subset \mathcal{D}_{L^p}(\Omega) \), which is essentially the case when \( \Omega \), the open set \( \Omega \) will play a role if this condition actually becomes necessary a dense subspace, which is essentially the case when \( \Omega \neq \mathbb{R}^d \) or \( p = \infty \). As seen later, the smoothness of the open set \( \Omega \) will play a role if this condition actually becomes \( \partial^\alpha \varphi(x) = 0 \) for every \( x \) on \( \partial \Omega \).

All these lines of analysis yield Fréchet spaces, but if \( \mathcal{D}(\Omega) \) is not dense then their dual spaces may contains elements that are not distribution. This gives rise to “traces” or distributions on the boundary.

For \( 1 \leq p < \infty \) we keep the notation

\[
\mathcal{D}_{L^p} = \{ \varphi \in C^\infty(\mathbb{R}^d) : \partial^\alpha \varphi \in L^p(\mathbb{R}^d), \ \forall \alpha \},
\]

and we use \( \mathcal{B} \) to denote the space of functions \( \varphi \) in \( C^\infty(\mathbb{R}^d) \) such that for any multi-index \( \alpha \) we have

\[
\partial^\alpha \varphi \in L^\infty(\mathbb{R}^d) \quad \text{and} \quad \sup_{|x| \geq r} |\partial^\alpha \varphi(x)| \to 0 \quad \text{as} \quad r \to \infty.
\]

The density of \( \mathcal{D} \) in \( \mathcal{B} \) is immediate shown, but the inclusion \( \mathcal{D}_{L^p} \subset \mathcal{B} \) is not quite obvious. This becomes clear after proving the following inequality: for any \( a > 0 \) there exists a constant \( C = C_{a,d} \) such that

\[
|\varphi(x)| \leq C \sum_{|\alpha| \leq d} \int_{|x-y| \leq a} |\partial^\alpha \varphi(y)| \, dy, \quad \forall \varphi \in C^d(\mathbb{R}^d). \tag{3.9}
\]

Indeed, if \( h \) denotes the Heaviside’s function (i.e., \( h(x) = h_1(x_1) \cdots h_d(x_d) \), with \( h_i(x_i) = 0 \) if \( x_i < 0 \) and \( h_i(x_i) = 1 \) if \( x_i \geq 0 \)) then the Dirac measure can be written as \( \delta = \partial^{(1,1,\ldots,1)} h \). Thus, for any distribution \( T \) with compact support we have \( T * \delta = h * \partial^{(1,1,\ldots,1)} T \). Hence, take a function \( k \) in \( \mathcal{D} \) such that \( k(x) = 1 \) if \( 2|x| \leq a \) and \( k(x) = 0 \) if \( |x| \geq a \), and for any fixed \( x \) consider the function \( y \mapsto k(x-y)\varphi(y) \) as the distribution \( T \) to deduce the equality

\[
k(x-\xi)\varphi(\xi) = \int_{\mathbb{R}^d} h(y)[\partial^{(1,1,\ldots,1)}(k(x-\xi + y)\varphi(\xi - y))] \, dy, \quad \forall x, \xi \in \mathbb{R}^d,
\]

which yields the inequality (3.9), upon choosing \( \xi = x \) and bounding \( k \) and all its first derivatives.

**Exercise 3.20.** By means of the inequality (3.9), prove the inclusions \( \mathcal{S} \subset \mathcal{D}_{L^p} \subset \mathcal{D}_{L^q} \subset \mathcal{B} \), for any \( 1 \leq p \leq q < \infty \) as well as the density of \( \mathcal{D} \) in any of those spaces. \( \square \)
At this point, we can take the dual spaces to deduce the reverse inclusions $S' \subset D'_{L_p} \subset \mathcal{B}' \subset D'$, where $D'_{L_p}$ denotes the dual of the space $D_{L^q}$ with $1/p + 1/q = 1$, e.g., see Schwartz [112, Section VI.8, pp. 199–205].

### 3.3 More Operations and Localization

We are interested in the pointwise product, tensor product and convolution of distributions and smooth functions. Also, a local expression for distribution is desired.

#### 3.3.1 Product of Distributions

It is clear the meaning of (pointwise) product and tensor product of functions, for instance, if $f = f(x)$, $g = g(x)$ and $h = h(y)$ then $(fg)(x) = f(x)g(x)$ and $(f \otimes h)(x, y) = f(x)h(y)$ define the pointwise product and the tensor product. Thus, for any $\varphi_i$ in $D(\Omega_i)$, $i = 1, 2$, the function $\varphi(x) = \varphi_1(x_1)\varphi_2(x_2)$ belongs to $D(\Omega_1 \times \Omega_2)$, with the notation $x = (x_1, x_2)$ and $\varphi = \varphi_1 \otimes \varphi_2$.

By localizing (i.e., multiplying by smooth functions with compact support which are equal to 1 in some compact set) polynomials in both variables we show that the vector space generated by the family $D(\Omega_1) \otimes D(\Omega_2)$ of functions $\varphi_1 \otimes \varphi_2$ with $\varphi_i$ in $D(\Omega_i)$ is dense in $D(\Omega_1 \times \Omega_2)$. Hence, a distribution on $\Omega_1 \times \Omega_2$ is uniquely determined by its values on functions of the form $\varphi_1 \otimes \varphi_2$ (this is analogous to the product of measures). Therefore, the tensor product $T_1 \otimes T_2$ of two distributions $T_i \in D'(\Omega_i)$ is first defined by

$$
\langle T_1 \otimes T_2, \varphi_1 \otimes \varphi_2 \rangle = \langle T_1, \varphi_1 \rangle \langle T_2, \varphi_2 \rangle, \quad \forall \varphi_i \in D(\Omega_i),
$$

and then extended (by linearity and continuity) to any test function of the form $\varphi(x_1, x_2)$, which uniquely determines $T_1 \otimes T_2$. To verify that $T_1 \otimes T_2$ can be extended by continuity, we may proceed as follows: if $\varphi$ belongs to $D(\Omega_1 \times \Omega_2)$ then the function $\varphi(\cdot, x_2)$ belongs to $D(\Omega_1)$, for any fixed $x_2$. Moreover, if $T_1$ is an element of $D'(\Omega_1)$ then the function $x_2 \mapsto \langle T_1, \varphi(\cdot, x_2) \rangle$ belongs to $D(\Omega_2)$ and $\partial_2(T_1, \varphi(\cdot, x_2)) = \langle T_1, \partial_2 \varphi(\cdot, x_2) \rangle$, where $\partial_2$ denotes any derivative in the variable $x_2$. Thus, we have $\langle T_1 \otimes T_2, \varphi \rangle = \langle T_1, \langle T_2, \varphi(\cdot, x_2) \rangle \rangle$, for every $\varphi$ in $D(\Omega_1 \times \Omega_2)$, and details (such as the verification that $T_1 \otimes T_2$ is a continuous functional) are left to the reader. The above arguments also show that the vector space generated by the family $D'(\Omega_1) \otimes D'(\Omega_2)$ of distributions $T_1 \otimes T_2$ with $T_i \in D'(\Omega_i)$ is dense in $D'(\Omega_1 \times \Omega_2)$. Sometimes, if $T_i$ is a distribution in $\Omega_i$ and $x_i$ denotes the variable in $\Omega_i$ then $T_1(x_1) \otimes T_2(x_2)$ indicates the tensor product $T_1 \otimes T_2$, with explicit mention to the variables.

**Exercise 3.21.** Complete the previous statements: show that (1) $D(\Omega_1) \otimes D(\Omega_2)$ is dense in $D(\Omega_1 \times \Omega_2)$; (2) $T_1 \otimes T_2$ can be uniquely extended to a distribution in $\Omega_1 \times \Omega_2$; and (3) the support of the tensor product of two distributions $T_1 \otimes T_2$ is the Cartesian product of their support $\text{supp}(T_1) \times \text{supp}(T_2)$.  

[Prepared by Menaldi]  
November 11, 2016
Remark 3.33. First, a partition of the unity with test functions shows that any two elements in \( D(\Omega) \) are equals if and only if they are locally equals (i.e., \( T = S \) iff \( \chi T = \chi S \) for any test function \( \chi \)). Therefore, combining this assertion with the density of the vector space \( D(\Omega_1) \otimes D(\Omega_2) \) on \( D(\Omega_1 \times \Omega_2) \), it should be clear that if \( T \) and \( S \) are elements in \( D(\Omega) \) with \( \Omega \subset \mathbb{R}^d \) then \( T = S \) if and only if \( \langle T, \varphi \rangle = \langle S, \varphi \rangle \) for any test function \( \varphi \) in \( D(\Omega) \) of the following product form \( \varphi(x) = \varphi_1(x_1) \ldots \varphi_d(x_d) \).

Exercise 3.22. Reconsider the previous question as follows: If \( T \) is a distribution in \( \Omega_1 \times \Omega_2 \) then we can define a continuous linear operator \( T : D(\Omega_2) \to D'(\Omega_1) \) by the formula

\[
\langle T \psi, \varphi \rangle = \langle T, \varphi \otimes \psi \rangle, \quad \forall \varphi \in D(\Omega_1), \quad \psi \in D(\Omega_2).
\]

Prove that the application \( T \mapsto T \) is injective and surjective.

Consider the multiplication (pointwise product) by a function \( f \) in \( C^\infty(\Omega) \) as an operation from \( D(\Omega) \) into itself, \( \varphi \mapsto f \varphi \). It is rather simple to check the linearity and continuity of this operation, so that, for a given element in \( D'(\Omega) \) we can define \( fT \) in \( D'(\Omega) \) as

\[
\langle fT, \varphi \rangle = \langle T, f \varphi \rangle, \quad \forall \varphi \in D(\Omega),
\]

we can multiply a distribution and a smooth function. Similarly, if \( T \) is a distribution of finite order, i.e., \( T \) belongs to the dual of \( C^n_0(\Omega) \) for some \( n = 0, 1, \ldots \), then the multiplication \( fT \) is also defined for any \( f \) in \( C^n(\Omega) \), not necessarily infinitely differentiable. In particular, we can multiply a continuous function and Radon measure. Moreover, from the definition of multiplication, we deduce that the derivative \( \partial(fT) = (\partial f)T + f(\partial T) \) follows the usual product rule. The support of \( fT \) is contained into the intersection of the support of \( f \) and the support of \( T \). Also, we have

\[
(f_1 \otimes f_2)(T_1 \otimes T_2) = (f_1T_1) \otimes (f_2T_2),
\]

combining the tensor product and the multiplication.

Exercise 3.23. Let \( x = (x', x_d) \) a point in \( \mathbb{R}^d \), with \( x' \) in \( \mathbb{R}^{d-1} \), and \( \mathbb{R}^d_+ = \mathbb{R}^{d-1} \times [0, \infty) \). If \( f \) belongs to \( C^\infty(\mathbb{R}^d_+) \) we denote its zero-extension to the whole \( \mathbb{R}^d \) by \( \overline{f} \), i.e., \( \overline{f}(x', x_d) = f(x', x_d) \) if \( x_d \geq 0 \), and \( \overline{f}(x', x_d) = 0 \) if \( x_d < 0 \). Consider \( \overline{f} \) as a distribution on \( \mathbb{R}^d \) and prove that its first derivative in the normal direction \( x_d \), is given by the formula \( \partial_d \overline{f} = \overline{\partial_d f} + J \), where

\[
\langle J, \varphi \rangle = \int_{\mathbb{R}^{d-1}} f(x', 0) \varphi(x', 0) \, dx', \quad \forall \varphi \in D(\mathbb{R}^d)
\]

Moreover, by means of the Dirac function, give a formula for the \( n \)-derivative in the normal direction \( x_d \), \( \overline{\partial^n_d \varphi} \), in term of a tensor product of distributions. Furthermore, obtain a similar formula in general, for any derivative \( \overline{\partial^\alpha \varphi} \) for any multi-index \( \alpha \).
It probably necessary to acknowledge that the ‘theory of distributions’ is a completely linear concept, in the sense that the multiplication of two distributions cannot consistently be defined as an extension of the product of a distribution (denoted with $\times$ to emphasize the point) and a test function preserving the ‘associative property’, e.g., multiplying the principal value $v.p.(1/x)$ with $x$ and $\delta$, the contradictory equalities
\[
(\delta \times x) \times (v.p.(1/x)) = 0 \quad \text{and} \quad \delta \times ((x \times v.p.(1/x))) = \delta
\]
appear.

### 3.3.2 Convolution of Distributions

As seen early, when working in the whole space $\Omega = \mathbb{R}^d$, the expression
\[
(f \ast g)(x) = \int_{\mathbb{R}^d} f(x - y) g(y) \, dy = \int_{\mathbb{R}^d} f(y) g(x - y) \, dy, \quad \forall x \in \mathbb{R}^d
\]
define the convolution of two functions, whenever this integral makes sense.

Recall first the space of infinity differentiable functions $E(\mathbb{R}^d)$ and next, the space $S(\mathbb{R}^d)$ of rapidly decreasing smooth functions, with its Fréchet topology given by the double sequence of seminorms $\{p_{n,k} : n, k \geq 0\},$
\[
p_{n,k}(\varphi) = \sup \{ (1 + |x|^2)^{k/2} |\partial^\alpha \varphi(x)| : x \in \mathbb{R}^d, |\alpha| \leq n \}\tag{3.10}
\]
and its dual $S'(\mathbb{R}^d)$, the space of tempered distributions.

If $\varphi, \psi$ belongs to $D(\mathbb{R}^d)$ then $\varphi \ast \psi$ belongs $D(\mathbb{R}^d)$ and supp$\{\varphi \ast \psi\} \subset$ supp$\{\varphi\}$ + supp$\{\psi\}$. To check that if $\varphi, \psi$ belongs to $S(\mathbb{R}^d)$ then $\varphi \ast \psi$ belongs to $S(\mathbb{R}^d)$, the argument is longer. Indeed, Peetre’s inequality
\[
\frac{(1 + |x|^2)^s}{(1 + |y|^2)^s} \leq 2^{|s|}(1 + |x - y|^2)^{|s|}, \quad \forall x, y \in \mathbb{R}^d, s \in \mathbb{R} \tag{3.11}
\]
yields
\[
\int_{\mathbb{R}^d} (1 + |x - y|^2)^{-k/2}(1 + |y|^2)^{-r/2} \, dy \leq \\
\leq 2^{k/2}(1 + |x|^2)^{-k/2} \int_{\mathbb{R}^d} (1 + |y|^2)^{(k-r)/2} \, dy.
\]
Therefore, if $\varphi, \psi$ belongs to $S(\mathbb{R}^d)$, $r > k + d$ and $|\alpha| \leq n$, then the estimate
\[
| (\partial^\alpha \varphi \ast \psi)(x) | \leq C p_{n,k}(\varphi)p_{0,r}(\psi) \int_{\mathbb{R}^d} (1 + |x - y|^2)^{-k/2}(1 + |y|^2)^{-r/2} \, dy
\]
implies that $p_{n,k}(\varphi \ast \psi)$ is finite, i.e., $\varphi \ast \psi$ belongs to $S(\mathbb{R}^d)$.

Now, if $\psi$ is a smooth function with compact support and $T$ is a distribution on $\mathbb{R}^d$ then we can define the function
\[
x \mapsto (T \ast \psi)(x) = \langle T, \psi(x - \cdot) \rangle, \quad \forall x \in \mathbb{R}^d, \tag{3.12}
\]
which belongs to $C^\infty(\mathbb{R}^d)$, and agrees with the usual definition if $T = f$ a locally integrable function. On the other hand, if $T$ has a compact support, i.e., $T$ belongs to $\mathcal{E}'(\mathbb{R}^d)$, then $T \ast \psi$ is also defined for any smooth $\psi$ non necessary with compact support, i.e., in $\mathcal{E}(\mathbb{R}^d)$. In any case, we can show the following relation on their supports: $\text{supp}\{T \ast \psi\} \subseteq \text{supp}\{T\} + \text{supp}\{\psi\}$. Analogously, the duality between $C^n$ or $C^n_0$ can be used. In particular, the convolution of a Radon measure $\mu$ and an element in $C_0^0$ (or $L^1_{\text{loc}}$) belongs to $C_0^0$ (or $L^1_{\text{loc}}$). If either (a) $\varphi$ and $\psi$ belongs to $\mathcal{D}(\mathbb{R}^d)$ and $T$ belongs to $\mathcal{D}'(\mathbb{R}^d)$ or (b) $\varphi$ belongs to $\mathcal{E}(\mathbb{R}^d)$ and $\psi$ belongs to $\mathcal{D}(\mathbb{R}^d)$ and $T$ belongs to $\mathcal{E}'(\mathbb{R}^d)$ (i.e., at least two of the three elements have compact supports), then it is not hard to show that $(T \ast \varphi) \ast \psi = T \ast (\varphi \ast \psi)$.

Below, we summarize the main properties of the convolution of a distribution and a smooth function:

**Proposition 3.34.** The convolution of a distribution $T$ and a smooth function $\psi$ is defined by (3.12) in each of the following cases: (a) $T$ in $\mathcal{D}'(\mathbb{R}^d)$ and $\psi$ in $\mathcal{D}(\mathbb{R}^d)$, (b) $T$ in $\mathcal{E}'(\mathbb{R}^d)$ and $\psi$ in $\mathcal{E}(\mathbb{R}^d)$ and (c) $T$ in $\mathcal{S}'(\mathbb{R}^d)$ and $\psi$ in $\mathcal{S}(\mathbb{R}^d)$. Moreover, in each case, $T \ast \psi$ is an element in $\mathcal{E}(\mathbb{R}^d)$ and its derivatives can be calculated by

$$
\partial^\alpha (T \ast \psi) = (\partial^\alpha T) \ast \psi = (T \ast \partial^\alpha \psi),
$$

for any multi-index $\alpha$. Furthermore, if $T$ belongs to $\mathcal{S}'(\mathbb{R}^d)$ and $\psi$ belongs to $\mathcal{S}(\mathbb{R}^d)$ then $T \ast \psi$ and all its derivatives have at most polynomial growth.

**Proof.** The verification of inclusion $\text{supp}\{T \ast \psi\} \subseteq \text{supp}\{T\} + \text{supp}\{\psi\}$ is not completely obvious, the compactness of one of the supports is necessary. Based on this inclusion, cases (a) and (b) are deduced. The identity with the derivatives is essentially obtained by the continuity and linearity of the distributions.

If $T$ is a tempered distribution, i.e., and element in $\mathcal{S}'(\mathbb{R}^d)$ then for some constants $C$ and $n$ we have

$$
|(T \ast \varphi)(x)| = \langle T, \varphi(x - \cdot) \rangle \leq C \rho_n(\tau_x \varphi) =
\sup \left\{ (1 + |x - y|^2)^{n/2} |\partial^\alpha \varphi(y)| : y \in \mathbb{R}^d, |\alpha| \leq n \right\}.
$$

In view of the inequality

$$
(1 + |x - y|^2)^{n/2} \leq 2^{n/2}(1 + |x|^2)^{n/2}(1 + |y|^2)^{n/2},
$$

we conclude that $T \ast \psi$ has indeed polynomial growth. \qed

• **Remark 3.35.** It is clear that the convolution is a continuous operation from $\mathcal{D}(\mathbb{R}^d) \times \mathcal{D}(\mathbb{R}^d)$ into $\mathcal{D}(\mathbb{R}^d)$, from $\mathcal{D}(\mathbb{R}^d) \times \mathcal{E}(\mathbb{R}^d)$ into $\mathcal{E}(\mathbb{R}^d)$, and form $\mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d)$ into $\mathcal{S}(\mathbb{R}^d)$. Therefore, the convolution results a continuous operation within the situation (a), (b) and (c) of the previous Proposition 3.34 \qed

The following approximation result shows that $\mathcal{D}(\mathbb{R}^d)$ is dense in $\mathcal{D}'(\mathbb{R}^d)$,
Proposition 3.36. If \( k \) is a smooth kernel with compact support, i.e., an element in \( \mathcal{D}(\mathbb{R}^d) \) such that
\[
\int_{\mathbb{R}^d} k(x) \, dx = 1,
\]
and \( k_\varepsilon(x) = \varepsilon^{-d} k(x/\varepsilon) \), for every \( x \) in \( \mathbb{R}^d \) and \( \varepsilon > 0 \), then we have \( T \ast k_\varepsilon \to T \) in \( \mathcal{D}'(\mathbb{R}^d) \) as \( \varepsilon \to 0 \), i.e.,
\[
\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d} (T \ast k_\varepsilon)(x) \varphi(x) \, dx = \langle T, \varphi \rangle, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d).
\]
If \( T \) has a compact support (or \( T \) is a tempered distribution), i.e., \( T \) belongs to \( \mathcal{E}'(\mathbb{R}^d) \) (or to \( \mathcal{S}'(\mathbb{R}^d) \)) and the convergence result holds for \( \varphi \) in \( \mathcal{E}(\mathbb{R}^d) \) (or in \( \mathcal{S}(\mathbb{R}^d) \)). Moreover, if \( T \) is a continuous linear functional on \( C^n(\mathbb{R}^d) \) then a kernel in \( \mathcal{C}^n(\mathbb{R}^d) \) suffices, and the converges takes place in the corresponding topology.

**Proof.** Indeed, the linearity and continuity of the distribution \( T \) imply the equality
\[
\langle T \ast k_\varepsilon, \varphi \rangle = \int_{\mathbb{R}^d} (T \ast k_\varepsilon)(x) \varphi(x) \, dx = \int_{\mathbb{R}^d} \langle T, k_\varepsilon(x - \cdot) \rangle \varphi(x) \, dx = \langle T, k_\varepsilon - \cdot \rangle \varphi(x) \, dx = \langle T, \hat{k}_\varepsilon \ast \varphi \rangle,
\]
where \( \hat{k}(x) = k(-x) \). Next, Remark B.87 yields the desired result.

Let consider the case when \( T \) belongs to \( \mathcal{S}(\mathbb{R}^d) \). This means that what we should establish is that if \( \varphi \) is any rapidly decreasing smooth function in \( \mathbb{R}^d \), i.e., it belongs to \( \mathcal{S}(\mathbb{R}^d) \), then \( \varphi \ast k_\varepsilon \to \varphi \) in \( \mathcal{S}(\mathbb{R}^d) \).

The whole point is to get an estimate to justify taking limit inside the integral, namely, a sup in \( x \) of
\[
| (\varphi \ast k_\varepsilon)(x) - \varphi(x) | = \left| \int_{\mathbb{R}^d} [\varphi(x - \varepsilon y) - \varphi(x)] k(y) \, dy \right|,
\]
with a weight \( (1 + |x|^2)^{n/2} \), for any \( n \geq 0 \). In this respect, note that the inequality
\[
(1 + |x|^2)^{n/2} \leq 2^{n/2} (1 + |x - z|^2)^{n/2} (1 + |z|^2)^{n/2}
\]
implies
\[
(1 + |x|^2)^{n/2} \leq 2^{n/2} (1 + |x - t \varepsilon y|^2)^{n/2} (1 + |y|^2)^{n/2}, \quad \forall 0 \leq t, \varepsilon \leq 1,
\]
which yields
\[
|\varphi(x - \varepsilon y) - \varphi(x)| (1 + |x|^2)^{n/2} \leq \varepsilon \int_0^1 |y \cdot \nabla \varphi(x - t \varepsilon y)| 2^{n/2} (1 + |x - t \varepsilon y|^2)^{n/2} (1 + |y|^2)^{n/2} \, dt,
\]
i.e.,
\[
\sup_x \{|\varphi(x - \varepsilon y) - \varphi(x)| (1 + |x|^2)^{n/2}\} \leq 2^{n/2}\varepsilon \left( \sup_x \{ |\nabla \varphi(x)| (1 + |x|^2)^{n/2}\} \right) (1 + |y|^2)^{(n+1)/2}.
\]

Hence, because \(\varphi\) and \(k\) belong to \(S(\mathbb{R}^d)\), the sup
\[
C_n(\varphi) = 2^{n/2} \sup_x \{ |\nabla \varphi(x)| (1 + |x|^2)^{n/2}\}
\]
and the integral
\[
I_n(k) = \int_{\mathbb{R}^d} (1 + |y|^{2(n+1)/2}) |k(y)| \, dy
\]
are finite, and
\[
\sup_x \{|(\varphi \ast k_{\varepsilon})(x) - \varphi(x)| (1 + |x|^2)^{n/2}\} \leq \varepsilon C_n(\varphi) I_n(k),
\]
valid for any \(0 < \varepsilon \leq 1\).

Finally, replacing \(\varphi\) with any derivative \(\partial^\alpha \varphi\), we deduce that \(\varphi \ast k_{\varepsilon} \to \varphi\) in the topology of \(S(\mathbb{R}^d)\), as desired. \(\square\)

Therefore, the convolution of two distributions makes sense if one of them has compact support. Making visible the variables, e.g., \(\langle T, \varphi \rangle = \langle T_x, \varphi(x) \rangle\), we have
\[
\langle T \ast S, \varphi \rangle = \langle T_x, \langle S_y, \varphi(x+y) \rangle \rangle = \langle S_x, \langle T_y, \varphi(x+y) \rangle \rangle = \langle T \otimes S, \varphi^{\oplus} \rangle,
\]
where \(\varphi^{\oplus}(x, y) = \varphi(x+y)\) and \(\varphi\) is any element in \(D(\mathbb{R}^d)\). For instance, if \(T\) has compact support then the function \(x \mapsto \langle T_y, \varphi(x+y) \rangle\) belongs to \(D(\mathbb{R}^d)\) and so \(\langle S_x, \langle T_y, \varphi(x+y) \rangle \rangle\) is well defined. Similarly, the function \(x \mapsto \langle S_y, \varphi(x+y) \rangle\) belongs only to \(E(\mathbb{R}^d)\), but \(\langle T_x, \langle S_y, \varphi(x+y) \rangle \rangle\) is also well defined since \(T\) belongs to \(E'(\mathbb{R}^d)\). The relation with the tensor product is also clear by definition.

Recall that the translation operator initially defined in \(D(\mathbb{R}^d)\) by \(\tau_h \varphi(x) = \varphi(x-h)\) is also considered as an operator on \(D'(\mathbb{R}^d)\) by setting \(\langle \tau_h T, \varphi \rangle = \langle T, \tau_{-h} \varphi \rangle\), for any test function \(\varphi\). The convolutions operation can be characterized as follow:

**Proposition 3.37.** Let \(L\) be a linear continuous mapping from \(D(\mathbb{R}^d)\) into \(E(\mathbb{R}^d)\) which commutes with translations, i.e.,
\[
L \tau_h \varphi = \tau_h L \varphi, \quad \forall \varphi \in D(\mathbb{R}^d), \ h \in \mathbb{R}^d.
\]

Then there exists a uniquely determined element \(T\) in \(D'(\mathbb{R}^d)\) such that \(L \varphi = T \ast \varphi\), for every \(\varphi\) in \(D(\mathbb{R}^d)\). Conversely, \(\varphi \mapsto T \ast \varphi\) defines a linear continuous mapping from \(D(\mathbb{R}^d)\), or \(E(\mathbb{R}^d)\), or \(S(\mathbb{R}^d)\) into \(E(\mathbb{R}^d)\) if \(T\) belongs to \(D'(\mathbb{R}^d)\), or \(E'(\mathbb{R}^d)\) or \(S'(\mathbb{R}^d)\).
Proof. First, let $L$ be a linear continuous mapping from $\mathcal{D}(\mathbb{R}^d)$ into $\mathcal{E}(\mathbb{R}^d)$ which commutes with translations. Since the mapping $\varphi \mapsto \hat{\varphi}$, with $\hat{\varphi}(x) = \varphi(-x)$ is a continuous linear operation on $\mathcal{D}(\mathbb{R}^d)$, the linear map $\langle T, \varphi \rangle = L\hat{\varphi}(0)$ defines a distribution satisfying $L\varphi(0) = \langle T, \varphi \rangle = (T * \varphi)(0)$. Replacing $\varphi$ with $\tau_h \varphi$ and using the fact that $L$ commutes with $\tau_h$, we deduce $L\varphi(h) = (T * \varphi)(h)$.

The converse assertion follows from the equalities $(T * \varphi) * \psi = T * (\varphi * \psi)$ and $\langle T, \tau_h \varphi \rangle = (T * \varphi)(h)$. Indeed, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$, the function $\psi : (x, h) \mapsto \tau_h \varphi$ can be considered as an element in $\mathcal{D}(\mathbb{R}^d \times \mathbb{R}^d)$, and if $\varphi_k \rightarrow \varphi$ in $\mathcal{D}(\mathbb{R}^d)$ then $\psi_k \rightarrow \psi$ in $\mathcal{D}(\mathbb{R}^d \times \mathbb{R}^d)$, and if $\phi_k(h) = \langle T, \tau_h \varphi_k \rangle$ and $\phi(h) = \langle T, \tau_h \varphi \rangle$ then $\phi_k \rightarrow \phi$ in $\mathcal{D}(\mathbb{R}^d)$. Hence, if $T$ is in $\mathcal{D}'(\mathbb{R}^d)$ and a sequence $\varphi_k \rightarrow \varphi$ in $\mathcal{D}(\mathbb{R}^d)$ then the equalities $\langle T, \tau_h \varphi_k \rangle = (T * \varphi_k)(h)$ and $\langle T, \tau_h \varphi \rangle = (T * \varphi)(h)$ show that $T * \varphi_k \rightarrow T * \varphi$ in $\mathcal{D}'(\mathbb{R}^d)$.

Certainly, an argument similar to the above complete the proof when $T$ belongs to either $\mathcal{E}'(\mathbb{R}^d)$ or $\mathcal{S}'(\mathbb{R}^d)$.

Recall that $C^n_0$ denotes the lctvs of all $n$-times continuously differentiable functions with compact support on $\mathbb{R}^d$, which dual space $(C^n_0)'$ is the space of all distribution of order at most $n$. In particular, $C^n_0$ is the space of continuous functions with compact support and the elements in its dual space $(C^n_0)'$ are called local signed measure.

**Corollary 3.38.** A distribution $T$ is a local signed measure (a distribution of order at most $n$ or a function in $L^q_{loc}$) if and only if the convolution $T * \psi$ can be identified to a continuous (locally bounded measurable) function for any function $\psi$ in $C^n_0$ (in $C^n_0$ or in $L^p_{loc}$, with $1 \leq p < \infty$, $1/p + 1/q = 1$).

**Proof.** First note that only the ‘necessity’ should be proved, and if $\hat{\varphi}(x) = \varphi(-x)$ then $(T * \varphi)(0) = \langle T, \varphi \rangle$. Now, if $K$ is a compact in $\mathbb{R}^d$ and $B$ is an open ball of centered at the origin, then the linear operator $\psi \mapsto T * \psi$ from $C^n_0(K)$ into $L^\infty(K)$ is continuous. Indeed, if $T$ is a continuous functions this is certainly true, which means that the linear operator $\psi \mapsto T_{\varepsilon} * \psi$ is continuous, for any $\varepsilon$ as in Proposition 3.37. Thus, Banach-Steinhaus Theorem 2.17 applied to the family $\psi \mapsto T_{\varepsilon} * \psi$ of continuous linear operators proves that its weak limit $\psi \mapsto T * \psi$ as $\varepsilon \rightarrow 0$ is also a continuous linear operator from $C^n_0(K)$ into $L^\infty(K)$. Remark that Proposition 3.37 is used to check that $T_{\varepsilon} * \psi \rightarrow T * \psi$, for any test function $\psi$.

Next, let us verify that $T * \psi$ is not only locally bounded, but it is actually continuous, for every $\psi$ in $C^n_0$. Indeed, it is clear that the assumption implies the continuity if $\psi$ is a test function. Therefore, use the assertion that $\psi \mapsto T * \psi$ is a continuous operator for the locally uniform convergence to check that if a sequence of test function $\psi_i \rightarrow \psi$ locally uniformly then $T * \psi_j \rightarrow T * \psi$ locally uniformly, which proves that $T * \psi$ is a continuous function, for any $\psi$ in $C^n_0$. Hence, $\psi \mapsto \langle T, \psi \rangle = (T * \psi)(0)$ is a local signed measure, i.e., an element of the dual space $(C^n_0)'$.

Finally, replace $C^n_0(K)$ with $C^n_0(K)$ or with $L^p(K)$ and redo the above argument to complete the proof. □
• **Remark 3.39.** Except for dimension $d = 1$, if a distribution $T$ is such that $T \ast \psi$ is a $n$-times continuously differentiable function for any function $\psi$ in $C_0^\infty$ then $T$ is not necessarily a locally signed measure, see Ornstein [97]. This is in contrast with the true assertion that if a distribution $T$ has all its first-order partial derivatives identifiable with functions in $L^p_{\text{loc}}(\mathbb{R}^d)$, $p > d$, then $T$ can also be identified with a continuous function, for instance the reader may check Schwartz [112, Section VI.6, Theorem XV, pp.181–184]. Certainly, this is discussed in more detail in the section about Sobolev spaces.

The convolution of two distributions $T$ and $S$ is also defined if for instance both $T$ and $S$ have supports in $[0, \infty)^d$, and in this case, $T \ast S$ will have also support in $[0, \infty)^d$. Based on the convolution, we have

**Proposition 3.40.** If $T$ is a distribution, i.e., $T$ belongs to $\mathcal{D}'(\Omega)$, such that $\partial^\alpha T$ is a locally integrable function, i.e., $\partial^\alpha T$ belongs to $L^1_{\text{loc}}(\Omega)$, for every multi-index $\alpha$ then $T$ is indeed a smooth function, i.e., $T$ belongs to $C^\infty(\Omega)$. Moreover, for every compact $K$ in $\Omega$ and for any $\varepsilon > 0$ strictly less than the distance from $K$ to the boundary $\partial \Omega$, there exists a constant $C_\varepsilon$ depending only on the constant $\varepsilon > 0$ and the dimension $d$ such that

$$\text{ess-sup}_{x \in K} |T(x)| \leq C_\varepsilon \sup_{|\alpha| \leq d} \left\{ \int_{K_\varepsilon} |\partial^\alpha T(x)| \, dx \right\},$$

where $K_\varepsilon = \{x : d(x, K) \leq \varepsilon\}$, with $d(x, K)$ being the distance from $x$ to $K$.

**Proof.** First, given a smooth kernel $k$ as in Proposition 3.36 and a function $f$ in $L^1_{\text{loc}}(\Omega)$, recall that the convolution $f \ast k_\varepsilon \to f$ in $L^1_{\text{loc}}(\Omega)$. Hence, if $\partial^\alpha T$ belongs to $L^1_{\text{loc}}(\Omega)$ then $\partial^\alpha (T \ast k_\varepsilon) = (\partial^\alpha T) \ast k_\varepsilon \to (\partial^\alpha T)$ in $\mathcal{D}'(\Omega)$ and in $L^1_{\text{loc}}(\Omega)$.

Also note that, for any smooth function $f$ with a compact support, the equality

$$f(x_1, \ldots, x_i, \ldots, x_d) = \int_{-\infty}^{x_i} \partial_{i} f(x_1, \ldots, t_i, \ldots, x_d) \, dt_i = \int_{-\infty}^{x_1} \, dt_1 \ldots \int_{-\infty}^{x_i} \, dt_i \ldots \int_{-\infty}^{x_d} \partial_1 \ldots \partial_d f(t_1, \ldots, t_i, \ldots, t_d) \, dt_d,$$

yields the estimate

$$\text{ess-sup}_{x \in \mathbb{R}^d} |f(x)| \leq \int_{\mathbb{R}^d} |\partial^d f(x)| \, dx, \quad \text{with} \quad \partial^d = \partial_1 \ldots \partial_d,$$

for any smooth function $f$ with compact support.

Hence, if $\partial^\alpha T$ is locally integrable for every multi-index $\alpha$ then the same holds true for the distribution $\chi T$ in $\mathcal{E}(\mathbb{R}^d)$ for any smooth function $\chi$ with compact support, moreover, now $\partial^\alpha (\chi T)$ is integrable in $\mathbb{R}^d$. Thus, apply the previous estimate to the functions $\partial^\alpha (\chi T) \ast k_\varepsilon$ to deduce that all the derivatives $\partial^\alpha (\chi T)$ are actually essentially bounded functions. Therefore, they are Lipschitz functions, actually, i.e., $\partial^\alpha (\chi T)$ are smooth functions with compact support, i.e., $\partial^\alpha T$ is in $C^\infty(\Omega)$.
Remark that the convolution of a characteristic functions with a smooth compactly support kernel provide the suitable cutting functions $\chi$, i.e., for any given a compact $K$ and constant $0 < \varepsilon < d(K, \Omega)$ there exists a smooth function $\chi$ such that (a) $\chi = 1$ on $K$, (b) $\chi = 0$ on $\Omega \setminus K_{\varepsilon}$ and (c) $|\partial^\alpha \chi| \leq C_{|\alpha|, d} \varepsilon^{-|\alpha|}$, for any multi-index $\alpha$ and some constant $C_d$ depending only on the order of derivative $|\alpha|$ and the dimension $d$.

Therefore, after choosing $\chi = 1$ on the compact $K$ and using Leibniz's rule for the derivatives of a product of functions, the desired estimate follows and the proof is completed.

\textbf{Remark 3.41.} When $\Omega = \mathbb{R}^d$ the convolution with a smooth and compactly supported kernel yields the construction of a sequence of cutting functions satisfying (a) $\chi_n(x) = 1$ if $|x| \leq n$, (b) $\chi_n(x) = 0$ if $|x| \geq n + 1$, (c) $|\partial^\alpha \chi_n(x)| \leq C_{|\alpha|, d}$ for any $x$, any multi-index $\alpha$, and some constant depending only on the order of derivative $|\alpha|$ and the dimension $d$. Hence, the argument of Proposition 3.40 with $\varphi \chi_n$ instead of just $\chi$ implies the following estimate: If $T$ is a distribution in $\mathbb{R}^d$ such that $\partial^\alpha T$ is an integrable function, for every multi-index $\alpha$, then $T$ is a smooth function and

$$\sup_{\mathbb{R}^d} |\varphi T| \leq C_d \sup_{|\alpha| \leq d} \left\{ \int_{\mathbb{R}^d} |\partial^\alpha (\varphi(x) T(x))| \, dx \right\}, \quad \forall \varphi \in C^\infty(\mathbb{R}^d),$$

for a constant $C_d$ depending only on the dimension $d$. In particular, take $\varphi = (1 + |x|^2)^{-k/2}$ with $k > d$, to obtain

$$\sup_{\mathbb{R}^d} \left| (1 + |x|^2)^{-k/2} T \right| \leq C_{d,k} \sup_{|\alpha| \leq d} \left\{ \int_{\mathbb{R}^d} |\partial^\alpha T(x)| \, dx \right\},$$

for some suitable constant depending only on the $d$ and $k$. Moreover, using Hölder inequality and $k$ sufficiently large, the $L^1$-norm of the $\partial^\alpha T$ can be replaced by the $L^p$-norm. This shows that if $T$ belongs to $\mathcal{D}'(\mathbb{R}^d)$ is such that $\partial^\alpha T$ can be identify to an $p$-integrable function in $\mathbb{R}^d$, for every multi-index $\alpha$, then $T$ and all its derivatives are smooth function with a polynomial growth.

\textbf{Exercise 3.24.} Recall the differential operator $\Delta = \sum_{i=1}^d \partial_i^2$. A fundamental distributional solution associated with the iterated Laplacian $\Delta^k$ is a distribution $E = E_{kd}$ on $\mathbb{R}^d$ such that $\Delta^k (E \ast \delta) = \delta$, where $\delta$ is the Dirac delta measure, $\langle \delta, \varphi \rangle = \varphi(0)$. Verify that $E = |x|^{2k-d} (a_{kd} \ln |x| + b_{kd})$ is a fundamental distributional solution associated $\delta^k$ in $\mathbb{R}^d$, where one of the constants $a_{kd}$ or $b_{kd}$ vanishes, namely, if $2k - d < 0$ or $d$ is odd then $a_{kd} = 0$, and otherwise $b_{kd} = 0$. Note that if $2k - d > 0$ then $E$ belongs to $C^{2k-d-1}$ and complete the following argument. First, consider a distribution $T$ with compact support and verify that $T = E \ast (\Delta T)$. Next, if $\Delta^k T$ is a distribution of order $n$ (i.e., it belongs to the dual space of $C^n$) with a compact support and $2k - d - 1 \geq n$ then $T$ is the distribution associated to the function $x \mapsto \langle \Delta^k T, E(x - \cdot) \rangle$, and therefore $T$ belongs to $C^n$.

\textbf{Exercise 3.25.} Verify the correctness of the following examples of convolutions:
3.3. More Operations and Localization

a.- Riesz potentials: $R_\alpha$, $0 < \alpha < d$, and for any $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$,

$$(-\Delta)^{-\alpha/2} \varphi(x) = R_\alpha \ast f(x) = C_{\alpha,d} \int_{\mathbb{R}^d} |x-y|^{-(d-\alpha)} \varphi(y) dy,$$

where $C_{\alpha,d} = \Gamma((d-\alpha)/2)/[2^{\alpha} \pi^{d/2} \Gamma(\alpha/2)]$ is a normalizing constant.

b.- Calderon-Zygmund integro-differential operator, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$,

$$(-\Delta)^{-1/2} \varphi(x) = C_d \sum_{i=1}^d \lim_{\varepsilon \to 0} \int_{|x| \geq \varepsilon} (x_i - y_i)|x-y|^{-d-1}\partial_i \varphi(y) dy,$$

where $C_d = \Gamma((d+1)/2)\pi^{-(d+1)/2}$ is again a normalizing constant. Note the singular integral and recall that the limit is called the principal value of the integral.

c.- The Newtonian potential for $d \geq 3$ is defined by

$$(-\Delta)^{-1} \varphi(x) = (N \ast \varphi)(x) = \frac{1}{(d-2)\omega_d} \int_{\mathbb{R}^d} |x-y|^{2-d} \varphi(y) dy,$$

where $\omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of the unit sphere. For $d=2$ ($d=1$) we use the kernel $(1/2\pi) \ln(|x-y|) (|x|/2)$. *If $\Delta = \partial_1^2 + \cdots + \partial_d^2$ is the usual Laplacian then verify that $\Delta(N \ast \varphi)(x) = 0$ for every $x$ in $\mathbb{R}^d$.

d.- Double layer potential, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^{d-1})$, with $x = (x', x_d)$

$$N \ast (\varphi(x') \otimes \delta'(x_d)) = \frac{1}{\omega_d} \int_{\mathbb{R}^{d-1}} x_d \varphi(y') (|x' - y'|^2 + x_d^2)^{-d/2} dy'.$$

*Verify that $u(x', x_d) = 2N \ast (\varphi(x') \otimes \delta'(x_d))$, which is called the Poisson integral formula, yields a solution of the Dirichlet problem $\Delta u = 0$ in $\mathbb{R}^d$ and $u(\cdot, 0) = \varphi$ in $\mathbb{R}^{d-1}$.

e.- Single layer potential, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^{d-1})$, with $x = (x', x_d)$

$$N \ast (\varphi(x') \otimes \delta(x_d)) = \frac{1}{(d-2)\omega_d} \int_{\mathbb{R}^{d-1}} \varphi(y') (|x' - y'|^2 + x_d^2)^{1-d/2} dy'.$$

*Verify that the $\partial_d$ of the single layer potential is equal to double layer potential.

Questions marked with * could not be so simple. The reader may check the book by Stein [113] for a detail account of Singular Integrals.

\[ \square \]

**Exercise 3.26.** Let $\{T_k\}$ be a sequence of distribution converging to 0 in $\mathcal{D}'(\mathbb{R}^d)$, and let $S$ be another distribution. Prove the if either (a) $S$ has a compact support or (b) the supports of $\{T_k\}$ are contained in a fixed compact set, then $S \ast T_k \to 0$ in $\mathcal{D}'(\mathbb{R}^d)$.

\[ \square \]
3.3.3 Local Structure

It has been established that the derivatives (of any order) of a function in $L^1_{\text{loc}}$ can be considered as a distribution or generalized function. The converse of this assertion is expressed as follows:

**Theorem 3.42.** Let $T$ be an element in $\mathcal{D}'(\Omega)$ and $K$ be a compact subset of the open domain $\Omega \subset \mathbb{R}^d$. Then there exists a positive integer $n = n(T, K)$ and a function $f = f(x; T, K)$ in $L^2(K)$ such that

$$\langle T, \varphi \rangle = (-1)^{|\beta|} \int_K f(x) \partial^\beta \varphi(x) \, dx, \quad \forall \varphi \in \mathcal{D}(\Omega), \quad \text{supp} \varphi \subset K,$$

where the multi-index $\beta = (\beta_1, \ldots, \beta_d)$, with $\beta_i = n$ and order $|\beta| = nd$, i.e., $T = \partial^\beta f$ in $\mathcal{D}'_K(\Omega)$.

**Proof.** Since the restriction of $T$ to $\mathcal{D}'_K(\Omega)$ is a continuous functional, there exist a positive integer $n$ and a constant $C_1 > 0$ such that

$$|\langle T, \varphi \rangle| \leq C_1 \sup \{ |\partial^\alpha \varphi(x) : x \in K, |\alpha| \leq n - 1 \}, \quad \forall \varphi \in \mathcal{D}_K(\Omega).$$

Now, because $K$ is compact there exist $a \geq 1$ such that $|x| \leq a$ for every $x$ in $K$. Hence, by iteration of the inequality

$$|\psi(x)| \leq \int_{-a}^{x_1} \left| \partial_1 \psi(t, x_2, \ldots, x_d) \right| \, dt \leq (2a) \int_{-a}^{x_1} \left| \partial_1 \partial_2 \cdots \partial_d \psi(t, x_2, \ldots, x_d) \right|^2 \, dt,$$

for every $x$ in $K$, we deduce

$$|\psi(x)| \leq \int_{\{|x| \leq a\}} \left| \partial_1 \partial_2 \cdots \partial_d \psi(x) \right| \, dx \leq (2a)^{d/2} \left\| \partial_1 \partial_2 \cdots \partial_d \psi \right\|_2,$$

where $\| \cdot \|_2$ is the norm in the Hilbert space $L^2(K)$. Moreover,

$$|\partial^\alpha \varphi(x)| \leq \int_{\{|x| \leq a\}} \left| \partial_1 \partial_2 \cdots \partial_d \partial^\alpha \varphi(x) \right| \, dx \leq (2a)^{nd} \left\| \partial_1 \partial_2 \cdots \partial_d \partial^\alpha \varphi \right\|_2,$$

for every multi-index $|\alpha| \leq n - 1$.

The mapping $\psi \mapsto \varphi$, with $\psi = (-1)^{|\beta|} \partial^\beta \varphi$, is one-to-one from $\mathcal{D}_K(\Omega)$ into itself and so, the previous estimates show that $\psi \mapsto \langle T, \varphi \rangle$, with $\psi = (-1)^{|\beta|} \partial^\beta \varphi$, can be extended (in view of Hahn-Banach Theorem 2.26) to a linear continuous functional on $L^2(K)$, which is initially defined for any $\psi$ belonging to a subspace of $\mathcal{D}_K(\Omega) \subset L^2(K)$.

Therefore, by Riesz’ representation theorem, there exists a function $f$ in $L^2(K)$ such that

$$\langle T, \varphi \rangle = \int_K f(x) \psi(x) \, dx,$$

i.e., the desired representation.
By means of Theorem 3.42, it is clear that if \( T \) is a distribution with compact support then the equality \( T = \partial^\beta f \) holds, for some multi-index \( \beta \) and some locally square integrable (actually continuous, see next result) function \( f \).

The context of the following result include: (a) partial derivatives of rank at most 1, i.e., distribution sense partial derivatives \( \partial^\alpha \) for any multi-index \( \alpha = (\alpha_1, \ldots, \alpha_d) \) with \( \alpha_i = 0 \) or \( \alpha_i = 1 \) for any \( i = 1, \ldots, d \); (b) the Heavidside’s function \( H \) on \( \mathbb{R}^d \), i.e., \( H(x) = 1 \) if \( x_i \geq 0 \) for every \( i = 1, \ldots, d \), and \( H(x) = 0 \) otherwise; (c) the Dirac measure \( \delta \) on \( \mathbb{R}^d \), i.e., \( \langle \delta, \varphi \rangle = \varphi(0) \) for every test function \( \varphi \); (d) locally bounded variation functions in \( \mathbb{R}^d \), i.e., function of the form \( \mu \ast H \), where \( \mu \) is a locally signed measure meaning a distribution of order zero in \( \mathbb{R}^d \), and \( H \) is Heavidside’s function; (e) absolutely continuous functions in \( \mathbb{R}^d \), i.e., locally bounded variation functions in \( \mathbb{R}^d \) where the locally signed measure \( \mu \) is absolutely continuous with respect to the Lebesgue measure.

**Proposition 3.43.** If \( T \) is an element in \( \mathcal{D}'(\mathbb{R}^d) \) such that all its partial derivatives \( \partial^\alpha T \), of rank at most 1, are either locally integrable functions or locally signed measures, then \( T \) is either an absolutely continuous function or a locally bounded functions with locally bounded variation.

**Proof.** To study the given distribution \( T \) on a relative compact open set \( \Omega \) of \( \mathbb{R}^d \), simply use a smooth cutting function \( \chi \) in \( \mathbb{R}^d \) satisfying \( \chi = 1 \) on a neighborhood of \( \Omega \) and replace the distribution \( T \) with the product \( \chi T \), which becomes a distribution of compact support and therefore of finite order. Indeed, Leibnitz formula for the \( n \)-order derivatives of a product of smooth functions shows that if all partial derivatives \( \partial^\alpha T \), of rank at most 1, are either locally integrable functions or locally signed measures then it is the same for the distribution \( \chi T \).

Therefore, it is sufficient to prove the conclusion of this proposition under the condition \( T \) is a distribution with compact support satisfying \( \partial_1 \ldots \partial_d T = \mu \), where \( \mu \) is either locally integrable function or locally signed measure. Thus, to this purpose, use the equality \( \partial_1 \ldots \partial_d H = \delta \) the Dirac measure to find that

\[
\mu \ast H = \partial_1 \ldots \partial_d T \ast H = T \ast \partial_1 \ldots \partial_d H = T \ast \delta = T,
\]

which certainly proves the desired conclusion. \( \square \)

**Remark 3.44.** The previous Proposition 3.43 makes clear that the function \( f \) that locally represent \( T \) in Theorem 3.42 could be taken in any class \( C^k_0(\mathbb{R}^d) \), it suffices to increase the order of the multi-index \( \beta \). Actually, for any distribution \( T \) of finite order and any nonnegative integer \( k \) there exists a function \( f \) in \( C^k(\mathbb{R}^d) \) and a multi-index \( \beta \) such that \( T = \partial^\beta f \) holds true. Indeed, the arguments in Theorem 3.42 shows that if \( T \) is a distribution of finite order, say \( m(T) \) then for any compact \( K \) there exists a locally square integrable function \( f_K \) and an index \( \beta = \beta(m) \) (depending only on the order \( m \leq m(T) \) of \( T \) on \( K \)) such that \( T = \partial^\beta f \) holds true. Therefore, if \( T \) is a distribution of finite order \( m = m(T) \) and \( \{\chi_i\} \) is a locally finite partition of the unity on \( \mathbb{R}^d \) of test functions, then \( T = \sum_i T_i \), with \( T_i = \chi_i T \), and each \( T_i \) has a compact support and order at most \( n \). Hence, take \( f = \sum_i f_i \), where \( f_i \) is a suitable function
corresponding to $T_i$, so that $\beta = \beta(n)$ is kept fixed to find the equality $T = \partial^\beta f$ holds true.

- **Remark 3.45.** Proposition 3.43 does not imply that if all first-order partial derivatives of a distribution $T$ are locally integrable functions then $T$ is also a locally integrable function. However, if a distribution $T$ is such that there exists a locally integrable function $f$ satisfying the equation $\partial^\alpha f = \partial^\alpha T$ in the distribution sense, for any $\alpha$ of order 1, then $T$ is equal to $f$ plus a constant. □

- **Remark 3.46.** As in Theorem B.88, let $\{\chi_i : i \geq 1\}$ be partition of the unity $\{x_i : i \geq 1\}$ subordinate to an open cover $\{\Omega_\alpha : \alpha\}$ of an open subset $\Omega$ of $\mathbb{R}^d$, i.e., (a) $\chi_i$ belongs to $C^\infty_0(\mathbb{R}^d)$, (b) for every $i$ there exists $\alpha = \alpha(i)$ such that $\chi_i(x) = 0$ for every $x$ in $\Omega \setminus \Omega_\alpha$, (c) $0 \leq \chi_i(x) \leq 1$ and $\sum_i \chi_i(x) = 1$, for every $x$ in $\Omega$, where the series is locally finite, namely, for any compact set $K$ of $\Omega$ the set of indices $i$ such that the support of $\chi_i$ intercept $K$, $\text{supp}(\chi_i) \cap K \neq \emptyset$, is finite. Note that by construction, $\chi_i$ belongs to $\mathcal{D}(\Omega_{\alpha(i)})$ but $\sum_i \chi_i$ does not belong to $\mathcal{D}(\Omega)$. Also, it is clear that if $T$ belongs to $\mathcal{D}'(\Omega)$ then $T = \sum_i (T\chi_i)$ and each $T\chi_i$ belongs to $\mathcal{D}'(\Omega_{\alpha(i)})$. Moreover, the converse is true, namely, if $\{T_i\}$ is a countable family of elements in $\mathcal{D}(\Omega_{\alpha(i)})$ with the property that $T_i = T_j$ on the intersection set $\Omega_{\alpha(i)} \cap \Omega_{\alpha(j)}$ then the expression $\langle T, \varphi \rangle = \sum_i \langle T_i, \chi_i\varphi \rangle$ defines a distribution $T$ on $\Omega$, i.e., we can “reconstruct” $T$ from the pieces $T_i$. A proof follows immediately from the fact that a linear functional $T$ on $\mathcal{D}(\Omega)$ is a distribution if and only if for every compact $K$ of $\Omega$ there exists a constant $C > 0$ and an index $n$ such that

$$|\langle T, \varphi \rangle| \leq C \sup \{|\partial^\alpha \varphi(x)| : x \in K, |\alpha| \leq n\}, \quad \forall \varphi \in \mathcal{D}_K(\Omega).$$

This is usually referred to as the principle of localization. □

### 3.3.4 Recap on Inductive Limits

Let us discuss a little more the topology on the spaces $\mathcal{D}(\Omega)$ and $\mathcal{E}(\Omega)$ as well as on their dual spaces $\mathcal{D}'(\Omega)$ and $\mathcal{E}'(\Omega)$. A key role is played by the seminorms

$$p_{n,K}(u) = \sup \{|\partial^\alpha u(x)| : x \in K, |\alpha| \leq n\},$$

for $n = 0, 1, \ldots$ and a subset $K$ of $\Omega$. Both are sequentially complete lctvs, the space $\mathcal{D}(\Omega)$ is not metrizable, i.e., the topology is given by an uncountable family of seminorms. However, $\mathcal{E}(\Omega)$ is metrizable, i.e., the topology is given by a sequence of seminorms, say a Fréchet space. Both spaces are separable and both have the Heine-Borel property, i.e., every closed and bounded set is compact (some times this is referred as being a perfect space or a Montel space. Moreover, both spaces are barrel lctvs, i.e., every convex, balanced, absorbing and closed set is a neighborhood of zero.

- **On $\mathcal{E}(\Omega)$:** (1) A sequence $\{\varphi_k\}$ converges to zero if and only if $p_{n,K}(\varphi_k) \to 0$, for every $n$ and any compact $K$ of $\Omega$. (2) A set $B$ is bounded if and only if for every $n$ and any compact $K$ of $\Omega$ there is a constant $C = C(n,K,B)$ such that $p_{n,K}(v) \leq C$ for every $v$ in $B$. [Preliminary]
3.3. More Operations and Localization

- On \(\mathcal{D}(\Omega)\): (1) A sequence \(\{\varphi_k\}\) converges to zero if and only if (a) there exists a compact set \(K\) of \(\Omega\) such that \(\varphi_k(x) = 0\) for every \(x\) in \(\Omega \setminus K\) and (b) \(p_{n,k}(\varphi_k) \to 0\), for every \(n\). (2) A set \(B\) is bounded if and only if (a) there exists a compact set \(K\) of \(\Omega\) such that \(\varphi_k(x) = 0\) for every \(x\) in \(\Omega \setminus K\) and (b) for every \(n\) there exists a constant \(C = C(n,B)\) such that \(p_{n,k}(v) \leq C\) for every \(v\) in \(B\).

Then, it is clear that if \(\varphi\) belongs to \(\mathcal{D}(\Omega)\) and \(\psi\) belongs to \(\mathcal{E}(\Omega)\) then the product \(\varphi\psi\) belongs to \(\mathcal{D}(\Omega)\). Thus, if an increasing sequence \(\{\varphi_k\}\) in \(\mathcal{D}(\Omega)\) satisfies \(\varphi_k \to 1\) in \(\mathcal{E}(\Omega)\) then \(\varphi_k\psi \to \psi\) in \(\mathcal{E}(\Omega)\), i.e., \(\mathcal{D}(\Omega)\) is dense in \(\mathcal{E}(\Omega)\).

For each compact subset \(K\) of \(\Omega\), the subspace
\[
\mathcal{D}_K(\Omega) = \{ \varphi \in \mathcal{D}(\Omega) : \text{supp}(\varphi) \subset K \}
\]
endowed with the topology induced by the countable family of seminorms \(\{p_{n,K} : n = 0, 1, \ldots\}\) becomes a complete metrizable lctvs, i.e., a Fréchet spaces. These subspaces yields the inductive limit topology on \(\mathcal{D}(\Omega)\). Again, this means that (1) a sequence \(\{\varphi_k\}\) converges to zero if and only if the sequence converges to zero in some \(\mathcal{D}_K(\Omega)\), i.e., (a) there exists a compact set \(K\) of \(\Omega\) such that \(\{\varphi_k\} \subset \mathcal{D}_K(\Omega)\) and (b) \(\varphi_k \to 0\) in \(\mathcal{D}_K(\Omega)\); (2) a set \(B\) is bounded if and only if \(B\) is bounded in some \(\mathcal{D}_K(\Omega)\), i.e., (a) there exists a compact set \(K\) of \(\Omega\) such that \(B \subset \mathcal{D}_K(\Omega)\) and (b) \(B\) is a bounded set in \(\mathcal{D}_K(\Omega)\).

A linear functional \(T\) on \(\mathcal{D}(\Omega)\) is a distribution, i.e., it belongs to \(\mathcal{D}'(\Omega)\) if and only if its restriction \(T \big|_K\) to \(\mathcal{D}_K(\Omega)\), on any compact subset \(K\) of \(\Omega\), is continuous, i.e., for every \(K\) there exist an index \(n = n(K,T)\) and a constant \(C = C(K,T)\) such that \(|\langle T, \varphi \rangle| \leq C p_{n,K}(\varphi)\) for every \(\varphi\) in \(\mathcal{D}_K(\Omega)\). If the index \(n = n(T)\) can be chosen independent of \(K\) then the distribution \(T\) is said to be of finite order. In contrast, the linear functional \(T\) can be extended to be an element of \(\mathcal{E}(\Omega)\) if and only if there exists a compact subset \(K = K(T)\) of \(\Omega\) and an index \(n = n(T)\) such that \(|\langle T, \varphi \rangle| \leq C p_{n,K}(\varphi)\) for every \(\varphi\) in \(\mathcal{D}(\Omega)\). The elements of \(\mathcal{E}'(\Omega)\) are distributions with compact support.

The dual spaces \(\mathcal{D}'(\Omega)\) and \(\mathcal{E}'(\Omega)\) are endowed with the uniform convergence on bounded sets, i.e., the lctvs given by the dual seminorms \(p'_B(T) = |\langle T, \varphi \rangle|\), for any bounded set \(B\). This means that a sequence \(\{T_k\}\) converges to zero if and only if for any bounded set \(B\) the numerical sequence \(\langle T_k, \varphi \rangle \to 0\) uniformly in \(\varphi\) in \(B\). Actually, because each bounded set \(B\) is relatively compact, the “strong” dual topology as above is equivalent to the “weak star” dual topology (or pointwise convergence), namely, given by the family of seminorms \(T \mapsto |\langle T, \varphi \rangle|\), for \(\varphi\) ranging over either \(\mathcal{D}(\Omega)\) or \(\mathcal{E}(\Omega)\). Both dual spaces are complete lctvs. This is

- On \(\mathcal{E}'(\Omega)\): (1) A sequence \(\{T_k\}\) converges to zero if and only if \(\langle T_k, \varphi \rangle \to 0\) for every \(\varphi\) in \(\mathcal{E}(\Omega)\). (2) A set \(B'\) is bounded if and only if for every \(\varphi\) in \(\mathcal{E}(\Omega)\) there exists a constant \(C = C_\varphi(B')\) such that \(|\langle T, \varphi \rangle| \leq C_\varphi\) for every \(T\) in \(B'\).

- On \(\mathcal{D}'(\Omega)\): (1) A sequence \(\{T_k\}\) converges to zero if and only if \(\langle T_k, \varphi \rangle \to 0\) for every \(\varphi\) in \(\mathcal{D}(\Omega)\). (2) A set \(B'\) is bounded if and only if for every \(\varphi\) in \(\mathcal{D}(\Omega)\) there exists a constant \(C = C_\varphi(B')\) such that \(|\langle T, \varphi \rangle| \leq C_\varphi\) for every \(T\) in \(B'\).
The dual spaces $\mathcal{E}'(\Omega)$ and $\mathcal{D}'(\Omega)$ are both barrel spaces and the inclusions

$$\mathcal{D}(\Omega) \subset \mathcal{E}(\Omega), \quad \text{and} \quad \mathcal{E}'(\Omega) \subset \mathcal{D}'(\Omega),$$

and using the natural embedding $L^1_{\text{loc}}(\Omega) \subset \mathcal{D}'(\Omega)$, also

$$\mathcal{E}(\Omega) \subset \mathcal{D}'(\Omega), \quad \text{and} \quad \mathcal{D}(\Omega) \subset \mathcal{E}'(\Omega),$$

all inclusions are continuous and dense (recall that $\mathcal{E}'(\Omega)$ are the distributions with compact support). Both spaces are reflexive, i.e., the bidual reproduce the initial space, $\mathcal{D}''(\Omega) = \mathcal{D}(\Omega)$ and $\mathcal{E}''(\Omega) = \mathcal{E}(\Omega)$. The convergence in $\mathcal{E}(\Omega)$ satisfies: (1) a sequence $\{T_k\}$ converges to zero if and only if (a) there exists a compact set $K$ of $\Omega$ such that the support of $T_k$ are included in $K$ and (b) $\langle T_k, \varphi \rangle \to 0$ for every $\varphi$ in $\mathcal{D}(\Omega)$; (2) a set $B'$ is bounded if and only if (a) there exists a compact set $K$ of $\Omega$ such that the support of $T_k$ are included in $K$ and (b) for every $\varphi$ in $\mathcal{D}(\Omega)$ there exists a constant $C = C_\varphi(B')$ such that $|\langle T, \varphi \rangle| \leq C_\varphi$ for every $T$ in $B'$.

It should be clear that “inductive limit” is only used for the space $\mathcal{D}(\Omega)$. When $\Omega = \mathbb{R}^d$ then $\mathcal{S} = \mathcal{S}'(\mathbb{R}^d)$ is a Fréchet space and its dual space $\mathcal{S}' = \mathcal{S}''(\mathbb{R}^d)$ is the space of tempered distributions. As discussed later, the space $\mathcal{S}$ can be presented as an intersection of a sequence $\{H_n\}$ of Hilbert spaces, with inclusion dense, continuous and compact (i.e., the norm $\| \cdot \|_n$ is dominated by the norm $\| \cdot \|_{n+1}$ and a bounded set in $H_n$ is pre-compact in $H_{n+1}$), which yields the inclusions

$$\mathcal{S} = \bigcap_n H_n \subset H_n \subset H_0 = H_0' \subset H_n' \subset \mathcal{S}'$$

as being continuous and dense. Certainly, $\mathcal{S}$ and $\mathcal{S}'$ are both Montel spaces and reflexive.

An easier situation is the space $\mathbb{R}^\infty$ of all real-valued (or complex-valued) sequences $u = \{u_i : i \geq 1\}$, which is thought as the union of Euclidean spaces, i.e., $\bigcup_k \mathbb{R}^k$ (or $\bigcup_k \mathbb{C}^k$), with the sup-norm $p(u) = \sup_i |u_i|$ and the seminorms $p_k(u) = \sup_{i \leq k} |u_i|$ and $q_k(u) = \sup_{i > k} |u_i|$. Note that $p(u)$ or $q_k(u)$ may be infinite, but they satisfy $q_k(u) = 0$ for every $u$ in $\mathbb{R}^k$, $q_k(u) > 0$ for every $u$ in $\mathbb{R}^\infty \setminus \mathbb{R}^k$, and $p_k^\infty(u) + q_k(u) = p(u)$ for any $u$ in $\mathbb{R}^\infty$. This space becomes a Fréchet space with the topology induced by the sequence of seminorms $\{p_k : k \geq 1\}$, which the same as the (infinite) product topology. Actually, this space can be viewed also as the space of polynomials and it could be denoted by $c$.

Other topologies can be used on subspaces of $\mathbb{R}^\infty$, for instance the Banach space $\ell^p$ of all elements in $\mathbb{R}^\infty$ with a finite $p$-norm $|u|_p = (\sum_k |u_i|^p)^{1/p}$, for some $1 \leq p \leq \infty$. Certainly, there are many other topologies that can be used on subspace of sequences, e.g., convergent sequences (the space $c$) and convergent sequences to zero (the space $c_0$), and all the connections with the so-called Schauder basis in a Banach space. Certainly, the dual space of $\ell^p$, $1 \leq p < \infty$ is $\ell^{q}$, with $1/p + 1/q = 1$, while the dual space of $\ell^\infty$ is more complicated, for instance, see Lindenstrauss and Tzafriri [83, Part I, Chapters 1–4, pp. 1–179].
Consider the subspace $s$ of all rapidly deceasing sequences, i.e., $u = \{u_i : i \geq 1\}$ satisfying $i^k|u_i| \to 0$ for every $k = 0, 1, \ldots$, which is a Fréchet space with the seminorms (weighted sup-norms) $p_k(u) = \sup_i\{i^k|u_i|\}$. This space is of particular interest since it has a Hilbertian structure, i.e., we can easily check that the scalar or inner products $(u, v)_k = \sum_i i^{2k}u_iv_i$ yield norms $\bar{p}_k(u) = \sqrt{(u, u)_k}$ and the locally convex topologies defined by either $\{p_k : k \geq 0\}$ or $\{\bar{p}_k : k \geq 0\}$ are equivalent. Moreover, if $h_k$ denotes the subspace of all sequences $u$ satisfying $\sqrt{(u, u)_k} < \infty$ (i.e., the space $\ell_2$ above) then $h_k$ is a Hilbert space and $s = \bigcap_k h_k$.

The (proper) inductive limit topology applied to $\mathbb{R}^\infty$ yields the subspace $\mathfrak{d}$ (lower case “eufrak” $d$, which could be denoted by $\mathbb{R}_0^\infty$, with the subindex 0 for “compact support”) of all sequences with only a finite number of non-zero terms, i.e., the space $\mathfrak{d}_k = \mathbb{R}^k$, after extending with zeros. The seminorms (actually sup-norm) $p$ and $q_k$ defined previously, induces a sequential locally convex and complete topology on $\mathfrak{d}$. In order to better distinguish an element in $\mathbb{R}^\infty$ (a real-valued sequence) from a sequence of elements in $\mathbb{R}^\infty$, let us use the notation $\{u(n)\}$ for a sequence of elements in $\mathbb{R}^\infty$ and $u_i^{(n)}$ for the $i$-component of the element $\{u^{(n)}\}$ in $\mathbb{R}^\infty$.

- **On $c$:** (1) A sequence $\{u^{(n)}\}$ converges to zero if and only if $u_i^{(n)} \to u_i$ for every $i$. (2) A set $B$ is bounded if and only if for every $k$ there is a constant $C = C_k$ such that $p_k^b(u)(u) \leq C$ for every $u$ in $B$, i.e., $|u_i| \leq C$, for every integer $i \leq k$ and $u$ in $B$.

- **On $\mathfrak{d}$:** (1) A sequence $\{u^{(n)}\}$ converges to zero if and only if (a) there exists an integer $k$ such that $u_i^{(n)} = 0$ for every $i > k$ and any $n$, and (b) $u_i^{(n)} \to u_i$ for every $i$. (2) A set $B$ is bounded if and only if there an integer $k = k$ and a constant $C$ such that $|u_i| \leq C$ for any $i \leq k$ and $|u_i| = 0$ for any $i > k$, for every $u$ in $B$.

- **On $s$:** (1) A sequence $\{u^{(n)}\}$ converges to zero if and only if $p_k(u^{(n)}) \to 0$, for every integer $k$. (2) A set $B$ is bounded if and only if for every $k$ there is a constant $C = C_k$ such that $p_k(u) \leq C$ for every $u$ in $B$, i.e., $i^k|u_i| \leq C$, for every integer $i$ and $u$ in $B$.

It is clear from the definition that if $u$ belongs to $\mathfrak{d}$ and $v$ belongs to $c$ then the product $uv = \{u_i v_i : u \geq 1\}$ belongs to $\mathfrak{d}$. Hence, if an increasing sequence $u_k$ in $\mathfrak{d}$ satisfies $u_k \to 1$ in $c$ then $u_kv \to v$ in $c$, i.e., $\mathfrak{d}$ is dense in $c$. Certainly, these remarks also apply with $s$ replacing either $c$ or $\mathfrak{d}$, i.e., $\mathfrak{d} \subset s \subset c$, with injective and dense inclusions. All this is very similar to the spaces of distributions $\mathcal{D}(\mathbb{R}) \subset \mathcal{S}(\mathbb{R}) \subset \mathcal{E}(\mathbb{R})$, but a key difference is the infinite-dimensional subspace $\mathcal{D}_K(\mathbb{R})$ and its equivalent, the finite-dimensional space $\mathfrak{d}_k$, which is $\mathbb{R}^k$ ‘augmented’ by zeros. The spaces $\mathcal{S}(\mathbb{R})$ and $s$ are actually the ‘same’, via an isometric transformation between the Hilbert spaces $H_n$ and $h_n$. Analogously, the pair of spaces $C^0(\mathbb{R})$ (as a subspace of $\mathcal{E}$) and $c$, and the pair $C^0_0(\mathbb{R})$ (as a subspace of $\mathfrak{d}$) and $\mathfrak{d}$ have many similarities and some differences.
Regarding the dual spaces $\mathfrak{d}'$, $\mathfrak{s}'$ and $\mathfrak{e}$, it is clear that $\ell^2$ is a Hilbert space, and therefore, these dual spaces satisfy the relations

$$\mathfrak{d} \subset \mathfrak{s} \subset \ell^2 = (\ell^2)' \subset \mathfrak{s}' \subset \mathfrak{d}'$$

and $\mathfrak{e}' \subset \mathfrak{d}'$, and certainly, these three spaces are reflexive, i.e., the bidual reproduce the initial space, namely, $\mathfrak{d}'' = \mathfrak{d}$, $\mathfrak{s}'' = \mathfrak{s}$ and $\mathfrak{e}'' = \mathfrak{e}$. Actually, consider the canonical basis $e_i = \{0, \ldots, 0, 1, 0, \ldots\}$ with the 1 is the $i$-element of the sequence $e_i$ in $\mathfrak{e}$. The spanned subspace of $\{u_i : i \geq 1\}$ is the whole space $\mathfrak{d}$, which is dense in $\mathfrak{e}$. Thus, the algebraic dual space of $\mathfrak{d}$ is clearly identified with $\mathfrak{e}$, and with the inductive topology, a natural isomorphism, $\mathfrak{d}' = \mathfrak{e}$, is obtained. This is very different from the distributions space $\mathcal{D}$ and $\mathcal{E}$, as one may expect.
Chapter 4

Introduction to Sobolev Spaces

A classic reference is the book by Adams [2], only the tip of the ice is seriously discussed on this section. Usually, this is introduced as a means to solve partial differential equations, e.g., the books by Evans [42] and Renardy and Rogers [104].

If Ω is an open subset of $\mathbb{R}^d$, $1 \leq p \leq \infty$ and $m = 0, 1, \ldots$, then the Sobolev space $W^{m,p}(\Omega)$ is defined as all functions $f$ in $L^p(\Omega)$ with weak derivative (see Definition 3.25) $\partial^\alpha f$ in $L^p(\Omega)$, for any multi-indexes of order $|\alpha| \leq m$. In particular $W^{0,p}(\Omega) = L^p(\Omega)$, and $W^{1,\infty}(\Omega)$ is the space of Lipschitz continuous functions, i.e., functions $f$ such that for some constant $M$ we have $|f(x) - f(x')| \leq M|x - x'|$, for every $x, x'$ in $\Omega$ (we need to recall Rademacher’s Theorem, namely, any Lipschitz continuous function is almost everywhere differentiable). The vector space $W^{p,m}(\Omega)$ becomes a Banach space with the norm

$$
\|f\|_{m,p} = \left( \sum_{|\alpha| \leq m} \int_\Omega |\partial^\alpha f(x)|^p \, dx \right)^{1/p}.
$$

Moreover, for $p = 2$ and the notation $W^{m,2}(\Omega) = H^m(\Omega)$, the norm $\| \cdot \|_{m,2}$ derives from an inner product and $H^m(\Omega)$ is a Hilbert space. As an exercise, the reader may verify the completeness of $W^{m,p}(\Omega)$. It is clear that if $\Omega$ is bounded then $C^m(\overline{\Omega}) \subset C^m_b(\Omega) \subset W^{m,p}(\Omega)$, for every $1 \leq p \leq \infty$. For $p = \infty$ we have the strict inclusion of $C^m_b(\Omega)$ into $W^{m,\infty}(\Omega)$, as a closed subspace. Thus, we define $W^{m,p}_0(\Omega)$, for $1 \leq p < \infty$, as the closure of $C^\infty(\Omega)$ (actually, $C^0_b(\Omega)$ suffices) in $W^{m,p}(\Omega)$. For now, this leave out the case $p = \infty$, i.e., $W^{m,\infty}_0(\Omega)$. Local Sobolev space can be defined too, i.e., $W^{m,p}_{loc}(\Omega)$ are functions in $W^{m,p}(\omega)$ for any open set $\omega$ with compact closure inside the open set $\Omega$.

Therefore, for any nonnegative integer $m$ and $1 \leq p \leq p$, we have a working definition of the Sobolev spaces $W^{m,p}(\Omega)$ and $W^{m,p}_0(\Omega)$. The dual space $(W^{m,p}_0(\Omega))'$ is denoted by $W^{-m,p'}(\Omega)$, with $1/p + 1/p' = 1$. These Sobolev
spaces of negative order are also Banach (Hilbert if \( p = 2 \)) spaces with the dual norm, and the continuous and dense inclusion \( D(\Omega) \subset W_{-m,p}(\Omega) \) becomes \( W_{-m,p}(\Omega) \subset D'(\Omega) \), by duality. On the contrary, because a priori, the test functions \( D(\Omega) \) are not dense in \( W_{m,p}(\Omega) \), the dual spaces \( (W_{m,p}(\Omega))^' \) are not spaces of distributions in \( \Omega \), i.e., their elements could have some distribution on the boundary \( \partial \Omega \) too. Indeed, the restriction to \( \Omega \) of functions defined on \( \mathbb{R}^d \) establishes a continuous inclusion \( W_{m,p}(\mathbb{R}^d)|_{\Omega} \subset W_{m,p}(\Omega) \), and as seen later, if \( \Omega \) is sufficiently smooth then the equality holds. In this case, by duality, \( (W_{m,p}(\Omega))^' = (W_{m,p}(\mathbb{R}^d)|_{\Omega})' \subset D'(\mathbb{R}^d)|_{\Omega} \), with a proper meaning, i.e., any element \( f' \) in \( (W_{m,p}(\Omega))^' \) can be regarded as the distribution \( \varphi \mapsto \langle f', \varphi \rangle_{\Omega} \) for every \( \varphi \) in \( D'(\mathbb{R}^d) \), where as usual. Note that \( \varphi|_{\Omega} \) means the restriction to \( \Omega \), which does not necessarily belong to \( D(\Omega) \). Actually, it can be shown that any element \( f' \) in the dual space \( (W_{m,p}(\Omega))^' \) admits a representation of the form

\[
\langle f', u \rangle = \sum_{|\alpha| \leq m} \int_{\Omega} v_\alpha(x) \partial^\alpha u(x) dx, \quad \forall u \in W_{m,p}(\Omega),
\]

where \( v_\alpha \) belongs to \( L^{p'}(\Omega) \), \( 1/p + 1/p' = 1 \), e.g., see Ziemer [140, Section 4.3, pp. 185–189]. It is clear that the above expression of \( f' \) can be used as a distribution in \( \Omega \) or a distribution in \( \mathbb{R}^d \) with \( v_\alpha \mathbb{1}_\Omega \) in lieu of \( v_\alpha \). If test functions \( D'(\mathbb{R}^d)|_{\Omega} \) are dense in \( W_{m,p}(\Omega) \) (as shown later if the domain \( \Omega \) is smooth), these two distributions are the same if and only if \( f' \) belongs to \( W_0^{m,p}(\Omega) \).

## 4.1 Density and Extension

Now that Sobolev spaces \( W_{m,p}(\Omega) \) and \( W_0^{m,p}(\Omega) \) have been defined for \( 1 \leq p \leq \infty \) and \( m \) a nonnegative integer, we begin with

**Theorem 4.1** (Meyers-Serrin). The subspace \( C^\infty(\Omega) \cap W_{m,p}(\Omega) \) is dense in \( W_{m,p}(\Omega) \), for any \( 1 \leq p < \infty \).

**Proof.** The arguments make use of the kernel convolutions and the partition of the unity. We have to show that for every \( \delta > 0 \) and any \( f \) in \( W_{m,p}(\Omega) \) there exists \( \varphi \) in \( C^\infty(\Omega) \) such that \( \| f - \varphi \|_{m,p} < \delta \).

To this purpose, define

\[
\Omega_n = \{ x \in \Omega : d(x, \partial \Omega) > 1/n, \quad |x| < n \}, \quad \Omega_1 = \Omega_2, \quad \text{and}
\]

\[
\Omega_n = \{ x \in \Omega : 1/(n + 1) < d(x, \partial \Omega) < 1/(n - 1), \quad n - 1 < |x| < n + 1 \},
\]

where \( d(x, \partial \Omega) \) denotes the distance from \( x \) to the boundary \( \partial \Omega \). Note these open sets satisfy \( \Omega_n \subset \Omega_{n+1} = \bigcup_{i=1}^n \Omega_i \) and \( \Omega = \bigcup_n \Omega_n \), so that \( \Omega = \bigcup_n \Omega_n \). Thus, by the partition of the unity Theorem B.88, there exists a sequence \( \{ \chi_n : n \geq 1 \} \) such that (a) \( \chi_n \) belongs to \( C^\infty(\Omega_n) \), (b) \( 0 \leq \chi_n(x) \leq 1 \) and \( \sum_n \chi_n(x) = 1 \), for every \( x \) in \( \Omega \), where the series is locally finite, namely, for any compact set \( K \) of \( \Omega \) the set of indices \( i \) such that the support of \( \chi_i \) intercept \( K \), \( \text{supp}(\chi_i) \cap K \neq \emptyset \), is finite.
If $k$ is a smooth kernel with support in the unit ball, $k_\varepsilon(x) = \varepsilon^{-d}k(x/\varepsilon)$, and $0 < \varepsilon(n+1)(n+2) < 1$ then the mollification $f \ast k_\varepsilon$ has support in $K_n \subset \Omega_{n+2} \cap (\Omega \setminus \Omega_{n-2})$. Hence, we may choose positive real numbers $\varepsilon_n$ such that $0 < \varepsilon_n(n+1)(n+2) < 1$ and

$$\int_{K_n} \left| \left((\chi_n f) \ast k_{\varepsilon_n}\right)(x) - (\chi_n f)(x) \right|^p \, dx \leq (2^{-n} \delta)^p,$$

for every $n \geq 1$.

Define $\varphi = \sum_{n=1}^{\infty} (\chi_n f) \ast k_{\varepsilon_n}$, which belongs to $C^\infty(\Omega)$ since the series is locally finite. Actually for every $x$ in $\Omega_n$ we have

$$\varphi(x) = \sum_{i=1}^{n+2} (\chi_i f) \ast k_{\varepsilon_i})(x) \quad \text{and} \quad f(x) = \sum_{i=1}^{n+2} (\chi_i f)(x),$$

with $(\chi_{n+2} f)(x) = 0$. Therefore, denoting by $\| \cdot \|_{m,p;\Omega}$ the norm within the open set $\mathcal{O} \subset \Omega$, we obtain

$$\|f - \varphi\|_{m,p;\Omega} \leq \sum_{i=1}^{n+2} \| (\chi_i f) \ast k_{\varepsilon_i} - \chi_i f\|_{m,p;\Omega} \leq \delta,$$

and we conclude as $n$ goes to $\infty$. \hfill $\square$

Continuity up to the boundary requires some regularity for the boundary, e.g., we look for sufficient conditions to deduce that the subspace $C^m(\overline{\Omega}) \cap W^{m,p}(\Omega)$ is dense in $W^{m,p}(\Omega)$, for $1 \leq p < \infty$.

### 4.1.1 Regularity on the Domain

Essentially, we need to force the domain $\Omega$ to lie locally in one side of its boundary, e.g., the unit disk from which a radius is removed does not satisfy the following property.

**Definition 4.2.** An open set $\Omega$ in $\mathbb{R}^d$ satisfies the *segment property* (or condition) if there exists a locally finite open covering $\{U_i\}$ of its boundary $\partial\Omega$ and corresponding vectors $\xi^i$ in $\mathbb{R}^d \setminus \{0\}$ such that $x + t\xi^i$ belongs to $\Omega$ for all $x$ in $\overline{\Omega} \cap U_i$ and $t$ in $(0, 1)$.

Note that a locally finite of cover $\{U_i : i \geq 1\}$ of $\partial\Omega$ means that $\partial\Omega \subset \bigcup_i U_i$ and that any compact subset of $\partial\Omega$ can intersect at most finitely many element of $\{U_i : i \geq 1\}$. Now, we state the following result, e.g., see Nečas [93].

**Theorem 4.3.** For any open subset $\Omega$ of $\mathbb{R}^d$ satisfying the segment property the subspace $C^\infty_0(\overline{\mathbb{R}^d}) \cap W^{m,p}(\Omega)$ is dense in $W^{m,p}(\Omega)$, for any $1 \leq p < \infty$. In particular we have $W^{m,p}(\mathbb{R}^d) = W^{m,p}_0(\mathbb{R}^d)$, for every $1 \leq p < \infty$ and $d \geq 1$. \hfill $\square$

Recall that a function $f$ satisfying

$$|f(x) - f(y)| \leq M|x - y|^{\alpha}, \quad \forall x, y \in \mathcal{O}$$
is called \(\alpha\)-Hölder continuous on \(\mathcal{O} \subset \mathbb{R}^d\) for \(0 < \alpha < 1\) and Lipschitz continuous for \(\alpha = 1\). Thus, a function of class \(C^m\) means continuous functions with continuous partial derivatives up to the order \(m\), while a function of class \(C^{m,\alpha}\) means a function of class \(C^m\) having \(\alpha\)-Hölder continuous partial derivatives of order \(m\). If \(\alpha = 1\), a function of class \(C^{m,1}\) means of class \(C^m\) having Lipschitz continuous partial derivatives of order \(m\).

If \(\Omega\) is an open subset of \(\mathbb{R}^d\) and \(f\) a function of class \(C^{m,\alpha}\) on \(\overline{\Omega}\) then we define the quantity

\[
\|f\|_{m,\alpha} = \sum_{|\beta| \leq m} \|\partial^\beta f\|_\alpha, \quad \|g\|_\alpha = \|g\|_\infty + |g|_\alpha,
\]

where \(\beta\) is an multi-index, \(\| \cdot \|_\infty\) is the sup-norm on \(\overline{\Omega}\) and \(| \cdot |_\alpha\) is the Hölder or Lipschitz seminorm given by

\[
|g|_\alpha = \sup \{|g(x) - g(y)|/|x - y| : x, y \in \overline{\Omega}, x \neq y\}.
\]

Thus, \(C^{m,\alpha}(\overline{\Omega})\) with an integer \(m\) and \(0 < \alpha \leq 1\) is the Banach space of functions \(f\) satisfying \(\|f\|_{m,\alpha} < \infty\). Recall also the Banach space \(C^{m}(\overline{\Omega})\) of bounded functions of class \(C^m\) in \(\overline{\Omega}\) having bounded derivatives, with the sup-type norm \(\| \cdot \|_{m,\infty}\). Note the ambiguity of the notation \(\| \cdot \|_{m,1}\) which may represent the norm on the Lipschitz space \(C^{m,1}\) or on the Sobolev space \(W^{m,1}\), but when necessary we will use \(\| \cdot \|_B\) for a Banach space \(B\).

Extending the definition of a function from \(\Omega\) into \(\mathbb{R}^d\) requires some regularity on the domain, for instance

**Definition 4.4 (smooth domain).** An open set \(\Omega\) in \(\mathbb{R}^d\) with boundary \(\partial \Omega\) is called a **domain of class** \(C^{m,\alpha}\), for a nonnegative integer \(m\) and \(0 \leq \alpha \leq 1\) if at each point \(\xi\) on the boundary \(\partial \Omega\) there exist neighborhood \(Q\) of \(\xi\), (local) orthogonal coordinates \(y = (y_1, \ldots, y_d)\) with origin at \(\xi\) such that \(Q = \{y : -r < y_i < r\}\) and there is a function \(\phi\) of class \(C^{m,\alpha}\) defined on \(Q' = \{y' = (y_1, \ldots, y_{d-1}) : -r < y_i < r\}\) such that \(|\phi(y')| \leq r/2\) on \(Q'\) and \(\Omega \cap Q = \{y = (y', y_d) : y_d < \phi(y')\}\) and \(\partial \Omega \cap Q = \{y = (y', y_d) \in Q : y_d = \phi(y')\}\).

Usually an open set is called a domain if it is connected and the function \(\phi\) referred to as a local hypograph of class \(C^{m,\alpha}\). Remark that if \(m \geq 1\) then the outward unit normal vector to \(\partial \Omega\) at the point \(x = (y', \phi(y'))\) is (locally) defined by

\[
\mathbf{n}(y', \phi(y')) = \frac{(-\partial_1 \phi(y'), \ldots, -\partial_{d-1} \phi(y'), 1)}{[1 + (\partial_1 \phi(y'))^2 + \cdots + (\partial_{d-1} \phi(y'))^2]^{1/2}}, \tag{4.2}
\]

which is a function of class \(C^{m-1,\alpha}\) in \(Q'\). Certainly, this yields \(d-1\) independent tangential unit vectors \(\mathbf{t}_i\), for \(i = 1, \ldots, d-1\), e.g.,

\[
\mathbf{t}_1(y', \phi(y')) = \frac{(1, 0, \ldots, 0, \partial_1 \phi(y'))}{[1 + (\partial_1 \phi(y'))^2]^{1/2}}.
\]
which are orthogonal to \( \mathbf{n} \) as expected. However, for a Lipschitz domain (i.e., of class \( C^{0,1} \)), the outward unit normal \( \mathbf{n} \) and the tangential unit vectors \( \mathbf{t}_i \), for \( i = 1, \ldots, d-1 \), are only defined almost everywhere with respect to the surface area

\[
d\sigma = \left[ 1 + (\partial_1 \phi(y'))^2 + \cdots + (\partial_{d-1} \phi(y'))^2 \right]^{1/2} dy',
\]

(or measure \( dy' \)) on the boundary \( \partial \Omega \). Also, note that it is not hard to check that a convex open set is a domain of class \( C^{0,1} \), i.e., a Lipschitz domain.

\begin{itemize}
\item \textbf{Remark 4.5.} For a domain of class \( C^{m,\alpha} \) with \( m \geq 1 \) and \( 0 \leq \alpha \leq 1 \) Definition 4.4 becomes: (1) at each point \( \xi \) on the boundary \( \partial \Omega \) there exists a tangent plane, i.e., a (outward) normal unit vector \( n(\xi) \), and therefore local coordinates \( y = (y_1, \ldots, y_d) \) with origin at \( \xi \) and \( y_d \) axis in the direction of \( n(\xi) \) are defined, (2) for each point \( \xi \) on the boundary \( \partial \Omega \) there exists a sphere with center \( \xi \) and radius \( r > 0 \), and a function \( f(y'), y' = (y_1, \ldots, y_{d-1}) \) in a \((d-1)\)-dimensional ball \( B = \{ y' : |y'| \leq r/2 \} \) such that the surface \( \partial \Omega \) is given in a local system of (orthogonal) coordinates by the equation \( y_d = f(y_1, \ldots, y_{d-1}) \), where the function \( f \) is of class \( C^{m,\alpha} \). Sometimes, an open set \( \Omega \) of \( \mathbb{R}^d \) is called a \textit{Lipschitz domain} if for every point \( \xi \) on the boundary \( \partial \Omega \) there exists an open ball \( B \) centered at \( \xi \) and a bijection \( h : B \to Q, Q \subset \mathbb{R}^d \) such that (a) \( h \) and its inverse \( h^{-1} \) are both Lipschitz continuous functions, (b) \( h(\partial \Omega \cap B) = \{ y \in Q : y_d = 0 \} \) and (c) \( h(\Omega \cap B) = \{ y \in Q : y_d > 0 \} \). In this case, the domain is locally given by an inequality of the form \( y_d < \phi(y_1, \ldots, y_{d-1}) \) and the boundary by the equality \( y_d = \phi(y_1, \ldots, y_{d-1}) \), where \( y' \mapsto \phi(y') = -h_d(h^{-1}(y',0)) \) is a Lipschitz function, i.e., it satisfies Definition 4.4 with \( m = 0 \) and \( \alpha = 1 \). Also observe that it can be proved that if \( \Omega \) is a domain of class \( C^{0,\alpha} \) and \( \Phi \) is a bi-Lipschitz homeomorphism on a neighborhood of the closure \( \overline{\Omega} \) then \( \Phi(\Omega) \) is also a domain of class \( C^{0,\alpha} \), e.g., see Grisvard [60, Section 1.2.1, pp.5–14].
\end{itemize}

It is also clear that a domain of class \( C^0 \) satisfies the segment property and therefore, the \( C_0^\infty(\mathbb{R}^d) \cap W^{m,p}(\Omega) \) is dense in \( W^{m,p}(\Omega) \), for any \( 1 \leq p < \infty \) and any domain of class \( C^0 \).

In general, for a domain \( \Omega \) of class \( C^{m,\alpha} \), we can find a locally finite open cover \( \{ \mathcal{O}_i \} \) of \( \Omega \) and construct a regular partition of unity subordinate to this covering (i.e., \( \sum_i \chi_i(x) = 1 \), \( \chi_i \) is \( C^\infty \) with compact support in \( \mathcal{O}_i \)) with the following properties:

\begin{itemize}
\item (a) For every \( i \), we have either \( d(\mathcal{O}_i, \partial \Omega) > 0 \) or \( \mathcal{O}_i \cap \partial \Omega \neq \emptyset \);
\item (b) There exists one-to-one transformations \( y = Y_i(x) \) of class \( C^{m,\alpha} \) mapping \( \mathcal{O}_i \) into either the open ball \( B = \{ y \in \mathbb{R}^d : |y| < 1 \} \) or the open half-ball \( B_+ = \{ y \in \mathbb{R}^d_+ : |y| < 1 \} \), where the image of \( \mathcal{O}_i \cap \partial \Omega \) is a flat part of \( \partial B_+ \).
\end{itemize}

If \( \Omega \) is bounded then the open cover \( \{ \mathcal{O}_i \} \) is finite and if only the boundary \( \partial \Omega \) is bounded then \( \partial \mathcal{O}_i \cap \partial \Omega \neq \emptyset \) only for finite many \( i \).

\textbf{Theorem 4.6.} Let \( \Omega \) be a domain in \( \mathbb{R}^d \) of class \( C^m \) or \( C^{m,\alpha} \) as above. Then there exists a linear and bounded extension operator \( E \) preserving the class, i.e., from either \( C^m_b(\overline{\Omega}) \) or \( C^{m,\alpha}_b(\overline{\Omega}) \) or \( W^{m,p}(\Omega) \) into either \( C^m_b(\mathbb{R}^d) \) or \( C^{m,\alpha}_b(\mathbb{R}^d) \) or \( W^{m,p}(\mathbb{R}^d) \), \( 1 \leq p < \infty \), satisfying \( Ef(x) = f(x) \), for almost every \( x \) in \( \Omega \). \( \square \)
Proof. We consider only the case when $\Omega$ is bounded, and we give details for the semi-space $\mathbb{R}^d_+ = \{ x = (\tilde{x}, x_d) \in \mathbb{R}^d : x_d > 0 \}$. An argument call by local coordinates goes as follows: we write $f = \sum_i f_i - f_i \chi_i$, for a functions $f$ defined on $\Omega$ and the regular partition of the unity subordinated to the open cover $\{ O_i \}$ corresponding to smooth domain $\Omega$. Thus $f_i$ inherits the same smoothness as $f$ and has support in $O_i$. If $O_i$ intersect the boundary $\partial \Omega$ then the change of variables $y = Y_i(x)$ maps $O_i$ into the open half-ball $B_+ = \{ y \in \mathbb{R}^d_+ : |y| < 1 \}$, where the image of $O_i \cap \partial \Omega$ is a flat part of $\partial B_+$. Hence, the function $g_i(y) = f_i(x)$ has support in $B_+$ and can be extended to the whole ball $B$, preserving its class.

Now, for $\mathbb{R}^d_+$, the construction is based of successive reflections in smooth boundary. We define the extensions

$$Ef(x) = \begin{cases} f(x) & \text{if } x_d \geq 0, \\ \sum_{k=1}^{m+1} \lambda_k f(x_1, \ldots, x_{d-1}, -kx_n) & \text{if } x_d < 0, \end{cases}$$

and for a multi-index $\alpha$,

$$E_\alpha f(x) = \begin{cases} f(x) & \text{if } x_d \geq 0, \\ \sum_{k=1}^{m+1} (-k)^{\alpha_d} \lambda_k f(x_1, \ldots, x_{d-1}, -kx_n) & \text{if } x_d < 0, \end{cases}$$

where the coefficients $\lambda_1, \ldots, \lambda_{m+1}$ are the unique solution of the $(m+1) \times (m+1)$ system of linear equations

$$\sum_{k=1}^{m+1} (-k)^i \lambda_k = 1, \quad \text{for } i = 0, 1, \ldots, m.$$ 

We can check that if $f$ belongs to $C^m(\mathbb{R}^d_+)$ then $Ef$ belongs to $C^m(\mathbb{R}^d)$, $E_\alpha f$ belongs to $C^{m-\alpha_d}(\mathbb{R}^d)$,

$$\partial^\alpha Ef = E_\alpha \partial^\alpha f, \quad \forall |\alpha| \leq m,$$

and

$$\|Ef\|_{C^m_b(\mathbb{R}^d_+)} \leq C(m) \|f\|_{C^m_b(\mathbb{R}^d)},$$

$$\|Ef\|_{C^m_b(\mathbb{R}^d_+)} \leq C(m, \alpha) \|f\|_{C^m_b(\mathbb{R}^d)},$$

$$\|Ef\|_{W^{m,p}(\mathbb{R}^d)} \leq K(m, p) \|f\|_{W^{m,p}(\mathbb{R}^d)},$$

for some constants $C(m)$, $C(m, \alpha)$ and $K(m, p)$ and for any $f$ in $C^m(\mathbb{R}^d_+)$. □

Sometimes we make use of only a domain $\Omega$ of class $C^{m,\alpha}$ only piecewise, this means that $\Omega$ is a finite intersection of domain $\Omega_i$ of class $C^{m,\alpha}$. Moreover, if we need to work with the exterior normal $n$ of the domain, we may use a
condition of the type $\Omega = \{ x \in \mathbb{R}^d : \omega(x) < 0 \}$ with $\partial \Omega = \{ x \in \mathbb{R}^d : \omega(x) = 0 \}$ and $|\nabla \omega| \geq 1$ on $\partial \Omega$, which yields $n = \nabla \omega / |\nabla \omega|$. Furthermore, the extension Theorem 4.6 remains valid under weaker conditions, e.g., the so-called $(\varepsilon, \delta)$-domains, see the paper Jones [71].

It is rather interesting to remark the chain rule in the sense that for any $u$ in $C^m_b(\mathbb{R})$ the composition $x \mapsto u(f(x))$ preserves also the class $W^{m,p}(\Omega)$ and $\partial^\alpha u(f)$ follows Leibniz formula, e.g., $\partial_i u(f)(x) = u'(f(x)) \partial_i f(x)$. Moreover, it is a good exercise to approximate $u(x) = x^+$, e.g., by $u_\varepsilon(x) = 1_{\{x > 0\}}(\sqrt{x^2 - \varepsilon^2} - \varepsilon)$, to show (based on the chain rule and dominated convergence) that $f^+, f^-$ and $|f|$ belong to $W^{1,p}(\Omega)$, for every $f$ in $W^{1,p}(\Omega)$. Actually, a more general result holds.

### 4.1.2 Lipschitz Transformation

If $\Omega$ is an open subset of $\mathbb{R}^d$ then a bi-Lipschitz change of variables (or homeomorphism) in $\Omega$ is a map $\Phi$ from $\Omega$ into $\mathbb{R}^d$ such that

$$c|x - y| \leq |\Phi(x) - \Phi(y)| \leq C|x - y|, \quad \forall x, y \in \Omega,$$

and some constants $C \geq c > 0$. This means that $\Phi$ is a Lipschitz continuous one-to-one function and its inverse $\Phi^{-1}$ is also Lipschitz.

**Proposition 4.7.** Let $\Omega$ be an open subset of $\mathbb{R}^d$ and $\Phi$ be a bi-Lipschitz change of variables in $\Omega$. If $u$ belongs to $W^{1,p}(\Omega)$ then the function $x \mapsto v(x) = u(\Phi^{-1}(x))$ belongs to $W^{1,p}(\mathcal{O})$, with $\mathcal{O} = \Phi(\Omega)$. Moreover, there exist a negligible set $N \subset \Omega$ such that

$$\nabla v(x) = \sum_{j=1}^d u'_j(\Phi^{-1}(x)) \nabla \Phi^{-1}_j(x),$$

for almost every $x$ in $\Omega \setminus N$, where $u'_j = \partial_j u$ and $\Phi^{-1} = (\Phi_1^{-1}, \ldots, \Phi_d^{-1})$.

**Proof.** First, recall that a Lipschitz transformation $T : \Omega \to \mathbb{R}^d$ preserves negligible and Lebesgue measurable sets. Moreover, if $T$ is bi-Lipschitz and the Jacobian is defined almost everywhere by $J_T = |\det(\nabla T)|$ then

$$\int_{T(\Omega)} f(T^{-1}(x))dx = \int_{\Omega} f(y)J_T(y)dy,$$

for any nonnegative Lebesgue measurable function $f$ in $\mathbb{R}^d$.

If $u$ is a smooth function in $\Omega$ then $v$ is Lipschitz continuous in $\mathcal{O}$ and Rademacher’s Theorem (e.g., see Part I) implies that $v$ is almost everywhere differentiable and the chain rule (4.4) holds.

Since $\Phi$ is Lipschitz, the Jacobian $x \mapsto J_\Phi(x)$ of the change of variables $x \mapsto \Phi^{-1}(x)$ is bounded. Thus, there is a constant $C$ depending only on the dimension $d$ and the Lipschitz constants of $\Phi$ and $\Phi^{-1}$ such that

$$|v(x)|^p + |\nabla v(x)|^p \leq C \left( |u(\Phi^{-1}(x))|^p + |\nabla u(\Phi^{-1}(x))|^p \right) J_\Phi(\Phi^{-1}(x)),$$
almost every \( x \) in \( \mathcal{O} \). Hence, apply equality (4.5) with \( f = |\nabla u|^p \mathbb{1}_{\Omega'} \) and any open set \( \Omega' \subset \Omega \) to deduce

\[
\int_{\Omega'} (|v(x)|^p + |\nabla v(x)|^p) \, dx \leq C \int_{\Omega} (|u(x)|^p + |\nabla u(x)|^p) \, dx,
\]

for any open set \( \Omega' \subset \Omega \).

Now, take a sequence \( \{u_k\} \) of smooth functions such that \( u_k \to u \) in \( W^{1,p}(\Omega) \) and almost everywhere, see Meyers-Serrin Theorem 4.1, and use the previous estimate with \((u_k - u_n)\) to obtain

\[
\|v_k - v_n\|_{W^{1,p}(\Omega')} \leq C \|u_k - u_n\|_{W^{1,p}(\Omega)}, \quad \forall k, n,
\]

for any open set \( \Omega' \) with compact closure contained in \( \mathcal{O} \) and some suitable constant \( C \) depending only on \( p \), the dimension \( d \) and the Lipschitz constants of \( \Phi \) and \( \Phi^{-1} \).

Finally, collect all estimates to deduce that \( \{v_n\} \) is a Cauchy sequence in \( W^{1,p}(\mathcal{O}) \) and \( v_n \to v \) almost everywhere. Hence, \( v \) belongs to \( W^{1,p}(\mathcal{O}) \) and the chain rule (4.4) holds.

It is clear (but perhaps tedious) that Proposition 4.7 can be generalized to \( W^{m,p}(\Omega) \), i.e., the Sobolev space \( W^{m,p}(\Omega) \) is preserved by any homeomorphism of class \( C^{m-1,1} \) in \( \Omega \).

### 4.2 Imbedding and Compactness

First, clearly \( W^{m,p}(\mathbb{R}^d) \subset L^p(\Omega) \) and a preliminary observation is given by the following

**Proposition 4.8.** Let \( K \) be an uniformly bounded set in \( W^{1,p}(\mathbb{R}^d) \) with \( 1 \leq p \leq \infty \), and with the property that for every \( \varepsilon > 0 \) there exists \( r > 0 \) such that \( \|\mathbb{1}_{|\cdot|>r} f(\cdot)\|_p < \varepsilon \), for every \( f \) in \( K \). Then \( K \) is a totally bounded set in \( L^p(\mathbb{R}^d) \).

**Proof.** By means of Theorem 2.38 and Remark 2.40 we have to show that the translations are uniformly continuous on \( K \), i.e., if \( (\tau_{r,i} f)(x) = f(x_1, \ldots, x_i + r, \ldots, x_d) \) then for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( \|\tau_{r,i} f - f\|_p < \varepsilon \), for every \( 0 < r < \delta \), \( i = 1, \ldots, d \) and every \( f \) in \( K \).

Holder inequality applied to

\[
(\tau_{r,i} f)(x) - f(x) = \int_0^1 r \partial_i f(x_1, \ldots, x_i + rt, \ldots, x_d) \, dt
\]

yields

\[
\left| (\tau_{r,i} f)(x) - f(x) \right|^p \leq r^p \int_0^1 \left| \partial_i f(x_1, \ldots, x_i + rt, \ldots, x_d) \right|^p \, dt.
\]
Hence, Fubini theorem implies
\[ \int_{\mathbb{R}^d} |(\tau_{r,i} f(x)) - f(x)|^p \, dx \leq r^p \int_0^1 dt \int_{\mathbb{R}^d} |(\tau_{r,i}(\partial_i f))(x)|^p \, dx, \]
which proves that \( \|\tau_{r,i} f - f\|_p \leq r \|\partial_i f\|_p \). This last estimate holds also for \( p = \infty \), we conclude. \( \square \)

**Remark 4.9.** The reader may verify that the previous Proposition 4.8 remains valid for an open set \( \Omega \subset \mathbb{R}^d \) with changes accordingly to Theorem 2.38 and involving the notation \( \| \cdot \|_{p; \Omega_\delta} \), for \( \Omega_\delta = \{ x \in \Omega : d(x, \partial \Omega) > \delta, |x| < 1/\delta \} \). This shows that if \( \Omega \subset \mathbb{R}^d \) is a bounded domain then the inclusion (operator) \( W^{1,p}(\Omega) \subset L^p(\Omega) \) transforms (uniformly) bounded sets into totally bounded (or pre-compact) sets. \( \square \)

Perhaps the simplest situation is the one-dimensional case, i.e.,

**Proposition 4.10.** We have (1) \( W^{1,p}(\mathbb{R}) \subset C^{1-1/p}(\mathbb{R}) \), with \( 1 < p < \infty \), and (2) \( W^{d,1}(\mathbb{R}^d) \subset C^0_0(\mathbb{R}^d) \).

**Proof.** Indeed, by means of Hölder inequality, for a 1-dimensional variable we obtain
\[ |f(x) - f(y)| \leq \left| \int_x^y f(t) \, dt \right| \leq |x - y|^{1-1/p} \left( \int_{\mathbb{R}} |f(t)|^p \, dt \right)^{1/p}, \]
which proves (1).

Now, if \( f \) is a smooth function \( f \) with compact support then we have
\[ f(x_1, \ldots, x_i, \ldots, x_d) = \int_{-\infty}^{x_i} \partial_i f(x_1, \ldots, t_i, \ldots, x_d) \, dt = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_i} \cdots \int_{-\infty}^{x_d} \partial_1 \cdots \partial_d f(t_1, \ldots, t_i, \ldots, t_d) \, dt, \]
which yields the estimate
\[ \|f\|_\infty \leq \|\partial^d f\|_1, \quad \text{with} \quad \partial^d = \partial_1 \cdots \partial_d, \quad \forall f \in C^d_0(\mathbb{R}^d). \quad (4.6) \]

By means of Theorem 4.3, for every \( f \) in \( W^{d,1}(\mathbb{R}^d) \) there exists a sequence \( \{f_n\} \) of functions in \( C^d_0(\mathbb{R}^d) \) such that \( \|f - f_n\|_{d,1} \to 0 \). Hence, estimate (4.6) shows that \( \|f - f_n\|_\infty \to 0 \), which yields (2). \( \square \)

For \( p > d \), part (1) of Proposition 4.10 is usually referred to as Morrey’s inequality (see Proposition 4.14 later),
\[ |u(x) - u(x')| \leq C|x - x'|^{1-d/p} \|\nabla u\|_{L^p(\mathbb{R}^d)}, \quad \text{a.e.} \quad x, x' \in \mathbb{R}^d, \quad (4.7) \]
\[ |u(x)| \leq C(\|u\|_{L^p(\mathbb{R}^d)} + \|\nabla u\|_{L^p(\mathbb{R}^d)}), \quad \text{a.e.} \quad x \in \mathbb{R}^d, \]
for some constant \( C = C(d, p) \) and for any function \( u \) in \( W^{1,p}(\mathbb{R}^d) \). Certainly, the statement \( W^{1,p}(\mathbb{R}^d) \subset C^{1-1/d/p}(\mathbb{R}^d) \) actually means that any \( u \) in \( W^{1,p}(\mathbb{R}^d) \) has a representative (in its equivalent class) which belongs to \( C^{0,1-1/d/p}(\mathbb{R}^d) \).
Remark 4.11. For $u$ in $W^{1,p}(\mathbb{R}^d)$ with $p > d$, we also have $\lim_{|x| \to \infty} u(x) = 0$ and $u$ is differentiable for almost every $x$ in $\mathbb{R}^d$. Indeed, approximating $u$ by a sequence $\{u_n\}$ of smooth functions with compact support, the estimate

$$\|u - u_n\|_{L^\infty(\mathbb{R}^d)} \leq C\|u - u_n\|_{W^{1,p}(\mathbb{R}^d)}, \quad \forall n,$$

show that $|u(x)| \leq |u_n(x)| + \|u - u_n\|_{L^\infty(\mathbb{R}^d)} \leq \varepsilon$ if $n$ is sufficiently large, i.e., $|u(x)| \leq \varepsilon$ if $x$ is outside of a bounded region. Now, to check that $u$ is a.e. differentiable, define the function $v(x) = u(x) - u(x_0) - (x - x_0) \cdot \nabla u(x_0)$, for a fixed $p$-Lebesgue point of the gradient $\nabla u$, i.e., $x_0$ in $\mathbb{R}^d$ such that

$$\lim_{r \to 0} K_{x_0}(r) = 0, \quad \text{with} \quad K_{x_0}(r) := r^{-d} \int_{Q_r} |\nabla u(x) - \nabla u(x_0)|^p dx,$$

where $Q_r = x_0 + (-r/2, r/2)^d$ is the open cube of size $r$ centered at $x_0$. Remark that a first step in proving the first part of estimate (4.7) is to establish the bound

$$|u(x) - r^{-d} \int_{Q'_r} u(y)dy| \leq \frac{2dp}{p-d} \left( \int_{Q'_r} |\nabla u(y)|^p dy \right)^{1/p},$$

for almost every point $x$ in the cube $Q'_r$ of size $r$, which applied twice to $v$ (instead of $u$) yields

$$|v(x) - v(x_0)| \leq C r^{1-d/p} \left( \int_{Q_r} |\nabla v(y)|^p dy \right)^{1/p}, \quad r = 2|x - x_0|,$$

for some constant $C = C(d,p)$, or equivalently

$$\frac{|u(x) - u(x_0) - (x - x_0) \cdot \nabla u(x_0)|}{|x - x_0|} \leq C \left[ K_{x_0}(2|x - x_0|) \right]^{1/p},$$

i.e., $u$ is differentiable at $x_0$. The reader may find convenient to check Chapter 11 of the recent book Leoni [79, pp 311-347], or even the whole book for a comprehensive introductory material to Sobolev spaces. □

The limiting case $d = p$ is more delicate, we have $W^{1,d}(\mathbb{R}^d) \subset L^q(\mathbb{R}^d)$ for every $q$ in $[d, \infty)$, with the estimate

$$\|u\|_{L^q(\mathbb{R}^d)} \leq C\|u\|_{W^{1,d}(\mathbb{R}^d)}, \quad d \leq q < \infty,$$

(4.8)

for some constant $C = C(d,p)$. Note that on the unit open ball $B = \{x \in \mathbb{R}^d : |x| < 1\}$, the function $u(x) = \ln \left( \ln(1 + 1/|x|) \right)$ belongs to $W^{1,d}(B)$ but does not belong to $L^\infty(B)$. However, we have the estimate

$$\|u\|_{L^\infty(\mathbb{R}^d)} \leq \|\partial^{1,1,\ldots,1} u\|_{L^1(\mathbb{R}^d)}, \quad \forall u \in W^{d,1}(\mathbb{R}^d),$$

(4.9)

where $\partial^{1,1,\ldots,1} u$ is the $d$-derivative, once in each variable.
4.2.1 Some Typical Estimates

By no means the above results are complete, but they are sufficient to realize part of the technique necessary to establish the so-called Sobolev imbedding. Indeed, the following estimate is known as Sobolev-Gagliardo-Nirenberg’s estimate:

**Proposition 4.12.** If $1 \leq p < d$ and $d \geq 2$ then the following estimate holds,

$$
\|f\|_{L^p(\mathbb{R}^d)} \leq C\|\nabla f\|_{L^p(\mathbb{R}^d)}, \quad \text{with} \quad p^* := \frac{dp}{d-p}, \quad C = \frac{pd-p}{d-p} (4.10)
$$

for any function $f$ in $L^1_{loc}(\mathbb{R}^d)$ vanishing at infinity (i.e., $|\{x \in \mathbb{R}^d : |f(x)| > r\}| \to 0$ as $r \to \infty$) and with (gradient) $\nabla f$ in $L^p(\mathbb{R}^d)$). In particular, this implies that $W^{1, p}(\mathbb{R}^d) \subset L^q(\mathbb{R}^d)$, for every $p \leq q \leq p^*$.

**Proof.** First, if $d \geq 2$ and $f_i$ belongs to $L^{d-1}(\mathbb{R}^{d-1})$, for $i = 1, \ldots, d$, then the function $f(x) = \prod_i f_i(\hat{x}_i)$, with $\hat{x}_i = (x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_d)$, belongs to $L^1(\mathbb{R}^d)$ and

$$
\int_{\mathbb{R}^d} |f(x)|dx \leq \prod_i \left( \int_{\mathbb{R}^{d-1}} |f_i(x)|^{d-1}dx \right)^{1/(d-1)}. (4.11)
$$

Indeed, this is obvious for $d = 2$ and by induction, several application of Hölder inequality complete the argument.

To establish the estimate (4.10), without loss of generality, we may assume that $f$ is smooth with a compact support. For the case $p = 1$ use the identity

$$
|f(x)| = \left| \int_{-\infty}^{x_i} \partial_i f((x_1, \ldots, x_{i-1}, \theta, x_{i+1}, \ldots, x_d)d\theta \right|
$$

to deduce $|f(x)| \leq |f_i(\hat{x}_i)|^{d-1}$ with

$$
f_i(\hat{x}_i) = \left( \int_{\mathbb{R}} |\partial_i f((x_1, \ldots, x_{i-1}, \theta, x_{i+1}, \ldots, x_d)|d\theta \right)^{1/d-1},
$$

for $i = 1, \ldots, d$. Hence, $|f(x)|^{d/(d-1)} \leq \prod_i f_i(\hat{x}_i)$ and estimate (4.11) yields

$$
\int_{\mathbb{R}^d} |f(x)|^{d/(d-1)}dx \leq \left( \int_{\mathbb{R}^d} |\nabla f(x)|dx \right)^{(d-1)/d},
$$

i.e., (4.10) with $p^* = d/(d-1)$.

In general, for $1 < p < d$, apply estimate (4.10) with $p = 1$ to the function $g_\varepsilon = |f|^{1+\varepsilon}$, with $\varepsilon > 0$ and $\nabla g_\varepsilon = (1+\varepsilon)|f|^{\varepsilon} |\nabla f|$, to obtain

$$
\left( \int_{\mathbb{R}^d} |f(x)|^{(1+\varepsilon)d/(d-1)}dx \right)^{(d-1)/d} \leq (1+\varepsilon) \int_{\mathbb{R}^d} |f(x)|^{\varepsilon} |\nabla f(x)|dx \leq (1+\varepsilon) \left( \int_{\mathbb{R}^d} |\nabla f(x)|^p dx \right)^{1/p},
$$

after using Hölder inequality with $1/p + 1/p' = 1$. Finally, if $\varepsilon = (p-1)/(d-1)$ then $(1 + \varepsilon)d/(d-1) = \varepsilon p' = dp/(d-p)$, which proves estimate (4.10).

Finally, the last assertion follows from the general inclusion $L^p \cap L^{p^*} \subset L^q$, for every $p \leq q \leq p^*$.

It may be useful to mention Poincaré’s inequality, i.e.,

**Proposition 4.13.** If $B_r$ is a ball with radius $r > 0$ in $\mathbb{R}^d$ then for each $1 \leq p < d$ there exits a constant $C$ depending only on $p$ and $d$ such that

$$\left| \frac{1}{|B_r|} \int_{B_r} f(y) - \frac{1}{|B_r|} \int_{B_r} f(x) dx \right|^{p^*} \leq C r \left| \frac{1}{|B_r|} \int_{B_r} |\nabla f(x)|^p dx \right|^{1/p},$$

with $p^* = \frac{dp}{d-p}$, and for any open ball $B_r \subset \mathbb{R}^d$ and any function $f$ in $W^{1,p}(B_r)$.

**Proof.** Let us first show that for each $1 \leq p < \infty$ there exists a constant $C$ depending only on $p$ and the dimension $d$ such that

$$\int_{B_r} |f(y) - f(z)|^p dy \leq C r^{d+p-1} \int_{B_r} |\nabla f(y)|^p |y - z|^{1-d} dy,$$

for any open ball $B_r$ in $\mathbb{R}^d$ of radius $r$, any continuously differentiable function $f$ in $B_r$ and any point $z$ in $B_r$. Indeed, if $B(z,s)$ is the ball centered at $z$ with radius $s > 0$ then, by means of polar coordinates centered at $z$ in $B_r$, we have

$$\int_{B_r} g(x) |x - z|^p dx = \int_0^{2r} s^p ds \int_{B_r \cap \partial B(z,s)} g(x') \ell_{d-1} (dx'),$$

with $\partial B(z,s) = \{x' \in \mathbb{R}^d : |x' - z| = s\}$, as well as

$$\int_{B_r \cap \partial B(z,s)} \ell_{d-1} (dy') \int_0^1 g(z + t(y' - z)) dt \leq \int_0^1 t^{1-d} dt \int_{B_r \cap \partial B(z,ts)} g(x') \ell_{d-1} (dx') = s^{d-1} \int_0^1 t^{1-d} dt \int_{B_r \cap \partial B(z,ts)} g(x') |x' - z|^{1-d} \ell_{d-1} (dx') = s^{d-2} \int_{B_r \cap B(z,s)} g(x) |x - z|^{1-d} dx,$$
for any integrable function $g$. This implies

\[
\int_0^{2r} s^p ds \int_{B_r \cap \partial B(z,s)} \ell_{d-1}(dy') \int_0^1 g(z + t(y' - z)) \, dt \leq \\
\leq \int_0^{2r} s^{d+p-2} ds \int_{B_r \cap B(z,s)} g(x) |x - z|^{1-d} \, dx \leq \\
\leq \left( \frac{2^{d+p-1}}{d + p - 1} \right) r^{d+p-1} \int_{B_r \cap B(z,s)} g(y) |y - z|^{1-d} \, dx.
\]

Hence taking $g(x) = |\nabla f(x)|^p$ and using the inequality

\[
|f(y) - f(z)|^p \leq |y - z|^p \int_0^1 |\nabla f(z + t(y - z))|^p \, dt, \quad \forall y, z,
\]

we deduce the estimate (4.12) with $C = 2^{d+p-1}/(d + p - 1)$.

To facilitate our writing, denote by $\langle f \rangle_{B_r}$ the average of $f$ with respect to the Lebesgue measure on an open ball $B_r$ of radius $r > 0$ in $\mathbb{R}^d$, i.e.,

\[
\langle f \rangle_{B_r} = \frac{1}{|B_r|} \int_{B_r} f(y) dy, \quad |B_r| = \frac{\pi^{d/2} r^d}{\Gamma(d/2 + 1)},
\]

so that Poincaré inequality becomes

\[
\left( \langle |f - \langle f \rangle_{B_r}|^p \rangle \right)^{1/p^*} \leq C r \left( \langle |\nabla f|^p \rangle_{B_r} \right)^{1/p}, \quad (4.13)
\]

for any open ball $B_r \subset \mathbb{R}^d$ and any function $f$ in $W^{1,p}(B_r)$. Thus, use polar coordinates as in proving estimate (4.12) to compute and to obtain the estimate

\[
\langle |f - \langle f \rangle_{B_r}|^p \rangle \leq C r^p \langle |\nabla f|^p \rangle_{B_r},
\]

for another suitable constant $C = C(d,p)$. Moreover, it suffices to assume that the function $f$ belongs to $W^{1,p}(B_r)$.

Now, use on Proposition 4.12 and the continuity of the extension to the whole space $\mathbb{R}^d$ to obtain

\[
\left( \int_{B_1} |g(x)|^{p^*} \right)^{1/p^*} dx \leq C \left( \int_{B_1} (|g(x)|^p + |\nabla g(x)|^p) dx \right)^{1/p}
\]

for a ball $B_1$ with radius 1 and some constant $C$ depending only on $p$ and the dimension $d$. Next, take $g(x) = g(ry)/r$ to deduce

\[
\langle |g|^p \rangle_{B_r} \right)^{1/p^*} \leq C \left( \langle |g(x)|^p \rangle_{B_r} + r^p \langle |\nabla g(x)|^p \rangle_{B_r} \right)^{1/p}.
\]

Finally, Poincaré’s inequality (4.13) follows by collecting all pieces and setting $g = f - \langle f \rangle_{B_r}$ in the above inequalities.

Another key technique is developed in proving Morrey’s inequality, i.e.,
Proposition 4.14. For each finite $p > d$ there exists a constant $C$ depending only on $p$ and the dimension $d$ such that

\[ |f(y) - f(z)| \leq Cr \left| \frac{1}{|B_r|} \int_{B_r} |\nabla f(x)|^p dx \right|^{1/p}, \tag{4.14} \]

for any open ball $B_r$ of radius $r > 0$ in $\mathbb{R}^d$, any function $f$ in $W^{1,p}(B_r)$, and almost every $y$ and $z$ in $B_r$. In particular, if $f$ belongs to $W^{1,p}(\mathbb{R}^d)$ then the limit

\[ \lim_{r \to 0} \frac{\pi^{d/2} r^d}{\Gamma(d/2 + 1)} \int_{\{y \in \mathbb{R}^d : |y-x|<r\}} f(y) = f^*(x) \]

exists for every $x$ on $\mathbb{R}^d$ and $f^*$ is Hölder continuous with exponent $1 - d/p$.

Proof. First note that with the same notation as in the proof of Proposition 4.13, Morrey’s inequality becomes

\[ |f(y) - f(z)| \leq Cr \left( \left\langle |\nabla f|^p \right\rangle_{B_r} \right)^{1/p}, \]

and the limit can be written as

\[ \lim_{r \to 0} \langle f \rangle_{B(x,r)} = f^*(x), \]

where $B(x,r)$ is the open ball centered at $x$ with radius $r$.

It is clear that $f$ may be assumed continuously differentiable to obtain the inequality (4.14), without any loss of generality. Estimate (4.12) with $p = 1$ yields

\[ |f(y) - f(z)| \leq r^{-d} \int_{B(x,r)} (|f(y) - f(\xi)| + |f(\xi) - f(z)|) d\xi \leq C \int_{B(x,r)} |\nabla f(\xi)| (|y - \xi|^{1-d} + |\xi - z|^{1-d}) d\xi, \]

with $C = 2^{d+p-1}/(d + p - 1)$. Hence, use Hölder inequality and the estimate

\[ \int_{B(x,r)} (|y - \xi|^{1-d} + |\xi - z|^{1-d})^{p/(p-1)} d\xi \leq Cr^{d-(d-1)p/(p-1)}, \]

for another constant $C = C(p, d)$, to get

\[ |f(y) - f(z)| \leq Cr^{[d-(d-1)p/(p-1)]/(p-1)} \left( \int_{B(x,r)} |\nabla f(\xi)|^p d\xi \right)^{1/p} \leq C r^{1-d/p} \left( \int_{B(x,r)} |\nabla f(\xi)|^p d\xi \right)^{1/p}, \]

i.e., the Morrey’s inequality holds true with some constant $C$ similar to (4.12).
Now, if \( f \) belongs to \( W^{1,p}(\mathbb{R}^d) \) then apply Morrey’s inequality for almost every \( y \) and \( z \) with \( z = x \) and \( r = |x - y| \) to obtain

\[
|f(y) - f(x)| \leq C|x - y|^{1 - d/p} \left( \int_{B(x,r)} |\nabla f(\xi)|^p d\xi \right)^{1/p} \leq C\|\nabla f\|_{L^p(\mathbb{R}^d)} |x - y|^{1 - d/p},
\]

i.e., \( f \) is almost everywhere equal to a Hölder continuous function and therefore the pointwise limits defining \( f^* \) exists and is equal to \( f \) almost everywhere. □

Remark the case \( p = \infty \) is understood by the assertion: \( f \) belongs to \( W^{1,\infty}_{\text{loc}}(\mathbb{R}^d) \) if and only if \( f \) is locally Lipschitz. The interested reader may take a look at Evans and Gariepy [43, Chapter 4, pp. 120–165].

### 4.2.2 General Imbedding

The following regularity on the boundary of the domain is useful (but not necessary, since \((\varepsilon, \delta)\)-domain are also qualified, see the paper Jones [71]).

**Definition 4.15.** We say that \( \Omega \) satisfies the (interior) cone property (or condition) if there exists a fixed spherical cone \( \mathbb{K} \) (of some height \( h \) and solid angle \( \beta \)) such that each point \( x \) in \( \partial \Omega \) is the vertex of a cone \( \mathbb{K}_x \) contained in \( \Omega \) and congruent to \( \mathbb{K} \).

We say that an open set \( \Omega \) satisfies the strong local Lipschitz property if the conditions about smooth domain of Definition 4.4 are satisfied with uniform Lipschitz local coordinates, i.e., the functions \( f \) have the same Lipschitz continuous constant.

**Theorem 4.16** (Sobolev imbedding). Let \( \Omega \subset \mathbb{R}^d \) be an open set satisfying the cone property. Suppose \( m \) and \( n \) are two nonnegative integers and \( 1 \leq p < \infty \). Then the following continuous imbedding hold: (1) if \( [mp < d \text{ and } p \leq q \leq \frac{dp}{d-mp}] \) or \( [mp = d \text{ and } p \leq q < \infty] \) then \( W^{n+m,p}(\Omega) \subset W^{n,q}(\Omega) \), and (2) if \( [p = 1 \text{ and } m = d] \) or \( [mp > d] \) then \( W^{n+m,p}(\Omega) \subset C^n_b(\Omega) \). Moreover, if \( \Omega \) has the strong local Lipschitz property, then the case \( mp > d \) can be refined and the space \( C^n_b(\Omega) \) is replaced by the space \( C^{n,\alpha}(\Omega) \), with \( 0 < \alpha < m - d/p \) if \( p > mp - d > 0 \) and \( 0 < \alpha < 1 \) if \( p = mp - d \).

As expected, we prove the imbedding results by establishing a priori estimates in the corresponding norms for smooth functions, and then by density and smooth extension (see Theorem 4.3 and Theorem 4.6) we conclude. Hence, the above conclusions are valid for arbitrary domains provided the \( W \)-spaces are replaced with the corresponding \( W_0 \)-spaces (the closure of \( C_0^\infty \) in the \( W \)-norm. For instance, the interested reader may check the books Adams and Fournier [3], Maz’ya [88] and Nečas [93] for great details.

Proposition 4.8 yields the essential arguments to deduce Rellich-Kondrachov Theorem, which states that the above imbedding operator are compact if the domain \( \Omega \) is bounded, i.e., a uniformly bounded set in the initial space \( W^{n+m,p}(\Omega) \) becomes a totally bounded set in the other space \( W^{n,q}(\Omega) \), \( C^n_b(\Omega) \) or \( C^{n,\alpha}(\Omega) \).
4.3 Traces on the Boundary

The trace on the boundary of a given bounded continuous function \( f \) on a closed domain \( \bar{\Omega} \) is defined as the restriction of \( f \) over \( \partial \Omega \), Tietze’s extension (see e.g., Part I) ensures that any bounded continuous function \( f \) defined on the boundary \( \partial \Omega \) of a closed domain \( \bar{\Omega} \) can be extended to the whole domain as a bounded continuous function with sup-norm controlled. Thus, the linear extension operator is continuous from \( C^0_b(\partial \Omega) \) into \( C^0_b(\bar{\Omega}) \) and the linear operator trace (i.e., its left-inverse) is continuous from \( C^0_b(\bar{\Omega}) \) onto \( C^0_b(\partial \Omega) \). This also applies to the spaces \( C^k_b \), but some regularity on the boundary is required.

Since a bounded uniformly continuous function defined on an open set \( \Omega \) admits a unique extension to the boundary \( \partial \Omega \), the Hölder spaces \( C^0_{\alpha b}(\bar{\Omega}) \) and \( C^0_{\alpha}(\Omega) \) are equivalent under the norm \( \| \cdot \|_0 + [\cdot]_{\alpha} \), with

\[
[f]_{\alpha} = \sup \{ |f(x) - f(y)| |x - y|^{-\alpha} : x, y \in \Omega, |x - y| \leq 1 \},
\]

and \( \| \cdot \|_0 \) being the sup norm in \( \Omega \). Certainly, this argument can be apply to the space \( C^{k,\alpha}_{b}(\Omega) \), with \( k \) a positive integer and norm

\[
\|f\|_{k,\alpha} = \sup_{|\beta| \leq k} \|\partial^\beta f\|_0 + \sup_{|\beta| = k} [f]_{\alpha}.
\]

Thus, \( C^{k,\alpha}_{b}(\bar{\Omega}) \) could be defined as the extension to the closed domain \( \bar{\Omega} \) of functions in \( C^{k,\alpha}_{b}(\Omega) \).

Therefore, if the domain \( \Omega \) is of class \( C^{0,\alpha} \) as in Definition 4.4, then the linear trace operator can be defined from \( C^{0,\alpha}_{b}(\bar{\Omega}) \) onto \( C^{0,\alpha}_{b}(\partial \Omega) \). In general, if a domain \( \Omega \) is of class \( C^{k,\alpha} \) then the trace of the normal derivative with respect to \( \partial \Omega \) of order at most \( k \) are defined, i.e., the iteration of the operator \( \partial_n = \nabla \cdot n \), where \( n \) is the unit normal vector to \( \partial \Omega \) (usually, the exterior direction is chosen).

If \( \Omega \) is a smooth domain of \( \mathbb{R}^d \) and \( d < mp \) the Sobolev space \( W^{m,p}(\Omega) \) is imbedded into the Hölder space \( C^{0,\alpha}_{b}(\Omega) \) with any \( 0 < \alpha < m - d/p \), and the previous arguments proves that the trace of a function in \( W^{m,p}(\Omega) \) is defined and belongs to \( C^{0,\alpha}_{b}(\partial \Omega) \). However, if \( d > mp \) then the trace (or values on the boundary \( \partial \Omega \)) becomes in itself a relevant problem, in that functions in \( L^p(\Omega) \) may not be defined on a negligible set like the boundary \( \partial \Omega \).

4.3.1 In Half-space

In general, if the traces have been defined for half-space then by local coordinates, the same concepts are passed to smooth domains. Thus, the remaining of this section concern the space \( W^{m,p}(\mathbb{R}^d_+) \), with \( 1 \leq p < \infty \) and where \( x = (x,x_d) \), \( x' \) belongs to \( \mathbb{R}^{d-1} \) and \( x_d > 0 \). The boundary \( \partial \mathbb{R}^d_+ \) is identified with \( \mathbb{R}^{d-1} \).

Proposition 4.17. The trace operator, which initially defined for smooth functions as \( f \mapsto f|_{\mathbb{R}^{d-1}} = f(\cdot,0) \), can be extended to the whole Sobolev space as a linear continuous operator from \( W^{1,p}(\mathbb{R}^d_+) \) into \( L^p(\mathbb{R}^{d-1}) \), \( 1 \leq p < \infty \).
Proof. The identity
\[ u(x', 0) = -\int_0^\infty \partial_d u(x', x_d) dx_d \]
is valid for any continuously differentiable function \( u \) having a compact support in \( \mathbb{R}^d \). This yields the inequality
\[ \int_{\mathbb{R}^{d-1}} |u(x', 0)| dx' \leq \int_{\mathbb{R}_+^d} |\partial_d u(x)| dx. \]
In particular, if \( f \) is a continuously differentiable function with compact support then chose \( u = |f|^p \) with \( 1 < p < \infty \) to obtain
\[ \int_{\mathbb{R}^{d-1}} |f(x', 0)|^p dx' \leq p \int_{\mathbb{R}_+^d} |f(x)|^{p-1} |\partial_d f(x)| dx. \]
After noting that \((p-1)p' = p \) with \( 1/p + 1/p' = 1 \), use Hölder inequality to deduce
\[ \int_{\mathbb{R}^{d-1}} |f(x', 0)|^p dx' \leq p \left( \int_{\mathbb{R}_+^d} |f(x)|^p dx \right)^{1/p'} \left( \int_{\mathbb{R}_+^d} |\partial_d f(x)|^p dx \right)^{1/p}, \]
which yields the estimate
\[ \| f(\cdot, 0) \|_{L^p(\mathbb{R}^{d-1})} \leq p^{1/p} \| f \|_{L^p(\mathbb{R}_+^d)}^{1/p'} \| \partial_d f \|_{L^p(\mathbb{R}_+^d)}^{1/p}, \tag{4.15} \]
for any smooth function \( f \) having a compact support. Next, for \( p = 1 \) take \( u = f \) to deduce
\[ \| f(\cdot, 0) \|_{L^1(\mathbb{R}^{d-1})} \leq \| \partial_d f \|_{L^1(\mathbb{R}_+^d)}, \]
i.e., the limiting case of (4.15) as \( p \to 1 \).

Finally, since the restriction to \( \mathbb{R}_+^d \) of functions in \( C_0^\infty(\mathbb{R}^d) \) forms a dense set in \( W^{1,p}(\mathbb{R}_+^d) \), see Theorem 4.3, the trace operator (initially defined for smooth functions) \( f \mapsto f|_{\mathbb{R}^{d-1}} = f(\cdot, 0) \) can be extended to the whole Sobolev space and the estimate remain valid for any function \( f \) in \( W^{1,p}(\mathbb{R}_+^d) \).

Estimate (4.15) is certainly not optimal in the sense that, as mentioned earlier, the trace of a function in \( W^{1,p}(\mathbb{R}_+^d) \) is better than just a function in \( L^p(\mathbb{R}^{d-1}) \). For instant, if \( 1 \leq p < d \) then essentially the same previous argument with \( u = |f|^q \) and \( p \leq q = (dp - p)/(d - p) \) yields
\[ \int_{\mathbb{R}^{d-1}} |f(x', 0)|^q dx' \leq q \left( \int_{\mathbb{R}_+^d} |f(x)|^{(q-1)p'} dx \right)^{1/p'} \left( \int_{\mathbb{R}_+^d} |\partial_d f(x)|^p dx \right)^{1/p}. \]
This, combined with Sobolev-Gagliardo-Nirenberg’s estimate (4.10) applied to the symmetric extension \( \tilde{f}(x) = f(x', |x_d|) \) of the function \( f \), implies the estimate
\[ \| f(\cdot, 0) \|_{L^q(\mathbb{R}^{d-1})} \leq C \| \nabla f \|_{L^p(\mathbb{R}_+^d)}, \tag{4.16} \]
for some constant $C = C(d, p)$ and any function $f$ in $W^{1,p}(\mathbb{R}^d_+)$. 

On the other hand, the trace can also be defined for functions of bounded variation instead of functions in $W^{1,1}$ (i.e., $\partial_i f$ is a signed measure instead of a function in $L^1$), e.g., see Leoni [79, Chapter 15, pp. 451–476].

All this can be iterated, i.e., for a smooth function $f$ in $W^{m,p}(\Omega)$ the vector trace operator $(f(\cdot, 0), \partial_d f(\cdot, 0), \ldots, \partial_d^{m-1} f(\cdot, 0))$, i.e.,

$$f \mapsto T f = (f|_{\mathbb{R}^{d-1}}, \partial_d f|_{\mathbb{R}^{d-1}}, \ldots, \partial_d^{m-1} f|_{\mathbb{R}^{d-1}}), \quad (4.17)$$

can be extended to the whole Sobolev space as a linear continuous operator from $W^{m,p}(\mathbb{R}^d_+)$ into the product space

$$W^{m-1,p}(\mathbb{R}^{d-1}) \times W^{m-2,p}(\mathbb{R}^{d-1}) \times \cdots \times L^p(\mathbb{R}^{d-1}),$$

for any $1 \leq p < \infty$.

The meaning of a function with zero-trace is considered in the next

**Proposition 4.18.** Let $f$ be a function in $W^{m,p}(\mathbb{R}^d_+)$, with $1 \leq p, m < \infty$ and $m$ integer. Then the function $f$ belongs to $W^{m,p}_0(\mathbb{R}^d_+)$ if and only if $T f = 0$, see definition (4.17).

**Proof.** Since $W^{m,p}_0(\mathbb{R}^d_+)$ is the closure of the space of test functions $D(\mathbb{R}^d_+)$ and the trace operator $T$ is a continuous, it is clear that $T f = 0$, for any function $f$ in $W^{m,p}_0(\mathbb{R}^d_+)$. 

Without any loss of generality, assume $m = 1$ and note that the converse statement is more delicate. Indeed take a function $f$ in $W^{1,p}(\mathbb{R}^d_+)$ with zero-trace $f|_{\mathbb{R}^{d-1}} = 0$ and a smooth function $\zeta$ of one variable with values in $[0, 1]$ such that $\zeta = 0$ on $[-1/2, 1/2]$ and $\zeta = 1$ outside of $[-1, 1]$ to consider the sequence $\{f_k, k \geq 1\}$, with $f_k(x) = f(x)\zeta_k(x_d)$ and $\zeta_k(x_d) = \zeta(kx_d)$. Since $f_k$ vanishes near $x_d = 0$, the function $f_k$ belongs to $W^{1,p}_0(\mathbb{R}^d_+)$ for every $k$. Moreover, because $\partial_i f_k = \zeta_k \partial_i f$ for $i = 1, \ldots, d-1$, we have $f_k \rightarrow f$ and $\partial_i f_k \rightarrow \partial_i f$ in $L^p(\mathbb{R}^d)$. However,

$$\partial_d f_k(x) = \zeta(k\theta) \partial_d f(x', \theta) + f(x', \theta) k \zeta'(k\theta).$$

Thus, to prove that $f_k \rightarrow f$ in $W^{1,p}(\mathbb{R}^d_+)$, we need to show that

$$f(x', x_d) k \zeta'(kx_d) \rightarrow 0, \quad \text{in} \quad L^p(\mathbb{R}^d_+),$$

i.e.,

$$k^p \int_0^{1/k} dx_d \int_{\mathbb{R}^{d-1}} |f(x', x_d)|^p dx' \rightarrow 0 \quad \text{as} \quad k \rightarrow \infty,$$

because $|\zeta'(kx_d)|^p = 0$ if $kx_d < 1$. This will establish that $f$ belongs to the subspace $W^{1,p}_0(\mathbb{R}^d_+)$. 

To this purpose, and since the restriction to $\mathbb{R}^d_+$ of smooth functions with compact supports is a dense set in $W^{1,p}(\mathbb{R}^d_+)$, use the continuity of the trace to obtain the identity

$$f(x', x_d) = \int_0^{x_d} \partial_d f(x', \theta) d\theta, \quad \text{a.e.,}$$

for every function in $W^{1,p}(\mathbb{R}^d_+)$ with zero-trace $f|_{\mathbb{R}^{d-1}} = 0$.

Now, integrate in $\mathbb{R}^{d-1}$ and apply Hölder inequality to get

$$\int_{\mathbb{R}^{d-1}} |f(x', x_d)|^p dx' \leq x_d^{p-1} \int_0^{x_d} d\theta \int_{\mathbb{R}^{d-1}} |\partial_d f(x', \theta)|^p dx', \quad \text{a.e.,}$$

which yields

$$\int_0^{1/k} dx_d \int_{\mathbb{R}^{d-1}} |f(x', x_d)|^p \leq \left( \int_0^{1/k} x_d^{p-1} dx_d \right) \left( \int_0^{1/k} d\theta \int_{\mathbb{R}^{d-1}} |\partial_d f(x', \theta)|^p dx' \right),$$

i.e.,

$$k^p \int_0^{1/k} dx_d \int_{\mathbb{R}^{d-1}} |f(x', x_d)|^p dx' \leq p^{p-1} \int_0^{1/k} d\theta \int_{\mathbb{R}^{d-1}} |\partial_d f(x', \theta)|^p dx',$$

and the desired convergence follows.

4.3.2 In a Smooth Domain

Recall that on a Lipschitz domain (see Definition 4.4, with $m = 0$ and $\alpha = 1$), the outward unit normal vector $n$ is defined almost everywhere with respect to the surface area $d\sigma$ (or measure $dx'$) on the boundary $\partial\Omega$. It is clear that if the domain is better, say of class $C^{m,\alpha}$ with $m \geq 1$, then the outward unit normal is of class $C^{m-1,\alpha}$.

Similarly to Theorem 4.3, it can be shown that actually the space of smooth functions with compact support in $\mathbb{R}^d$ restricted to $\Omega$, i.e., $C^\infty_0(\mathbb{R}^d)|_{\Omega}$, is dense in $W^{s,p}(\Omega)$ for any $s \leq 0$, while $s > 0$ this density holds for continuous domains, i.e., Definition 4.4, with $m = 0$ and $\alpha = 0$.

Note that for a Lipschitz domain the (outward) unit normal direction $n$ is defined almost everywhere with respect to the surface measure (denoted by either $d\sigma$ or just $dx'$) on the boundary $\partial\Omega$. Moreover this definition is invariant under a bi-Lipschitz homeomorphism.

Therefore, once everything have been done on the half-space $\mathbb{R}^d_+$ by means of local coordinates everything is transport to a smooth domain $\Omega$. For instance, the trace on the boundary $\partial\Omega$ becomes a linear continuous operator from $W^{1,p}(\Omega)$ into $L^p(\partial\Omega)$, where the surface area $d\sigma$ or (image) measure $dx'$
is used on $\partial \Omega$, i.e., Propositions 4.17 and 4.18 are valid with $\Omega$ and $\partial \Omega$ in lieu of $\mathbb{R}^d_+$ and $\mathbb{R}^{d-1}$. Equation (4.16) becomes

$$
\|f\|_{L^q(\partial \Omega)} \leq C \|\nabla f\|_{L^p(\Omega)}, \quad \forall f \in W^{1,p}(\Omega), \ 1 \leq p < \infty,
$$
(4.18)

for any $q = (dp - p)/(d - p)$ if $1 \leq p < d$ and some constant $C = C(d, p)$.

**Remark 4.19.** Recall that the above estimate is valid also for $q = p$ and therefore, because $W^{1,r}(\Omega) \supset W^{1,q}(\Omega) \cap W^{1,p}(\Omega)$ with $q^* = (dp - p)/(d - p)$ if $1 \leq p < d$, this estimate holds for any $q$ in $[p, q^*]$. Moreover, since $W^{1,r}(\Omega) \subset W^{1,p}(\Omega)$ when $1 \leq r < d \leq p$ and $\Omega$ is bounded, we can use this estimate with $q = (dr - r)/(d - r)$. This means that estimate (4.18) is valid (a) for any $p \leq q \leq (dp - p)/(d - p)$ if $1 \leq p < d$ or (b) for any $1 \leq q < \infty$ if $\Omega$ is bounded and $p \geq d$. \hfill $\Box$

The vector trace operator $\mathbb{T}$ (also denoted by $\gamma$) (4.17) becomes

$$
f \mapsto \mathbb{T}f = (f|_{\partial \Omega}, \partial_n f|_{\partial \Omega}, \ldots, \partial_n^{m-1} f|_{\partial \Omega}),
$$
(4.19)

where $\gamma_0 f = f|_{\partial \Omega}, \gamma_1 f = \partial_n f|_{\partial \Omega}$, etc., are the derivatives in the outward unit normal direction $n$ to the boundary $\partial \Omega$. The outward unit normal direction is chosen by convention, but remark that in $\mathbb{R}^d_+$ we used the inner unit normal direction $\partial_d$.

However, it is clearly interesting to obtain similar results with minimal use of localization arguments, i.e., working directly on the boundary. To this purpose, first we need a preliminary

**Lemma 4.20.** If $\Omega$ is a Lipschitz domain of $\mathbb{R}^d$ with a bounded boundary $\partial \Omega$ then there exists a smooth almost everywhere unit normal, i.e., there is $\delta > 0$ and $\nu = (\nu_1, \ldots, \nu_d)$ with $\nu_i$ in $D(\mathbb{R}^d)$ such that $\nu \cdot n \geq \delta$ almost everywhere in $\partial \Omega$, where $n$ is the outward unit normal direction defined almost everywhere with respect to the surface measure on the boundary $\partial \Omega$.

**Proof.** There are several way of construction a smooth almost everywhere normal, for instance a regularization of the gradient of the distance to the boundary or just using the local coordinates. Since the boundary $\partial \Omega$ is locally (on a neighborhood $B$) given by $\partial \Omega \cap B = \{y_d = \phi(y')\}$ and $\Omega \cap B = \{y_d < \phi(y')\}$, with $y' = (y_1, \ldots, y_{d-1})$ and a Lipschitz function $\phi$, it is clear that by choosing $\nu$ equal to the unit vector in the direction of $y_d$, the component of the outward unit normal $n$ in the direction $y_d$ is $[1 + |\nabla \phi(y')|^2]^{-1/2}$ and therefore, $\nu \cdot n \geq (1 + L^2_{\phi})^{-1/2}$, where $L_{\phi}$ is the Lipschitz constant of $\phi$ on $B$. Next, because the boundary $\partial \Omega$ is bounded, we can find a finite number of open balls $B_i$ covering $\partial \Omega$ where the previous construction takes place and $\nu|_{B_i}$ is a smooth function. Now, invoking a partition of the unity, there are smooth functions $0 \leq \chi_i \leq 1$ with support in $B_i$ such that $1 = \sum_i \chi_i$, and defining $\nu = \sum_i \chi_i \nu|_{B_i}$, we check that $\nu$ is a smooth function with a compact support such that $\nu \cdot n \geq \delta$ with $\delta = \inf_i (1 + L^2_{\phi_i})^{-1/2} > 0$. \hfill $\Box$
• Remark 4.21. For any closed set $F$ of $\mathbb{R}^d$ the distance $x \mapsto d(x, F)$ is a Lipschitz continuous function, which vanishes on $F$ and in general, not very smooth on the complement $\mathbb{R}^d \setminus F$. However there exists a regularized distance function $\delta(x, F)$, which is Lipschitz continuous and infinitely differentiable on the open complement set $\mathbb{R}^d \setminus F$ with the properties: (a) $cd(x, F) \leq \delta(x, F) \leq C d(x, F)$, and (b) $|\partial^\alpha \delta(x, F)| \leq C_\alpha[d(x, F)]^{-|\alpha|}$, for every $x$ in $\mathbb{R}^d \setminus F$, any non-zero multi-index $\alpha$, and some constants $c, C$ and $C_\alpha$ independent of $F$. The construction of $\delta(x, F)$ is based on the decomposition of open sets in cubes. This technique produces a continuous linear extension operator in several spaces, beginning with Lipschitz/Hölder functions defined on a closed set $F$ and (b) $|d|d_{\partial\Omega}$ on the complement $\mathbb{R}^d \setminus F$, which vanishes on $F$ and in general, not very smooth on the complement $\mathbb{R}^d \setminus F$. However there exists a regularized distance function $\delta(x, F)$, which is Lipschitz continuous and infinitely differentiable on the open complement set $\mathbb{R}^d \setminus F$ with the properties: (a) $cd(x, F) \leq \delta(x, F) \leq C d(x, F)$, and (b) $|\partial^\alpha \delta(x, F)| \leq C_\alpha[d(x, F)]^{-|\alpha|}$, for every $x$ in $\mathbb{R}^d \setminus F$, any non-zero multi-index $\alpha$, and some constants $c, C$ and $C_\alpha$ independent of $F$. The construction of $\delta(x, F)$ is based on the decomposition of open sets in cubes. This technique produces a continuous linear extension operator in several spaces, beginning with Lipschitz/Hölder functions defined on a closed set $F$ to Lipschitz/Hölder functions defined on the whole space $\mathbb{R}^d$, e.g., see Stein [113, Chapter VI, Theorem 2, pp. 167–171].

Certainly, if the boundary $\partial \Omega$ is unbounded then some extra condition (on the domain $\Omega$) could be added to retain the existence of a smooth (not necessarily with a compact support) almost everywhere outward unit normal direction.

**Theorem 4.22.** Let $\Omega$ be a Lipschitz domain of $\mathbb{R}^d$ with a bounded boundary $\partial \Omega$. (1) The trace on (or the restriction to) the boundary of a function $f$ in $W^{1,p}(\Omega)$ with $1 \leq 1 < \infty$ can be defined almost everywhere with respect to the surface area (or measure) $\sigma$ on $\partial \Omega$, and there exists a constant $K$ depending only on the norm in $C^1(\overline{\Omega})$ of $\nu$ and the constant $\delta$ of Lemma 4.20 such that

$$\int_{\partial \Omega} |f(x)|^p d\sigma(x) \leq K \left[ \varepsilon^{1-1/p} \int_\Omega |\nabla f(x)|^p \, dx + p \varepsilon^{-1/p} \int_\Omega |f(x)|^p \, dx \right],$$

for every $\varepsilon$ in $(0, 1)$. (2) If $f$ belongs to $W^{1,p}(\Omega)$ and $g$ belongs to $W^{1,p'}(\Omega)$ with $1/p + 1/p' = 1$ then the following integration-by-part formula holds

$$\int_\Omega (\partial_i f(x)) g(x) \, dx + \int_\Omega f(x)(\partial_i g(x)) \, dx = \int_{\partial \Omega} f(x)g(x)\nu_i(x) \, d\sigma(x),$$

where $\nu_i$ is the $i$-component in the outward unit direction which is defined almost everywhere with respect to the surface area or measure $\sigma$ on the boundary $\partial \Omega$, and values on the boundary of the functions $f$ and $g$ are understood in the sense of the trace over $\partial \Omega$.

**Proof.** First, note that in view of the density, it suffices to prove the estimate for continuously differentiable functions $f$ and $g$ with support in some compact ball of $\mathbb{R}^d$. Essentially with the same localization arguments used in Lemma 4.20 and a bi-Lipschitz change of variables, we can check that the integration-by-part formula and Green Theorem remain valid on Lipschitz domains, for smooth functions $f$ and $g$ as above. Thus, only the estimate needs to be proved.

Therefore, Green Theorem yields

$$\int_\Omega \nabla |f(x)|^p \cdot \nu(x) \, dx = \int_{\partial \Omega} |f(x)|^p (\nu(x) \cdot \nu(x)) \, d\sigma(x) - \int_\Omega |f(x)|^p (\nabla \cdot \nu(x)) \, dx,$$
where $\nu$ is the smooth outward unit normal as in Lemma 4.20. Since $\nabla |f|^p = p|f|^{p-2}f \nabla f, \nu \cdot n \geq \delta$, $|\nu| \leq C_0$ and $|\nabla \cdot \nu| \leq C_1$, we obtain

$$\delta \int_{\partial \Omega} |f|^p d\sigma(x) \leq pC_0 \int_{\Omega} |f(x)|^{p-1}|\nabla f(x)| dx + C_1 \int_{\Omega} |f(x)|^p dx.$$ 

Hence, note that if $1/p + 1/p' = 1$ then $(p - 1)p' = p$ and use Hölder inequality to get

$$p \int_{\Omega} |f(x)|^{p-1}|\nabla f(x)| dx \leq p \left( \int_{\Omega} |f(x)|^p dx \right)^{1/p'} \left( \int_{\Omega} |\nabla f(x)|^p dx \right)^{1/p}.$$ 

Next, apply the inequality $ab \leq a^p/p + b^{p'}/p'$ to deduce

$$p \int_{\Omega} |f(x)|^{p-1}|\nabla f(x)| dx \leq \varepsilon^{1-1/p} \int_{\Omega} |\nabla f(x)|^p dx + \frac{p\varepsilon^{-1/p}}{p'} \int_{\Omega} |f(x)|^p dx.$$ 

Finally, choose $K = \max\{C_0, C_1\}/\delta$ to conclude the argument. \qed

The interested reader may take a look at Green’s Formula in Tartar [121, Lectures 12–14, pp. 59–72] to compare arguments and other insides.

### 4.3.3 Spaces on the Boundary

Since bi-Lipschitz change-of-variables are allowed in the integrals, global properties on the Sobolev spaces $W^{m,p}(\Omega)$ can be studied by means of local coordinates in Lipschitz domains. However, some more specific properties may need more regularity, e.g., if the unit normal vector of class $C^{k,\alpha}$ is involved then a domain of class $C^{k+1,\alpha}$ is required.

For a Lipschitz domain $\Omega$ in $\mathbb{R}^d$, the unit normal vector $n$ is defined almost everywhere with respect to the surface as $d\sigma(x)$ or measure $dx'$ on the boundary $\partial \Omega$. Thus, the trace $\partial_n^k$ is a linear continuous operator from $W^{m,p}(\Omega)$ into $L^p(\partial \Omega)$ for any nonnegative integer numbers $m \geq k + 1$ and $1 \leq p < \infty$. As mentioned early, the case $p = \infty$ is easier and treated independently.

If $f$ is a smooth function in $W^{m,p}(\Omega)$ then $\partial_n^1 f = f|_{\partial \Omega}$ is the restriction to the boundary of the function $f$, $\partial_n^1 f = n \cdot \nabla f|_{\partial \Omega}$ is the restriction to the boundary of the first derivative of the function $f$ in the direction of the unit normal vector $n$, $\partial_n^2 f = n \cdot (\nabla^2 f) n|_{\partial \Omega}$ is the restriction to the boundary of the second derivative of the function $f$ in the direction of the unit normal vector $n$, etc., where $\nabla f$ is the vector of all first derivatives (the gradient) and $\nabla^2 f$ is the matrix of all second derivatives (the hessian). It is clear that everything reduces to the case of the trace $\partial_n^0$ considered as linear continuous operator from $W^{1,p}(\Omega)$ into $L^p(\partial \Omega)$. Based on the extension operator considered in Theorem 4.6, we would need a function in the space $W^{1,p}(\partial \Omega)$ to obtain an extension in $W^{1,p}(\Omega)$, which trace belongs to $L^p(\partial \Omega)$, i.e., we cannot expect to have a surjective trace operator from $W^{1,p}(\Omega)$ onto $L^p(\partial \Omega)$.

Based on the density of smooth functions with compact support in $\mathbb{R}^d$ restricted to $\Omega$ (i.e., $C_0^\infty(\mathbb{R}^d)|_{\Omega}$ as in Theorem 4.3), we could define the Sobolev
space on the boundary $W^{m,p}(\partial \Omega)$ as the closure of the space of smooth functions $f$ with compact support in $\mathbb{R}^d$ such that all normal derivatives $\partial^k_n f$ vanish on the boundary $\partial \Omega$, $k \geq 1$, with the norm

$$
\|f\|_{m,p} = \left( \sum_{|\alpha| \leq m} \int_{\partial \Omega} |\partial^\alpha_x f(x)|^p \, dx \right)^{1/p}, \quad (4.20)
$$

where $\partial^\alpha_x$ are the tangential derivatives corresponding to the $(d-1)$-dimensional multi-index $\alpha$ of order $|\alpha|$, i.e., if $t_i$, are $(d-1)$ linear independent tangential unit vectors (see (4.2)) and $\alpha_i = 1$ and $\alpha_j = 0$ for $j \neq i$, then $\partial^\alpha_x = t_i \cdot \nabla$. Actually, because all normal derivatives satisfy $\partial^k_n f = 0$ on the boundary, $k \geq 1$, the above norm can be written as

$$
\|f\|_{m,p} = \left( \sum_{|\alpha| \leq m} \int_{\partial \Omega} |\partial^\alpha f(x)|^p \, dx \right)^{1/p},
$$

where now, the sum is over $d$-dimensional multi-index $\alpha$.

Alternatively, if $\Omega$ is a bounded domain of class $C^{m,0}$ then the boundary $\partial \Omega$ is locally mapped into the flat part of the boundary of $\{ y \in \mathbb{R}^d_+ : |y| < 1 \}$, i.e., a bounded domain $B'$ of $\mathbb{R}^{d-1}$. Thus, the space $W^{m,p}(\partial \Omega)$ could be defined as a local image of $W^{m,p}(B')$. In any case, we expect that trace operator will map $W^{1,p}(\Omega)$ into some space between $L^p(\partial \Omega)$ and $W^{1,p}(\partial \Omega)$.

Since a function $f$ belongs to $W^{1,p}_0(\Omega)$ if and only if $f$ belongs to $W^{1,p}(\Omega)$ and the trace $f|_{\partial \Omega} = 0$ (see Proposition 4.18), to study the image of the trace reduces to the space of equivalence classes of functions $\tilde{f}$ defined by the relation $f = g$ if $f - g$ belongs to $W^{1,p}_0(\Omega)$, i.e., the null space of the trace operator. For reasons that will be clarify later, this quotient space is denoted by

$$
W^{1/p-1/p}(\partial \Omega) = W^{1,p}(\Omega)/W^{1,p}_0(\Omega).
$$

The quotient norm is given by

$$
\|f\|_{W^{m-1/p,p}(\partial \Omega)} = \inf \{ \|g\|_{W^{m,p}(\Omega)} : g - f \in W^{m,p}_0(\Omega) \}, \quad m \geq 1, \quad (4.21)
$$

and therefore, $W^{m-1/p,p}(\partial \Omega)$ becomes a Banach space. For a Lipschitz domain $\Omega$, this quotient space is isomorphic to the closure of the space $C^{0,0}_0(\mathbb{R}^d)|_{\partial \Omega}$ with the quotient norm (4.21). Note that if $\partial_t$ is a tangential derivative on the boundary (i.e., the direction $t$ is orthogonal to the normal direction $n$) then $\partial_t$ is a continuous linear operator from $W^{m-1/p,p}(\partial \Omega)$ into $W^{m-1/p,p}(\partial \Omega)$, for $m \geq 2$. Also, in view of Remark 4.19, for some constant $C > 0$,

$$
\|f\|_{L^q(\partial \Omega)} \leq C \|f\|_{W^{1-1/p,p}(\partial \Omega)}, \quad \forall f \in W^{1-1/p,p}(\partial \Omega),
$$

for any $p \leq q \leq (dp - p)/(d - p)$ if $1 \leq p < d$ and any $1 \leq q < \infty$ if $p \geq d$ and $\Omega$ is bounded.

This argument shows that the vector trace operator (4.19) is a linear continuous operator from $W^{m,p}(\Omega)$ onto the product space

$$
W^{m-1/p,p}(\partial \Omega) \times W^{m-1-1/p,p}(\partial \Omega) \times \cdots \times W^{1-1/p,p}(\partial \Omega).
$$
Moreover, if $\Gamma$ is a closed part of the boundary $\partial \Omega$ with the same smoothness and dimension as boundary, then we can define $W^{m,p}_0(\Omega \cup \Gamma) \subset W^{m,p}(\Omega)$ as the closure of smooth functions with compact support separated from $\partial \Omega \setminus \Gamma$. Therefore, we can also define the quotient space $W^{m-1/p,p}(\partial \Omega)$, which is isomorphic to the subspace of $W^{m-1/p,p}(\partial \Omega)$ with trace vanishing on $\partial \Omega \setminus \Gamma$. For instance, the read may take a look at Troianiello [131, Section 1.7, pp. 64-76).

4.4 Fractional Order Spaces

Firstly, let us recall the space $C^{\alpha}_{b}(\Omega)$, with $0 < \alpha < 1$, of bounded Hölder continuous functions, which is sometimes denoted by $C^{0,\alpha}_{b}(\Omega)$, for $0 < \alpha \leq 1$. The subindex $b$ stands for bounded, and is unnecessary when $\Omega$ is bounded. The relevant seminorm may takes several forms (all equivalent when $\Omega$ is a relatively smooth domain), e.g., $\| \cdot \|_{0}$ is the sup-norm in $\Omega$, and for $0 < \alpha \leq 1$,

$$\|f\|_{\alpha} = \sup\{ |f(x) - f(y)| |x - y|^{-\alpha} : x, y \in \Omega, |x - y| \leq 1 \},$$

(4.22)

with the norm $\| \cdot \|_{\alpha} = \| \cdot \|_{0} + |\cdot|_{\alpha}$. The limiting case $C^{0,1}_{b}(\Omega)$ produces the Lipschitz functions, which is larger than the typical space $C^{1}_{b}(\Omega)$ of bounded and continuously differentiable functions with bounded derivatives on the open set $\Omega$, with the sup-norm $\| \cdot \|_{0} + \|\nabla \cdot\|_{0}$. Since a uniformly continuous function on the open set $\Omega$ can be uniquely extended to the closure $\overline{\Omega}$ by continuity, there is not difference between the spaces $C^{0,\alpha}_{b}(\Omega)$ and $C^{0,\alpha}_{b}(\overline{\Omega})$. However, $C^{1}_{b}(\Omega)$ may be strictly larger than $C^{1}_{b}(\overline{\Omega})$, where all first derivative can be uniquely extended to the closure $\overline{\Omega}$ by assumption. If $\Omega$ is a reasonable smooth domain then $C^{1}_{b}(\overline{\Omega}) \subset C^{0,\alpha}_{b}(\overline{\Omega})$, for any $\alpha$. For instance, this holds if $\Omega$ is convex, or in general, if for every two points $x$ and $y$ in $\overline{\Omega}$ there exists a Lipschitz curve $\gamma$ entirely contained in $\overline{\Omega}$ such that $\gamma(0) = x$, $\gamma(1) = y$, and the the Lipschitz constant of $\gamma$ is bounded by a constant depending only on $\Omega$. Certainly, the local version are denoted by $C^{\alpha}_{\text{loc}}(\Omega)$, which means functions in $C^{\alpha}_{b}(\omega)$ for any open set $\omega$ with compact closure inside the open set $\Omega$.

4.4.1 Discussion and Definition

There are several ways of defining Sobolev spaces of fractional order (sometimes called Sobolev-Slobodeckij spaces), but in any case, to define $W^{s,p}(\Omega)$ and $W^{s,p}_0(\Omega)$, for any real number, it suffices to consider only $0 < s < 1$. This means that (a) for $s < 0$ the definition is by duality, (b) if $s = m + \sigma$, $m$ a positive integer and $0 < \sigma < 1$ then $W^{s,p}(\Omega)$ is defined as the functions in $W^{m,p}(\Omega)$ such that all derivative of order $m$ belong to $W^{\sigma,p}(\Omega)$, and analogously for $W^{s,p}_0(\Omega)$. Thus our discussion focus on $W^{s,p}(\Omega)$ and $W^{s,p}_0(\Omega)$, for $0 < s < 1 \leq p \leq \infty$.

For $p = \infty$, these spaces $W^{s,\infty}(\Omega)$, $0 < s < 1$ are the spaces of bounded Hölder continuous functions $C^{0,s}_{b}(\Omega)$ just discussed. The limiting case $W^{1,\infty}(\Omega)$ produces the Lipschitz functions. Using the fact that a function is Lipschitz if and only it all its first weak derivatives are essentially bounded, the notation
$W^{1,\infty}(\Omega)$ is equivalent to the bounded Lipschitz space $C^{0,1}_b(\Omega)$, if $\Omega$ is a reasonable smooth domain, see above. Because of the continuity up-to the boundary, the space $W^{s,\infty}_0(\Omega)$ could be defined as the subspace in $W^{s,\infty}(\Omega)$ of all functions with zero-value on the boundary. Therefore, the case $s = 1$ and $p = \infty$ are not included in the following analysis.

Based on the above discussion, consider the following $L^p$-type Hölder seminorm

$$|f|_{s,p} = \left( \int \int_{|x-y| \leq 1} |f(x) - f(y)|^p |x-y|^{-d-ps} \, dx \, dy \right)^{1/p}, \quad (4.23)$$

for any $0 < s < 1 \leq p < \infty$. Clearly, with a change of variables, the above integral becomes

$$\int_{|z| \leq 1} |z|^{-d-ps} \, dz \int_{\mathbb{R}^d} |f(y + z) - f(y)|^p \mathbb{1}_{\{y \in \Omega : y + z \in \Omega\}} \, dy,$$

and due to the integrability of $z \mapsto |z|^{-d-ps}$ over the region $\{z \in \mathbb{R}^d : |z| > 1\}$, the integral could be extended to the whole product space $\Omega \times \Omega$ instead of just the band $\{(x, y) \in \Omega \times \Omega : |x - y| \leq 1\}$. It is also clear that

$$\int_{\Omega} |f(y + z) - f(y)|^p \, dy \leq 2^p |z|^p \int_{\Omega} |\nabla f(x)|^p \, dx,$$

which implies that $|f|_{s,p} < \infty$ if $f$ belongs to $W^{1,p}(\Omega)$ and $\Omega$ is a reasonable smooth domain, as defined early. It is clear that $\alpha = 1$ is not a desirable choice in definition (4.23). Sometimes, it may relevant to include the set $\Omega$, i.e., the notation $| \cdot |_{s,p,\Omega}$ could used. For the limiting case $p = \infty$, remark that $| \cdot |_{s,\infty}$ or $[\cdot]_{\alpha,\Omega}$ or $| \cdot |_{s,\infty,\Omega}$ means the Hölder seminorm $[\cdot]_{\alpha}$ with $\alpha = s$.

Therefore, if $\| \cdot \|_p$ denotes the norm in the Lebesgue space $L^p(\mathbb{R}^d)$ then the expression

$$\| \cdot \|_{s,p} = \left( \| \cdot \|_p^p + | \cdot |_{s,p}^p \right)^{1/p}$$

is a norm on $W^{1,p}(\Omega)$ if $\Omega$ is a reasonable smooth domain. Hence, the Sobolev space $W^{s,p}(\Omega)$, $0 < s < 1 \leq p < \infty$, is defined as all functions $f$ in $L^p(\Omega)$ such that $|f|_{s,p} < \infty$, which becomes a Banach space with the norm $\| \cdot \|_{s,p}$, and a Hilbert space if $p = 2$. While, the Sobolev space $W^{s,p}_0(\Omega)$ is the closure of $C^{0,\infty}(\Omega)$ or $C^1(\Omega)$ in $W^{s,p}(\Omega)$. Local Sobolev space of fractional order can be defined too, i.e., $W^{s,p}_{lo}(\Omega)$ are functions in $W^{s,p}(\omega)$ for any open set $\omega$ with compact closure inside the open set $\Omega$. Under the above definition, if $\Omega$ is a reasonable smooth domain then $W^{s,p}(\Omega) \subset W^{1,p}(\Omega)$, $0 < s < 1 \leq p < \infty$, as expected (for $p = \infty$, this inclusion holds always by definition).

### 4.4.2 Basic Properties

It is clear that the product of a function in $C_b^1(\Omega)$ and a function in $W^{1,p}(\Omega)$ belongs again to $W^{1,p}(\Omega)$, i.e.,

$$\| \nabla (fg) \|_p \leq \| f \|_\infty \| \nabla g \|_p + \| \nabla f \|_\infty \| g \|_p, \quad \forall f \in C^1(\overline{\Omega}), \ g \in W^{1,p}(\Omega),$$
for any $1 \leq p \leq \infty$. In particular, $C^1_{\text{loc}}(\Omega) \subset W^{1,p}_{\text{loc}}(\Omega)$.

Similarly, the inequality
\[ |f(x)g(x) - f(y)g(y)| \leq |f(x)||g(x) - g(y)| + |f(x) - f(y)||g(y)| \]
implies that
\[ |fg|_{s,p} \leq \|f\|_{\infty} |g|_{s,p} + 2[f]_\alpha \|g\|_{p} \left( \int_{|z| \leq 1} |z|^{-d+p(\alpha-s)}dz \right)^{1/p} \]
where the integral is equal to $\omega_d/p(\alpha-s)$ if $\alpha > s$, with the area of the unit sphere $\omega_d = 2\pi^{d/2}/\Gamma(d/2)$. Thus, the product $(f,g) \mapsto fg$ is a continuous operation from $C^0,\alpha(\Omega) \times W^{s,p}(\Omega)$ into $W^{s,p}(\Omega)$, if $0 < s < \alpha \leq 1 \leq p \leq \infty$. In particular $C^0,\alpha(\Omega) \subset C^1_{\text{loc}}(\Omega)$. Since $W^{0,p}(\Omega) = L^p(\Omega)$, note that the limiting case $s = 0$ is the known fact that the multiplication is a continuous operation from $L^\infty(\Omega) \times L^p(\Omega)$ into $L^p(\Omega)$.

By the definition and by duality, it is also clear that the differentiation $f \mapsto \partial^\alpha f$ is a continuous operator from $W^{s,p}(\Omega)$ into $W^{s-|\alpha|,p}(\Omega)$, for any multi-index $\alpha$ with $|\alpha| \leq s$ or $s \leq 0$. However the case $s < |\alpha|$ is not so obvious. Following Grisvard [60, Section 1.4, pp. 20–36] and Nečas [93]. First discuss a convolution kernel approximation results interest by itself.

**Lemma 4.23.** Let $\rho$ be continuously differentiable function supported in the unit ball of $\mathbb{R}^d$ with integral equal to one, and consider the convolution $u_\rho(x,t) = (\rho_t \ast u)(x)$ with $\rho_t(x) = t^{-n}\rho(tx^{-1})$, $x \in \mathbb{R}^d$ and $t > 0$, for any function $u$ in $W^{s,p}(\mathbb{R}^d)$, $0 < s < 1 \leq p < \infty$. Then $u_\rho(\cdot,t) \rightarrow u$ as $t \rightarrow 0$ and $u_\rho(\cdot,t) \rightarrow 0$ as $t \rightarrow \infty$ in the norm of $L^p(\mathbb{R}^d)$. Moreover, the function $(x,t) \mapsto t^{-s-1/p} \partial_i u_\rho(x,t)$ belongs to $L^p(\mathbb{R}^d \times [0,\infty[)$ and the following estimate holds
\[
\int_0^\infty t^{p-ps-1}dt \int_{\mathbb{R}^d} |\partial_i u_\rho(x,t)|^p dx \leq \frac{\|\partial_i \rho\|_q}{sp+n} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|u(x) - u(y)|^p}{|x-y|^{n+sp}} dxdy,
\]
with $1/p + 1/q = 1$ and for any first partial derivative $\partial_i$ with respect to either $x_i$, $i = 1,\ldots,d$. Furthermore, $\partial_i u_\rho = d\rho \partial_t + \rho_t \ast u$, where $\rho_t(x) = x\nabla \rho(x)$ and the above estimate holds for $\partial_t$ replacing $\partial_i$ and $\|d\rho + \rho_t\|_q$ replacing $\|\partial_i \rho\|_q$.

**Proof.** Since the integral of $\rho$ is one, the integral of $\partial_i \rho$ is zero and
\[
\partial_i u_\rho(x,t) = t^{-1} \int_{\mathbb{R}^d} \partial_i \rho_t(x-y)[u(x) - u(y)]dy
\]
Hence, use the change of variable $x - y = tz$, the fact that $\rho$ vanish outside the unit ball and Hölder inequality to deduce that
\[
|\partial_i u_\rho(x,t)|^p \leq t^{-p}\|\partial_i \rho\|_q^p \left( \int_{|z| \leq 1} |u(x) - u(x-tz)|^p dz \right).
\]
Therefore,
\[
\|\partial_i \rho\|_q^{-p} \int_0^\infty t^{p-sp-1} dt \int_{\mathbb{R}^d} |\partial_i u_\rho(x,t)|^p dx \leq \\
\leq \int_0^\infty t^{sp-1} dt \int_{\mathbb{R}^d} dx \int_{|z| \leq 1} |u(x) - u(x - tz)|^p dz = \\
= \int_0^\infty t^{-d-sp-1} dt \int_{|x-y| \leq t} |u(x) - u(y)|^p dx dy 
\]
and exchanging the order of the integrals
\[
\|\partial_i \rho\|_q^{-p} \int_0^\infty t^{p-sp-1} dt \int_{\mathbb{R}^d} |\partial_i u_\rho(x,t)|^p dx \leq \\
\leq \int \int_{\mathbb{R}^d \times \mathbb{R}^d} |u(x) - u(y)|^p dx dy \int_0^\infty t^{-d-sp-1} dt = \\
= (d + sp)^{-1} \int \int_{\mathbb{R}^d \times \mathbb{R}^d} |u(x) - u(y)|^p |x-y|^{d+sp} dx dy,
\]
as desired.

Now, calculate the derivative in \( t \),
\[
\partial_t \rho_t(x) = dt^{-1} \rho_t(x) + t^{-1} \rho'_t(x)
\]
to deduce the identity \( \partial_t u_\rho = du_\rho + \rho'_t * u \). Next, an integration by parts shows that
\[
d \int_{\mathbb{R}^d} \rho(x) dx + \int_{\mathbb{R}^d} x \cdot \nabla \rho(x) dx = 0,
\]
which allows us to obtain the estimate for \( \partial_t u_\rho \) similarly to the \( x_i \)-derivative.

Finally, the convergence as \( t \to 0 \) and \( t \to \infty \) can be proved first for smooth function and then extended by density to any \( p \)-integrable function. \( \square \)

**Proposition 4.24.** The differentiation \( f \mapsto \partial^\alpha f \) is a continuous operator from \( W^{s,p}(\Omega) \) into \( W^{s-|\alpha|,p}(\Omega) \), for any multi-index \( \alpha \) and any real number \( s \).

**Proof.** As mentioned early, only the case \( 0 < s < 1 \leq p < \infty \) needs to be considered. To prove that \( \partial_t \) maps \( W^{s,p}(\mathbb{R}^d) \) into \( W^{s-1,p}(\mathbb{R}^d) \), the dual space of \( W^{1-s,q}(\mathbb{R}^d) \) with \( 1/p + 1/q = 1 \), we have to show that the bilinear form
\[
\langle u, v \rangle = \int_{\mathbb{R}^d} u(x) v(x) dx
\]
is defined and continuous on the product space \( W^{s,p}(\mathbb{R}^d) \times W^{1-s,q}(\mathbb{R}^d) \).

To this purpose, use Lemma 4.23 to write
\[
\langle \partial_i u_\rho(\cdot,t), v_\rho(\cdot,t) \rangle = \\
= \int_t^\infty dr \int_{\mathbb{R}^d} [\partial_t u_\rho(x,r) \partial_i v_\rho(x,r) - \partial_i u_\rho(x,r) \partial_t v_\rho(x,r)] dr.
\]
Now, use $1 = t^{1-s-1/p}t^{s-1/q}$, with $1/p + 1/q = 1$, and Hölder inequality to obtain

$$\left| \langle \partial_i u_p(\cdot,t), v_p(\cdot,t) \rangle \right| \leq \left( \|t^{1-s-1/p}\partial_t u(x,t)\|_p \|t^{s-1/p}\partial_t v(x,t)\|_q + \|t^{1-s-1/p}\partial_i u(x,t)\|_p \|t^{s-1/p}\partial_i v(x,t)\|_q \right),$$

with $\| \cdot \|_p$ and $\| \cdot \|_q$ are the norms in $L^p(\mathbb{R}^d \times ]0,\infty[)$ and $L^q(\mathbb{R}^d \times ]0,\infty[)$. Again, apply Lemma 4.23 to deduce

$$\left| \langle \partial_i u_p(\cdot,t), v_p(\cdot,t) \rangle \right| \leq C \|u\|_{W^{s,p}(\mathbb{R}^d)} \|v\|_{W^{1-s,q}(\mathbb{R}^d)}, \quad \forall t > 0,$$

and some suitable constant $C$ independent of $u$, $v$ and $t > 0$. Hence, the desired estimate follows as $t \to 0$. \hfill \Box

The following estimate is related to Lemma 4.23

**Proposition 4.25 (Hardy’s inequality).** If $f$ is a nonnegative Lebesgue integrable function on $]0,\infty[$ and for any $\gamma \neq 0$ then

$$\int_0^\infty \left[ t^\gamma \int_t^\infty f(\tau)d\tau \right] \frac{dt}{t} \leq \gamma^{-p} \int_0^\infty \left[ t^{\gamma+1} f(t) \right] \frac{dt}{t}, \quad \gamma > 0,$$

$$\int_0^\infty \left[ t^\gamma \int_0^t f(\tau)d\tau \right] \frac{dt}{t} \leq (-\gamma)^{-p} \int_0^\infty \left[ t^{\gamma+1} f(t) \right] \frac{dt}{t}, \quad \gamma < 0,$$

for any $1 \leq p < \infty$.

**Proof.** First, it is convenient to write

$$F_\gamma(x) = 1_{\gamma > 0} \int_x^\infty f(y)dy + 1_{\gamma < 0} \int_0^x f(y)dy, \quad \forall x > 0, \ \gamma \neq 0,$$

so that Hardy’s inequality takes the form

$$\int_0^\infty [x^{\gamma} F_\gamma(x)] \frac{dx}{x} \leq |\gamma|^{-p} \int_0^\infty [x^{\gamma+1} f(x)] \frac{dx}{x}.$$

Hence, a change of variable yields

$$\int_0^x f(y)dy = x \int_0^1 f(xy)dy$$

$$\int_x^\infty f(y)dy = x \int_1^\infty f(xy)dy = x \int_0^1 f(x/y)y^{-2}dy,$$

and therefore

if $\gamma < 0$ then $x^{\gamma-1/p} F_\gamma(x) = \int_0^1 x^{\gamma+1-1/p} f(xy)dy$

if $\gamma > 0$ then $x^{\gamma-1/p} F_\gamma(x) = \int_0^1 x^{\gamma+1-1/p} f(x/y)y^{-2}dy$. 
Integrate these expressions and Minkowski inequality for integrals (see Remark B.59) to deduce

if $\gamma < 0$ then
$$\| (\cdot)^{\gamma-1/p} F_\gamma \|_p = \int_0^1 \| (\cdot)^{\gamma+1-1/p} f(\cdot) \|_p dy$$

if $\gamma > 0$ then
$$\| (\cdot)^{\gamma-1/p} F_\gamma \|_p = \int_0^1 \| (\cdot)^{\gamma+1-1/p} f(\cdot/y) y^{-2} \|_p dy,$$

where $\| \cdot \|_p$ means the $L^p([0, \infty[)$-norm. Next, use the expressions

$$\left( \int_0^\infty [x^{\gamma+1-1/p} f(xy)]^p dx \right)^{1/p} = y^{-\gamma-1} \left( \int_0^\infty [x^{\gamma+1} f(x)]^p \frac{dx}{x} \right)^{1/p},$$

$$\left( \int_0^\infty [x^{\gamma+1-1/p} f(x/y) y^{-2}]^p dx \right)^{1/p} = y^{\gamma-1} \left( \int_0^\infty [x^{\gamma+1} f(x)]^p \frac{dx}{x} \right)^{1/p}$$

to conclude. \qed

For instance, if $L^p_\alpha(\mathbb{R}^+) \equiv \{ f \mid \| f \|_{L^p_\alpha(\mathbb{R}^+)} = \int_0^\infty |f(t)| t^\alpha |p| dt < \infty \}$ such that

$$\| f \|_{L^p_\alpha} = \int_0^\infty |f(t)| t^\alpha |p| dt < \infty,$$

then

$$f \mapsto \frac{1}{t} \int_0^t f(x) \, dx$$

defines a linear continuous operator in $L^p_\alpha(\mathbb{R}^+)$ if $\alpha + 1/p < 1$, while the

$$f \mapsto \frac{1}{t} \int_t^\infty f(x) \, dx = \frac{1}{t} \int_0^{1/t} f(y) y^{-2} \, dy$$

is a linear continuous operator in $L^p_\alpha(\mathbb{R}^+)$ if $\alpha + 1/p > 1$, and in both cases, the norm of the operator is bounded by $1/|\alpha + 1/p - 1|$.

Write $x = (x', x_d)$ with $x_d > 0$ when $x$ belongs to $\mathbb{R}^n_+$. Thus, another consequence of Hardy’s inequality is the following

**Proposition 4.26.** If $s - 1/p$ is not an integer and $f$ belongs to $W_0^{s,p}(\mathbb{R}^d_+)$ then the function $(x', x_d) \mapsto x_d^{-s+|\alpha|} \partial^\alpha f(x', x_d)$ belongs to $L^p(\mathbb{R}^d_+)$ for every multi-index $\alpha$ with $|\alpha| \leq s$.

**Proof.** Certainly, only the case $|\alpha| = 0$ and $d = 1$ should be considered. Thus if $s = m$ is a positive integer and $f$ belongs to $\mathcal{D}(\mathbb{R}^+)$ then

$$f(x) = \int_0^x \frac{(x-y)^{m-1}}{(m-1)!} f^{(m)}(y) \, dy,$$
which yields
\[ \frac{|f(x)|}{x^m} \leq \frac{1}{(m-1)!} \frac{1}{x} \int_0^x |f^{(m)}(y)| \, dy. \]

Hence, Hardy’s inequality implies
\[ \|(-m f)(\cdot)\|_{L^p(\mathbb{R}^+)} \leq \frac{p}{(m-1)!} \frac{p}{(p-1)} \|f^{(m)}\|_{L^p(\mathbb{R}^+)} \]
and by density, this holds for any \( f \) in \( W^{m,p}_0(\mathbb{R}^+) \).

Now, if \( s = m + \sigma \) with \( m \) integer and \( 0 < \sigma < 1 \) then consider the \( m \) derivative \( g = f^{(m)} \) and make use of the identity
\[
g(x) = -h(x) + \int_x^\infty h(y) \frac{dy}{y}, \quad \text{with} \quad h(x) = \frac{1}{x} \int_0^x [g(y) - g(x)] \, dy.
\]

To check that \( x \mapsto x^{-\sigma} h(x) \) belongs to \( L^p(\mathbb{R}^+) \), use Hölder inequality to get
\[
\int_0^\infty x^{-p\sigma} \left| \frac{1}{x} \int_0^x [g(y) - g(x)] \, dy \right|^{1/p} \, dx \leq \int_0^\infty x^{-p\sigma-1} \, dx \int_0^x |g(y) - g(x)|^p \, dy \leq \int_0^\infty \int_0^x \frac{|g(y) - g(x)|^p}{|y-x|^{1+p\sigma}} \, dx \, dy,
\]
which finite for \( g \) in \( W^{\sigma,p}(\mathbb{R}^+) \). Next, Hardy’s inequality shows that
\[
x \mapsto x^{-\sigma} \int_x^\infty h(y) \frac{dy}{y}
\]
belongs to \( L^p(\mathbb{R}^+) \) when \( \sigma < 1/p \). On the other hand, if \( \sigma > 1/p \) then make use of the identity
\[
g(x) = -h(x) + \int_0^x h(y) \frac{dy}{y}, \quad \text{with} \quad h(x) = \frac{1}{x} \int_0^x [g(y) - g(x)] \, dy.
\]
to conclude, in a similar way, that \( x \mapsto x^{-\sigma} f^{(m)}(x) \) belongs to \( L^p(\mathbb{R}^+) \). Finally, repeat the argument for the case of integer order to deduce that the function \( x \mapsto x^{-\sigma-(k)} f^{(m-k)}(x) \) is in \( L^p(\mathbb{R}^+) \), \( k = 1, 2, \ldots, m \).

Note that the weight \( x^{-s^{+}|\alpha|} \) represent the distance to the boundary. Thus, if \( \Omega \) is a bounded Lipschitz domain (see Definition 4.4) and \( x \mapsto \rho(x) \) is the distance to the boundary \( \partial \Omega \) then Proposition 4.26 shows that \( x \mapsto \rho^{-s^{+}|\alpha|}(x) f(x) \) belongs to \( L^p(\Omega) \).

It is clear that if \( s = m + s' \) with \( m \) integer and \( 0 < s' < 1 \) then
\[
f \mapsto \|f\|_{s,p} = \left( \sum_{|\alpha| \leq m} \|\partial^\alpha f\|_p^p + \sum_{|\alpha| = m} |\partial^\alpha f|_{s',p}^p \right)^{1/p}
\]
is a norm on $W^{s,p}(\Omega)$.

It should be clear by now that all results concerning *Density and Extension* and *Imbedding and Compactness* of the previous Sections 4.1 and 4.2, can be extended to fractional order Sobolev spaces. For instance, $W^{s,p}(\mathbb{R}^d) \subset \subset W^{t,q}(\mathbb{R}^d)$, for $t \leq s$ (possible negative) and $q \geq p \geq 1$ such that $s - d/p = t - d/q$, and $W^{s,p}(\mathbb{R}^d) \subset C^{k,\alpha}(\mathbb{R}^d)$, for $k < s - d/p < k + 1$, $\alpha = s - k - d/p$, with $k$ a nonnegative integer, e.g., see Grisvard [60, Chapter 1, pp. 1–80] and references therein.

Many other related subjects needs to be discussed, e.g., interpolation techniques, more details on the traces and the Sobolev spaces on the boundary, and several fine estimates, but those are left for a more advanced course. For instance, the whole book by Adams [2] is dedicated to Sobolev spaces and certainly, detailed proofs of quoted assertions (and many more results) can be found there. Other references, such as DiBenedetto [31], Leoni [79] and Ziemer [140], have a clean approach. Also, the recent English translation of Nečas [93] is a classic source.

Still to cover are the Sobolev spaces adapted to parabolic equations, i.e., on a domain of the form $\Omega \times [0, T]$, where a particular variable $t$ is distinguished, e.g., see the books Ladyzhenskaya et al. [77], Lieberman [81], and Wu et al. [134]. In later sections, the Fourier transform is used to given another characterization of Sobolev spaces.
Chapter 5

Basic Fourier Transform

The Fourier transform can be initially defined in various function spaces, perhaps the most natural we are $S(\mathbb{R}^d)$, the space of rapidly decreasing smooth functions. In its definition, the constant $\pi$ can be placed conveniently, for instance, in harmonic analysis

$$(\mathfrak{F} f)(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i x \cdot \xi} \, dx, \quad \forall \xi \in \mathbb{R}^d,$$  \hspace{1cm} (5.1)$$

where $x \cdot \xi$ is the Euclidean scalar product in $\mathbb{R}^d$, or

$$(F f)(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-i x \cdot \xi} \, dx,$$

is used, while

$$(F f)(\xi) = \int_{\mathbb{R}^d} f(x) e^{i x \cdot \xi} \, dx,$$

is used in probability (so-called characteristic function), in any case, the constant $\pi$ plays an important role in the inversion formula. In this section, we retain the expression (5.1), as well as the simplified notation $\mathfrak{F} f = \hat{f}$. For instance, the textbook by Stein and Shakarchi [114] is an introduction to this topic.

Essentially, by completing the square, the following one-dimensional calculation

$$\int_{\mathbb{R}} e^{-\pi \lambda x^2 - 2\pi i x \cdot \xi} \, dx = e^{-\pi \xi^2 / \lambda} \int_{\mathbb{R}} e^{-\pi (x \sqrt{\lambda} + i \xi / \sqrt{\lambda})^2} \, dx,$$

$$\partial \xi \int_{\mathbb{R}} e^{-\pi (x \sqrt{\lambda} + i \xi / \sqrt{\lambda})^2} \, dx = (i / \lambda) \int_{\mathbb{R}} \partial_x e^{-\pi (x \sqrt{\lambda} + i \xi / \sqrt{\lambda})^2} \, dx = 0,$$

$$\int_{\mathbb{R}} e^{-\pi \lambda x^2 / 2} \, dx = (1 / \sqrt{\lambda}) \int_{\mathbb{R}} e^{-\pi x^2 / 2} \, dx = 1 / \sqrt{\lambda},$$
shows that
\[
\int_{\mathbb{R}} e^{-\pi \lambda x^2} e^{-2\pi i x \cdot \xi} \, dx = (1/\sqrt{\lambda}) e^{-\pi \xi^2 / \lambda}.
\]
Using the product form the exponential (and a rotation in the integration variable), this yields
\[
\int_{\mathbb{R}^d} e^{-x \cdot ax + 2\pi i x \cdot \xi} \, dx = \frac{\pi^{d/2}}{\sqrt{\det(a)}} e^{-\xi \cdot a^{-1} \xi}, \quad \forall \xi \in \mathbb{R}^d,
\]
for any (complex) symmetric matrix \( a = (a_{ij}) \) whose real part is positive definite, i.e., \( \Re \{ x \cdot ax \} > 0 \), for every \( x \) in \( \mathbb{R}^d \). Therefore, in particular,
\[
(e^{-\pi |x|^2})(\xi) = e^{-\pi |\xi|^2}, \quad \forall \xi \in \mathbb{R}^d,
\]
i.e., the function \( x \mapsto e^{-\pi |x|^2 / 2} \) is a fixed point for the Fourier transform, which is a key fact used below.

For instance, an introduction at the beginning of the graduate level can be found in the book Pinsky [101], among others.

### 5.1 Smooth Functions

Recall the space \( S(\mathbb{R}^d) \) of rapidly decreasing smooth functions, with its Fréchet topology given by the double sequence of seminorms \( \{ p_{n,k} : n, k \geq 0 \} \) given by (3.10).

**Proposition 5.1.** The Fourier transform \( \mathfrak{F} \) defined by (5.1) is a continuous linear bijective application from \( S(\mathbb{R}^d) \) onto itself. The expression
\[
(\mathfrak{F}^{-1} \varphi)(x) = \int_{\mathbb{R}^d} \varphi(\xi) e^{2\pi i x \cdot \xi} \, d\xi, \quad \forall x \in \mathbb{R}^d.
\]
defines its inverse, which is also continuous.

**Proof.** The linearity comes from the definition, and an integration by parts shows that
\[
(2\pi i)^{|\beta|} |\xi|^{\beta} \partial_\xi^\alpha \varphi(\xi) = (-2\pi i)^{|\alpha|} |\partial_\xi^\beta (x^\alpha \varphi(x))|, \quad \forall \varphi \in S(\mathbb{R}^d),
\]
for every multi-indices \( \alpha = (\alpha_1, \ldots, \alpha_d) \) and \( \beta \), where \( |\alpha| = \alpha_1 + \cdots + \alpha_d \) and \( x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d} \).

Since a sequence \( \varphi_n \to 0 \) in \( S(\mathbb{R}^d) \) if and only if \( x^\alpha \partial^\beta \varphi(x) \to 0 \) uniformly in \( \mathbb{R}^d \) for any multi-indices \( \alpha \) and \( \beta \), or equivalently (by Leibniz’ formula), if and only if \( \partial^\beta (x^\alpha \varphi(x)) \to 0 \) uniformly in \( \mathbb{R}^d \) for any multi-indices \( \alpha \) and \( \beta \), the multiplication by a polynomial is a continuous operation in \( S(\mathbb{R}^d) \). This fact and the previous identity prove that \( \mathfrak{F} \) is a continuous linear mapping from \( S(\mathbb{R}^d) \) into itself.
A simple change of variables and an exchange of the order of integration yields the identity
\[
\int_{\mathbb{R}^d} \psi(\xi) \hat{\varphi}(\xi) e^{2\pi i x \cdot \xi} d\xi = \int_{\mathbb{R}^d} \hat{\psi}(y) \varphi(x + y) dy, \quad \forall \psi, \varphi \in S(\mathbb{R}^d).
\]
Replacing \(\psi(\xi)\) with \(\psi_\epsilon(\xi) = \psi(\epsilon \xi)\) and noting that \(\hat{\psi}(\epsilon \xi)(y) = \epsilon^{-d} \hat{\psi}(y/\epsilon)\), we deduce
\[
\int_{\mathbb{R}^d} \psi(\epsilon \xi) \hat{\varphi}(\xi) e^{2\pi i x \cdot \xi} d\xi = \int_{\mathbb{R}^d} \hat{\psi}(y) \varphi(x + \epsilon y) dy, \quad \forall \psi, \varphi \in S(\mathbb{R}^d),
\]
after a change of variables.
Hence, as \(\epsilon \to 0\) we obtain
\[
\psi(0) \int_{\mathbb{R}^d} \hat{\varphi}(\xi) e^{2\pi i x \cdot \xi} d\xi = \varphi(x) \int_{\mathbb{R}^d} \hat{\psi}(y) dy, \quad \forall \psi, \varphi \in S(\mathbb{R}^d), \forall x \in \mathbb{R}^d.
\]
In particular, if \(\psi(\xi) = e^{-\pi |\xi|^2}\) then \((\mathcal{F}^{-1} \varphi)(x) = \varphi(x)\).

Remarking that \(\overline{\mathcal{F}\varphi} = \mathcal{F}^{-1} \overline{\varphi}\) where the bar \(\cdot\) means complex conjugate, and denoting by \((\cdot, \cdot)\) the inner product in the complex-valued square integrable functions \(L^2(\mathbb{R}^d)\), i.e.,
\[
(f, g) = \int_{\mathbb{R}^d} f(x) \overline{g(x)} dx, \quad \forall f, g \in L^2(\mathbb{R}^d), \quad (5.4)
\]
we obtain

**Corollary 5.2** (Parseval-Plancherel). *If \(\varphi\) and \(\psi\) are two elements belonging to \(S(\mathbb{R}^d)\) then \((\varphi, \psi) = (\hat{\varphi}, \hat{\psi})\). Moreover, we have the convolution property \(\mathcal{F}(\varphi \ast \psi) = \hat{\varphi} \hat{\psi}\).*

**Proof.** A particular case, \(x = 0\), in the proof of Proposition 5.1 yields
\[
\int_{\mathbb{R}^d} \psi(\xi) \hat{\varphi}(\xi) d\xi = \int_{\mathbb{R}^d} \hat{\psi}(y) \varphi(y) dy, \quad \forall \psi, \varphi \in S(\mathbb{R}^d).
\]
Replacing \(\psi\) with its complex conjugate \(\overline{\psi}\) and noting that \(\mathcal{F}(\overline{\varphi}) = \overline{\mathcal{F}^{-1} \varphi}\), we obtain \((\mathcal{F}\varphi, \phi) = (\varphi, \mathcal{F}^{-1} \phi)\). Now, take \(\psi = \mathcal{F}^{-1} \phi\) to get \(\mathcal{F}\psi = \phi\) and \((\mathcal{F}\varphi, \mathcal{F}\psi) = (\varphi, \psi)\), i.e., Parseval-Plancherel’s equality.

Again, by exchanging the order of the integrals and by means of a change of variables, we deduce
\[
\mathcal{F}(\varphi \ast \psi)(\xi) = \int_{\mathbb{R}^d} (\varphi \ast \psi)(x) e^{-2\pi i x \cdot \xi} dx = \int_{\mathbb{R}^d} \psi(y) e^{-2\pi i y \cdot \xi} \left( \int_{\mathbb{R}^d} \varphi(x - y) e^{-2\pi i (x - y) \cdot \xi} dx \right) dy = \hat{\psi}(\xi) \hat{\varphi}(\xi),
\]
for every \(\xi\) in \(\mathbb{R}^d\), i.e., the convolution property. \(\square\)
Clearly, this is to say that the Fourier transform can be considered as an isometry from the (complex) $L^2(\mathbb{R}^d)$ onto itself, i.e., $\mathfrak{F}$ preserves the scalar product in $L^2$.

- **Remark 5.3.** The arguments in Corollary 5.2 also show that $\mathfrak{F}^{-1}$ is the (complex) adjoint operator of $\mathfrak{F}$ and that

$$\int_{\mathbb{R}^d} \hat{\varphi}(x) \psi(x) \, dx = \int_{\mathbb{R}^d} \varphi(y) \hat{\psi}(y) \, dy, \quad \forall \psi, \varphi \in \mathcal{S}(\mathbb{R}^d),$$

and by density, for any $\varphi$ and $\psi$ in $L^2(\mathbb{R}^d)$. Since $\mathfrak{F}^{-1}(\varphi) = \mathfrak{F}(\hat{\varphi})$, where $\hat{\varphi}(x) = \varphi(-x)$, we deduce $\mathfrak{F}^{-1}(\overline{\varphi \psi}) = \overline{\varphi \hat{\psi}}$, which yields the product-convolution formula $\overline{\varphi \psi} = \varphi \hat{\psi}$. Also, this shows that the Fourier transform has period 4, i.e., $\mathfrak{F}^4$ is the identity. Certainly, we also have the following relation with the partial derivatives, $(\partial^\alpha \hat{\varphi})(\xi) = \left( (-2\pi i x)^\alpha \varphi(x) \right)(\xi)$ and $(\partial^\alpha \varphi)(\xi) = (2\pi i \xi)^\alpha \hat{\varphi}(\xi)$, for every $\varphi$ in $\mathcal{S}(\mathbb{R}^d)$.

Before going further, besides the key Parseval-Plancherel’s equality $||\hat{\varphi}||_{L^2} = ||\varphi||_{L^2}$ and

$$\int_{\mathbb{R}^d} |\nabla \varphi(x)|^2 \, dx = 4\pi^2 \int_{\mathbb{R}^d} |\xi|^2 |\hat{\varphi}(\xi)|^2 \, d\xi, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d),$$

let us mention some useful properties valid for any functions in $\mathcal{S}(\mathbb{R}^d)$:

1. Convolution and multiplication: $\mathfrak{F}[\varphi \ast \psi] = \mathfrak{F}[\varphi] \mathfrak{F}[\psi]$ and $\mathfrak{F}[\varphi \psi] = \mathfrak{F}[\varphi] \ast \mathfrak{F}[\psi]$.

2. Translation: $\mathfrak{F}[\varphi(x - h)](\xi) = e^{-2\pi i h \cdot \xi} \mathfrak{F}[\varphi](\xi)$ and $\mathfrak{F}[e^{-2\pi i h \cdot x} \varphi(x)](\xi) = \mathfrak{F}[\varphi](\xi - h)$.

3. Differentiation: If $\alpha$ is a multi-index of order $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and $(2\pi i \xi)^\alpha = (2\pi i)^{\alpha_1} \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}$ then $\mathfrak{F}[\partial^\alpha \varphi](\xi) = (2\pi i \xi)^\alpha \mathfrak{F}[\varphi](\xi)$ and $\partial^\alpha \mathfrak{F}[\varphi](\xi) = \mathfrak{F}[(-2\pi i x)^\alpha \varphi(x)](\xi)$.

4. Transformations: Scaling, if $\varphi_\varepsilon(x) = \varepsilon^{-d} \varphi(x/\varepsilon)$ then $\mathfrak{F}[\varphi_\varepsilon](\xi) = \mathfrak{F}[\varphi](\varepsilon \xi)$ and $\mathfrak{F}[\varphi(\varepsilon x)](\xi) = \varepsilon^{-d} \mathfrak{F}[\varphi](\xi/\varepsilon)$, as well as

$$\mathfrak{F}[\varphi](0) = \int_{\mathbb{R}^d} \varphi(x) \, dx \quad \text{and} \quad \varphi(0) = \int_{\mathbb{R}^d} \mathfrak{F}[\varphi](\xi) \, d\xi.$$

Moreover, if $\mathfrak{r}$ is an orthogonal real matrix (i.e., either the transposed matrix give the inverse or $|\mathfrak{r}x| = |x|$ for every $x$ in $\mathbb{R}^d$) then $\mathfrak{F}[\varphi(\mathfrak{r}x)](\xi) = \mathfrak{F}[\varphi](\mathfrak{r} \xi)$.

5. Conjugation and period: If the bar is the conjugate of a complex number then $\mathfrak{F}[\overline{\varphi}](\xi) = \overline{\mathfrak{F}[\varphi](\xi)}$. Moreover, if the tilde is parity operator, i.e., $\mathfrak{P}[\varphi](x) = \tilde{\varphi}(x) = \varphi(-x)$, then $\mathfrak{F}^2 = \mathfrak{P}$, $\mathfrak{F}^3 = \mathfrak{F}^{-1} = \mathfrak{F} = \mathfrak{F}^*$ and $\mathfrak{F}^4 = I$, where $I$ is the identity operator and $\mathfrak{F}^*$ is the adjoint operator of $\mathfrak{F}$.

To end this subsection, let us consider Fourier image of the test functions $\mathcal{D}(\mathbb{R}^d)$. If $K$ is a compact set in $\mathbb{R}^d$ then the indicator function $I_K$ is defined as the function $\eta \mapsto \sup\{x \cdot \eta : x \in K\}$. Hahn-Banach extension theorem
shows that $I_K$ determines the convex envelop $\text{co}(K)$ of $K$, and that $I_K = I_{\text{co}(K)}$. Without going into details, an entire function in $d$-complex variables is an analytic functions in the whole $\mathbb{C}^d$ space, i.e., it can be expressed locally as a complex series.

**Theorem 5.4 (Paley-Wiener).** If $K$ is a convex compact set in $\mathbb{R}^d$ then an entire function $\Phi$ in the $d$-complex variables is the Fourier transform of a test function in $\mathcal{D}(\mathbb{R}^d)$ with support in $K$ if and only if for every $n \geq 0$ there exists a constant $C_n$ such that $|\Phi(\xi + i\eta)| \leq C_n(1 + |\xi|^2 + |\eta|^2)^{-n}e^{2\pi I_K(\eta)}$, for any $\xi + i\eta$ in the complex space $\mathbb{C}^d$.

**Proof.** To check the necessity of the estimate, use the equality
\[
(2\pi i\xi)^\alpha \hat{\varphi}(\xi) = \int_{\mathbb{R}^d} e^{2\pi i x \cdot \xi} \partial^\alpha \varphi(x) dx,
\]
for any multi-index $\alpha$, and replace $\xi$ by $\xi + i\eta$ to obtain
\[
(|\xi|^2 + |\eta|^2)^{\alpha/2} |\hat{\varphi}(\xi)| \leq C_\alpha e^{2\pi I_K(\eta)} \sup_{x \in K} |\partial^\alpha \varphi(x)|,
\]
which implies the desired estimate for $\Phi = \hat{\varphi}$.

Conversely, if $\Phi$ is an entire function satisfying the 'growth' estimate, then the function $\xi \mapsto \Phi(\xi + 0i)$ of the real variable $\xi$ is rapidly decreasing and smooth, i.e., $\Phi|_{\mathbb{R}^d}$ belongs to $\mathcal{S}(\mathbb{R}^d)$, an thus, $\Phi = \hat{\varphi}$ for some function $\varphi$ in $\mathcal{D}$. Therefore, choose any point $y$ in $\mathbb{R}^d \setminus K$ and prove that $\varphi(y) = 0$.

To this end, use Hahn-Banach theorem to find an hyperplane separating $y$ from $K$, i.e., there exists $\eta_0$ in $\mathbb{R}^d$ with $|\eta_0| = 1$ and $c$ in $\mathbb{R}$ such that $K \subset \{x \in \mathbb{R}^d : x \cdot \eta_0 < c\}$ and $y \cdot \eta_0 > c$. Moreover, after an orthogonal change of coordinates, $\eta_0 = (0,0,\ldots,0,1)$, and $y_d > I_K(\eta_0)$.

For $\xi = (\xi', \xi_d)$ and $t > 0$, use Cauchy’s theorem to write the Fourier inversion formula for $\varphi$ as
\[
\varphi(y) = \int_{\mathbb{R}^{d-1}} e^{2\pi iy' \cdot \xi'} d\xi' \int_{\mathbb{R}} e^{2\pi iy_d(\xi_d + it)} \Phi(\xi', \xi_d + it) d\xi_d,
\]
since $|\Phi(\xi', \xi_d + i\eta_d)| \to 0$ as $|\xi| \to \infty$ and $|\eta_d| \leq 1$. Combine this with the growth estimate (with $n=d+1$) to deduce the upper bound
\[
|\varphi(y)| \leq C \int_{\mathbb{R}^d} e^{-t(y_d - I_K(\eta_0))} (1 + |\xi|^2)^{-n-1} d\xi,
\]
for a suitable constant $C$. Hence, the condition $y_d > I_K(\eta_0)$ implies that the right-hand side vanishes as $t \to \infty$, i.e., $\varphi(y) = 0$. \qed

## 5.2 Tempered Distributions

Before considering tempered distribution, recall the slowly growth (at infinity) $L^1_{\text{loc}}$ functions, i.e., functions $f$ such that $f(x)$ is bounded by a polynomial outside of a compact region. These class of functions can be regarded as tempered...
distribution by the inclusion \( f \mapsto T_f \),
\[
    \langle T_f, \varphi \rangle = \int_{\mathbb{R}^d} f(x) \varphi(x) \, dx, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).
\]

Now, as in Parseval-Plancherel’s equality, essentially by exchanging the order of integrations, we have the identity
\[
    \int_{\mathbb{R}^d} \psi(x) \hat{\varphi}(x) \, dx = \int_{\mathbb{R}^d} \psi(x) \, dx \int_{\mathbb{R}^d} \varphi(y) e^{2\pi i x \cdot y} \, dy = \int_{\mathbb{R}^d} \varphi(y) \, dy \int_{\mathbb{R}^d} \psi(x) e^{2\pi i x \cdot y} \, dx = \int_{\mathbb{R}^d} \hat{\psi}(y) \varphi(y) \, dy,
\]
for every \( \psi, \varphi \) in \( \mathcal{S}(\mathbb{R}^d) \), which can be extended to holds for a pair of functions not necessarily rapidly decreasing and smooth. Indeed, the integral on the right makes sense for every \( \varphi \) in \( \mathcal{S}(\mathbb{R}^d) \) if \( \psi = f \) is a slowly increasing \( L^1_{\text{loc}} \) function. In particular, if \( f \) is integrable in \( \mathbb{R}^d \), the formula (5.1) also defines a function \( \hat{f} \) and in this case, \( \hat{T_f} = T_{\hat{f}} \).

In view of the above, the Fourier transform \( \mathcal{F} \) of a tempered distribution, i.e., an element \( T \) in \( \mathcal{S}'(\mathbb{R}^d) \), is defined by duality, namely,
\[
    \langle \hat{T}, \varphi \rangle = \langle T, \hat{\varphi} \rangle, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d). \tag{5.5}
\]

Thus, in particular this holds for any slowly increasing \( L^1_{\text{loc}} \) function.

The identity (5.3) remains true for tempered distribution, and \( \mathcal{F} \) becomes a bijection between \( \mathcal{S}'(\mathbb{R}^d) \) and itself, where its inverse is given by
\[
    \langle \mathcal{F}^{-1} T, \varphi \rangle = \langle T, \mathcal{F}^{-1} \varphi \rangle, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d).
\]

Moreover, if the convolution of two distributions \( T \) and \( S \) is defined (e.g., one of them has compact support) then \( \hat{T} \ast \hat{S} = \mathcal{F} T \ast S \) holds.

Any finite Radon measure \( \mu \) on \( \mathbb{R}^d \) can be regarded as a tempered distribution and its Fourier transform \( \hat{\mu} \) is actually given by the expression
\[
    \hat{\mu}(\xi) = \int_{\mathbb{R}^d} e^{-2\pi i x \cdot \xi} \mu(dx), \quad \forall \xi \in \mathbb{R}^d
\]
and we have

**Theorem 5.5 (Bochner).** If \( \Psi : \mathbb{R}^d \to \mathbb{C} \) is the characteristic function of a finite Radon measure \( \mu \) on \( \mathbb{R}^d \), i.e., \( \Psi(\xi) = \hat{\mu}(\xi) \), then (a) \( \Psi \) is continuous and (b) \( \Psi \) is positive definite, i.e., for every natural number \( k \), any \( \zeta_i \) in \( \mathbb{R}^d \) and any complex number \( z_i, i = 1, \ldots, k \) we have
\[
    \sum_{i,j=1}^k \Psi(\zeta_i - \zeta_j) z_i \overline{z_j} \geq 0,
\]
where \( \overline{z} \) is the conjugate of a complex number. Conversely, an arbitrary function \( \Psi : \mathbb{R}^d \to \mathbb{C} \) satisfying the above properties (a) and (b) is the characteristic function of a finite Radon measure \( \mu \) on \( \mathbb{R}^d \).
Proof. The continuity follows from the dominated convergence theorem, and to show (b), we remark that
\[
\sum_{i,j=1}^{k} \hat{\mu} (\zeta_i - \zeta_j) z_i \bar{z}_j = \int_{\mathbb{R}^d} \left| \sum_{i=1}^{k} z_i e^{-2\pi i \zeta_i} \right|^2 \mu(dx) \geq 0,
\]
for every \( z_i \) and \( \zeta_i \).

If \( \Psi \) is a positive definite function then by taking \( k = 2, \zeta_1 = 0 \) and \( \zeta_2 = \xi \) we deduce that the matrix
\[
\begin{bmatrix}
\Psi(0) & \Psi(\xi) \\
\Psi(-\xi) & \Psi(0)
\end{bmatrix}
\]
is semi-positive definite, which yields
\[
\Psi(\xi) = \Psi(-\xi), \quad \Psi(\xi) \geq 0, \quad |\Psi(x)| \leq \Psi(0), \quad \forall \xi \in \mathbb{R}^d,
\]
i.e., \( \Psi \) is a bounded function.

To prove the converse, let \( \Psi \) a continuous positive definite function. Because \( \Psi \) is necessarily bounded, it can be considered as a tempered distribution and since the Fourier transform is invertible on \( S'(\mathbb{R}^d) \), there exists a tempered distribution \( T \) such that \( \Psi = \hat{T} \). Now to show that \( T \) is indeed a Radon measure, it suffices to check that \( T \) is a positive distribution, see Proposition 3.20.

To this purpose, for any \( \phi \) in \( S(\mathbb{R}^d) \) denote \( \tilde{\phi}(x) = \varphi(-x) \), suppose \( \varphi \) real-valued and compute \( \langle T, \varphi^2 \rangle \) using Parseval-Plancherel’s equality to get
\[
\langle T, \varphi^2 \rangle = \langle \hat{T}^{-1} \Psi, \varphi^2 \rangle = \langle \Psi, \hat{T}^{-1}(\varphi^2) \rangle = \langle \Psi, (\hat{T}^{-1} \varphi) \ast (\hat{T}^{-1} \varphi) \rangle = \langle \Psi, \tilde{\phi} \ast \phi \rangle,
\]
where \( \phi = \hat{T} \varphi \). Now, since \( \Psi \) is positive definite, for \( \phi \) in \( D(\mathbb{R}^d) \) we have
\[
\int_{\mathbb{R}^d} \Psi(x) (\phi \ast \tilde{\phi})(x) \, dx = \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} \Psi(x + y) \tilde{\phi}(x) \phi(y) \, dy =
\]
\[
= \int_{\mathbb{R}^d} dx \int_{\mathbb{R}^d} \Psi(y - x) \phi(x) \phi(y) \, dy =
\]
\[
= \lim_{\varepsilon \to 0} \varepsilon^2 \sum_{n,m} \Psi(m \varepsilon - n \varepsilon) \tilde{\phi}(n \varepsilon) \phi(m \varepsilon) \geq 0,
\]
where the sum is finite (because \( \phi \) has a compact support) and extended to all \( d \)-dimensional integers \( n, m \). These calculations and the density of \( D(\mathbb{R}^d) \) in \( S(\mathbb{R}^d) \) yield
\[
\langle T, \varphi^2 \rangle = \langle \Psi, \phi \ast \tilde{\phi} \rangle \geq 0, \quad \forall \varphi \in D(\mathbb{R}^d).
\]

Hence, for any nonnegative smooth function \( \psi \) with support in \( K \subset \mathbb{R}^d \), we can find a nonnegative element \( \chi \) in \( D(\mathbb{R}^d) \) such that \( \chi = 1 \) on \( K \). Thus the function
\( \varphi = \sqrt{\chi(\psi + \varepsilon)} \) belongs to \( D(\mathbb{R}^d) \), for any constant \( \varepsilon > 0 \). Since \( \langle T, \varphi ^2 \rangle \geq 0 \), as \( \varepsilon \) vanishes we deduce \( \langle T, \psi \rangle \geq 0 \), i.e., \( T \) is a positive distribution.

It remains to check that \( T = \mu \) is a bounded or finite measure. Indeed, if \( \varphi_n = e^{-\pi |x|^2/n} \) then \( \varphi_n(x) \) increases to 1 as \( n \to \infty \), for every \( x \) in \( \mathbb{R}^d \). Moreover, \( \varphi_n \) belongs to \( \mathcal{S}(\mathbb{R}^d) \) and it Fourier transform satisfies \( \hat{\varphi}_n(\xi) = n^{d/2}e^{-\pi n|x|^2} \), i.e., \( \hat{\varphi}_n(\xi) = n^{d/2}\hat{\varphi}_1(\xi\sqrt{n}) \). Therefore

\[
\mu(\mathbb{R}^d) = \lim_{n} \int_{\mathbb{R}^d} \varphi_n(x) \mu(dx) = \lim_{n} \langle T, \varphi_n \rangle = \lim_{n} \langle \Psi, \hat{\varphi}_n \rangle = \lim_{n} \int_{\mathbb{R}^d} \Psi(\xi/\sqrt{n}) \hat{\varphi}_1(\xi) d\xi = \Psi(0) < \infty,
\]

after a change of variables on the last integral and remarking the \( \hat{\varphi}_1 \) is a kernel.

Recall that since \( \mathcal{S}(\mathbb{R}^d) \) is dense in \( \mathcal{S}'(\mathbb{R}^d) \) with the pairing \( \langle \cdot, \cdot \rangle \), the space of \( L^p(\mathbb{R}^d) \), \( 1 \leq p \leq \infty \) is also continuously embedded with dense image \( \mathcal{S}'(\mathbb{R}^d) \). Therefore, the Fourier transform \( \mathcal{F} \) can be considered as a continuous linear transformation from \( L^p(\mathbb{R}^d) \), \( 1 \leq p \leq \infty \) into \( \mathcal{S}'(\mathbb{R}^d) \). However, we can be more specify about the image \( \mathcal{F}(L^p) \).

\begin{remark}
\textit{Remark 5.6.} The Fourier transform could be considered on the space of entire functions, see Exercise 3.7, and take a look at Friedman [45, Chapter 5].
\end{remark}

\begin{exercise}
Exercise 5.1. First, prove that Fourier transform commute with the tensor product, i.e., if the \( T_1 \) and \( T_2 \) are two tempered distributions then \( T_1 \otimes T_2 = T_1 \ast T_1 \). Secondly, for the convolution of two distributions, prove that if \( T \) belongs to \( \mathcal{S}'(\mathbb{R}^d) \) and \( S \) belongs to \( \mathcal{E}'(\mathbb{R}^d) \) then the convolution \( T \ast S \) belongs to \( \mathcal{S}'(\mathbb{R}^d) \), the Fourier transform \( \hat{T} \ast S \) is identified with a smooth function, namely, \( \xi \mapsto \langle S, e^{-2\pi i \xi} \rangle \), and \( \hat{T} \ast \hat{S} = \hat{T} \hat{S} \).
\end{exercise}

\section{5.3 Integrable Functions}

As mentioned early, the expression (5.1) of the Fourier transform makes also sense for any function in \( L^1(\mathbb{R}^d) \). Recalling the space \( C_*(\mathbb{R}^d) \) of continuous functions vanishing at infinity, i.e., for every \( \varepsilon > 0 \) there exists a compact set \( K = K_\varepsilon \) in \( \mathbb{R}^d \) such that \( |f(x)| \leq \varepsilon \), for every \( x \) in \( \mathbb{R}^d \setminus K \), which is a Banach space with the sup-norm

\[
\| f \|_\infty = \sup \{ |f(x)| : x \in \mathbb{R}^d \}
\]

we have

\begin{theorem}[Riemann-Lebesgue]
\textit{Fourier transform} \( \mathcal{F} \) \textit{satisfies}

\[
\| \hat{f} \|_\infty \leq \| f \|_1, \quad \forall f \in L^1(\mathbb{R}^d),
\]

\textit{and it is a continuous linear mapping from} \( L^1(\mathbb{R}^d) \) \textit{into} \( C_*(\mathbb{R}^d) \).
\end{theorem}
Proof. By means of the equality \(|e^{-2\pi i xy}| = 1\), the \(L^\infty\) bound is immediately obtained. Next, to check that \(\hat{f}\) is continuous and that \(\hat{f}(\xi) \to 0\) as \(|\xi| \to \infty\), we remark that the space \(S(\mathbb{R}^d) \subset C_*(\mathbb{R}^d)\) of rapidly decreasing smooth functions is dense in \(L^1(\mathbb{R}^d)\), and that \(\mathfrak{F}\) maps \(S(\mathbb{R}^d)\) into itself. Therefore, for any given \(f\) in \(L^1(\mathbb{R}^d)\) there exists a sequence \(\{f_k\}\) in \(S(\mathbb{R}^d)\) such that \(f_k \to f\) in \(L^1\). By means of estimate on the sup-norm we have in \(\mathfrak{F}\) is dense in \(S\). We remark that the space \(L^\infty(\mathbb{R}^d)\) is a complete space, its limit \(\hat{f}\) belongs also to \(C_*(\mathbb{R}^d)\).

By considering \(f\) as a tempered distribution and weak derivatives, we can prove several properties valid for any \(f\) in \(L^1(\mathbb{R}^d)\):

1. If for some \(n \geq 0\), \(x \mapsto x^\alpha f(x)\) belongs to \(L^1(\mathbb{R}^d)\), for any \(|\alpha| \leq n\), then \(\partial^\alpha \hat{f}\) belongs to \(C_*(\mathbb{R}^d)\), for any \(|\alpha| \leq n\), and also \((\partial^\alpha \hat{f})(\xi) = ((-2\pi i)^\alpha f(x))(\xi)\).

2. If for some \(n \geq 0\), \(\partial^\alpha f\) belongs to \(L^1(\mathbb{R}^d)\), for any \(|\alpha| \leq n\), then \((\partial^\alpha \hat{f})(\xi) = (2\pi i)^\alpha \hat{f}(\xi)\).

3. If \(g\) belongs to \(L^1(\mathbb{R}^d)\) then \(\hat{f} * g = \hat{f} \hat{g}\) and if \(\hat{g}\) also belongs to \(L^1(\mathbb{R}^d)\) then \(g\) belongs to \(L^\infty(\mathbb{R}^d)\) and \(\hat{f} \hat{g} = \hat{f} * \hat{g}\).

4. If \(\tau_h\) denotes the translation operator, i.e., \(\tau_h f(x) = f(x-h)\), then \(\tau_h \hat{f}(\xi) = e^{-2\pi i h \xi} \hat{f}(\xi)\) and also \([\tau_h(\hat{f})](\xi) = [e^{2\pi i h x} f(x)](\xi)\).

5. If \(\theta\) is a nonzero real number then \([[f(x/\theta)](\xi) = \theta^d \hat{f}(\theta \xi)\).

6. If \(\hat{f}\) belongs to \(L^1(\mathbb{R}^d)\) then the inversion formula for \(\mathfrak{F}\) of Proposition 5.1 is valid.

The Fourier transform can also be considered in \(L^2(\mathbb{R}^d)\).

**Proposition 5.8.** The expressions

\[
(\mathfrak{F} f)(\xi) = \lim_{r \to \infty} \int_{|x| < r} f(x) e^{-2\pi i x \cdot \xi} \, dx, \quad \forall f \in L^2(\mathbb{R}^d),
\]

\[
(\mathfrak{F}^{-1} g)(x) = \lim_{r \to \infty} \int_{|x| < r} g(\xi) e^{2\pi i x \cdot \xi} \, d\xi, \quad \forall g \in L^2(\mathbb{R}^d),
\]

where the limit is understood in the \(L^2\)-sense, defines the Fourier transform \(\mathfrak{F}\) as an unitary operator on the complex \(L^2(\mathbb{R}^d)\).

**Proof.** By means of Parseval-Plancherel’s equality obtained in Corollary 5.2, we see that \(\mathfrak{F}\) and \(\mathfrak{F}^{-1}\) are isometries defined on a dense set, namely \(S(\mathbb{R}^d)\), and therefore, they can be extended to isometries from \(L^2(\mathbb{R}^d)\) into itself, being \(\mathfrak{F}^{-1}\) the inverse of \(\mathfrak{F}\). Now, if \((\cdot,\cdot)\) denotes the inner product in the complex \(L^2(\mathbb{R}^d)\) then from the equality \((\mathfrak{F} f, g) = (f, \mathfrak{F}^{-1} g)\) we deduce \(\mathfrak{F}^* = \mathfrak{F}^{-1}\), i.e., \(\mathfrak{F}\) is an unitary isomorphism on \(L^2(\mathbb{R}^d)\).

To show the validity of the above limiting expressions, for a given \(f\) in \(L^2\), we denote by \(f_r(x) = 1_{|x| < r} f(x)\), which has the property \(f_r \to f\) in \(L^2\) as
By continuity, we have $\hat{f}_r \to \hat{f}$. Since $f_r$ belongs also to $L^1$, we obtain the expression

$$\hat{f}_r(\xi) = \int_{|x| < r} f(x) e^{-2\pi i x \cdot \xi} \, dx.$$ 

Similarly with $g$ and $g_r$, we complete the proof. \hfill $\Box$

If $f$ is only locally integrable with slowly growth (at infinite) function then we are forced to consider $f$ as a tempered distribution, in particular, when $f$ is a periodic function in $\mathbb{R}^d$, i.e., $f(x + P) = f(x)$ for every $x$ in $\mathbb{R}^d$ and some constant (of which the smallest is called period) $P \neq 0$.

At this point we have shown that the Fourier transform $\mathcal{F}$ maps $L^1$ into $L^\infty$ and $L^2$ into itself. Interpolation theorem implies that $\mathcal{F}$ is also a continuous linear mapping from $L^p$ into $L^q$ with $1/p + 1/q = 1$ and $1 < p < 2$. Actually the so called Hausdorff-Young inequality $\|f\|_q \leq \|f\|_p$ holds. Regarding eigenvalues and eigenfunctions, note that the sequence $\{h_n\}$ of Hermite functions (obtained from Hermite polynomials) forms a complete orthogonal system of eigenfunctions for the Fourier transform on $L^2(\mathbb{R}^d)$, i.e., $h_n$ is orthogonal to $h_k$ if $n \neq k$, and $\widehat{h_n} = (-i)^n h_n$. For dimension $d = 1$,

$$h_n(x) = \frac{(-1)^n}{n!} e^{\pi x^2} \frac{d^n e^{-\pi x^2}}{dx^n}, \quad n = 0, 1, \ldots,$$

and

$$e_n = \frac{h_n}{\|h_n\|_{L^2}} = \sqrt{\frac{\sqrt{2(4\pi)}^{-n} n!}{h_n}}$$

is an orthonormal system. For instance, the reader may consult the book Duoandikoeztxea [40] or Grafakos [57, 58] for a comprehensive study on Fourier analysis.

Most of the topics developed in previous chapters can be found in the books Knapp [74, 73], covering much more material with another twist. Perhaps, the reader may want to take a look at Wheeden and Zygmund [133, Chapter 13].

## 5.4 Periodic Functions

A periodic function $f$ in each of the $d$ variables is called a multiply periodic function in $\mathbb{R}^d$, and for simplicity, the period 1, i.e., $f(x + k) = f(x)$ for every $k$ in $\mathbb{Z}^d$, $d$-dimensional integer numbers. It is convenient to consider the semi-open (or semi-closed) unit cube $Q = [-1/2, 1/2]^d$ in $\mathbb{R}^d$ and write $\mathbb{R}^d = \sum_k (Q + k)$, and use the quotient space $\mathbb{R}^d / \mathbb{Z}^d \sim (\mathbb{R}/\mathbb{Z})^d$ of equivalent classes (or cosets of $\mathbb{Z}^d$), which is usually called the $n$-dimensional torus $\mathbb{T}$. Clearly, $\mathbb{R}^d$ and $\mathbb{T}^d$ are Abelian groups under addition, the $d$-dimensional torus $\mathbb{T}^d$ is a compact Hausdorff space, geometrically identified with the 'complex torus' $\{z = (z_1, \ldots, z_d) \in \mathbb{C}^d : |z_1| = 1, \ldots, |z_d| = 1\}$, via the mapping $(x_1, \ldots, x_d) \mapsto (e^{2\pi i x_1}, \ldots, e^{2\pi i x_d})$; and endowed the Lebesgue measure,
by identifying $\mathbb{T}^d$ with the unit cube $Q$, and thus, the Lebesgue spaces $L^p(\mathbb{T}^d)$ are defined, i.e., $p$-integrable complex-valued functions on the cube $Q$, which are extended to $\mathbb{R}^d$ by periodicity and considered as functions defined on $\mathbb{T}^d$, in particular, the Hilbert space $L^2(\mathbb{T}^d)$, with

$$(f,g) = (f,g)_{L^2(\mathbb{T}^d)} = \int_{\mathbb{T}^d} f(x)\overline{g(x)}dx = \int_{Q} f(x)\overline{g(x)}dx,$$

as its inner product. In both spaces, $\mathbb{R}^d$ and $\mathbb{T}^d$, complex-valued functions $e$ with the property that $e(x+y) = e(x)e(y)$ are of special interest, actually, any measurable function $e$ on $\mathbb{R}^d$ (or $\mathbb{T}^d$) satisfying $e(x+y) = e(x)e(y)$ and $|e(x)| = 1$ for every $x, y$ is necessarily of the exponential form, i.e., there exits $\xi$ in $\mathbb{R}^d$ (or $\mathbb{T}^d$) such that $e : x \mapsto e^{2\pi i x \cdot \xi}$.

If $e_k(x) = e^{2\pi i k \cdot x}$ with $x$ in $\mathbb{T}^d$ and $k$ in $\mathbb{Z}^d$ then it is simple to show that the countable family $\{e_k : k \in \mathbb{Z}^d\}$ is an orthonormal basis in $L^2(\mathbb{T}^d)$. Indeed, easy computations check that

$$\int_0^1 e^{2\pi i k t} dt = 1 \text{ if } k = 0 \text{ and } = 0 \text{ otherwise;}$$

while to deduce that the family is complete, Stone-Weierstrass Theorem is invoked, i.e., first the equality $e_{k+n} = e_k e_n$ implies that the subspace of finite linear combinations of the $\{e_k\}$ is an algebra, and it is also clear that separate points on $\mathbb{T}^d$, and $e_0 = 1$, $e_k = e_{-k}$. This can be restated by affirming that if $f$ belongs to $L^2(\mathbb{T}^d)$ then the Fourier transform $\mathfrak{F}(f)$ is defined as the function from $Z$ into $\mathbb{C}$ given by

$$(\mathfrak{F}f)(k) = \hat{f}(k) = \int_{\mathbb{T}^d} f(x)e^{-2\pi i k \cdot x}dx = \int_{Q} f(x)e^{-2\pi i k \cdot x}dx,$$

and the series

$$x \mapsto \sum_{k \in \mathbb{Z}^d} \hat{f}(k)e^{2\pi i k \cdot x}, \text{ with } x \in \mathbb{T}^d \text{ or } x \in Q,$$

is called the Fourier series of $f$. The Fourier transform $\mathfrak{F}$ maps the space of square integrable function on $\mathbb{T}^d$, $L^2(\mathbb{T}^d)$, onto the space of square convergent series, $\ell^2(\mathbb{Z}^d)$, with indexes in $\mathbb{Z}^d$. Moreover, the Fourier series of $f$ converges to $f$ in $L^2$-norm, i.e.,

$$f(x) = \sum_{k \in \mathbb{Z}^d} \hat{f}(k)e^{2\pi i k \cdot x}, \text{ in } L^2(\mathbb{T}), \quad \forall f \in L^2(\mathbb{T}),$$

and Parseval’s identity becomes

$$\|\mathfrak{F}f\|_{\ell^2(\mathbb{Z}^d)} = \|f\|_{L^2(\mathbb{T}^d)}, \quad \forall f \in L^2(\mathbb{T}^d).$$

It is clear that the ‘coefficients’ $\hat{f}(k)$ make sense when $f$ is merely in $L^1(\mathbb{T}^d)$, i.e., the Fourier transform $\mathfrak{F}$ also maps $L^1(\mathbb{T}^d)$ into $\ell^\infty(\mathbb{Z}^d)$ and $\|\hat{f}\|_{\ell^\infty} \leq \|f\|_{L^1}$.
Actually, Hausdorff-Young inequality shows that $\mathbb{F}$ maps $L^p(\mathbb{T}^d)$, $1 \leq p \leq 2$, into $\ell^{p'}(\mathbb{Z}^d)$, $1/p + 1/p' = 1$, and $\|\hat{f}\|_{\ell^{p'}} \leq \|f\|_{L^p}$.

A Fourier series with coefficients $c_k$, i.e., $x \mapsto \sum_{k \in \mathbb{Z}^d} c_k e^{2\pi i k \cdot x}$, is convergent to a function in $L^2(\mathbb{T}^d)$ if the numerical series $\sum_{k \in \mathbb{T}^d} |c_k|^2 < \infty$. Since functions defined on $\mathbb{T}^d$ are periodic, a Fourier series is a good way of representing periodic functions. Another way of producing periodic functions is the expression

$$Pf(x) = \sum_{k \in \mathbb{Z}^d} f(x - k), \quad \text{i.e.,} \quad Pf = \sum_{k \in \mathbb{Z}^d} \tau_k f,$$

(5.6)

where $\tau_k f(x) = f(x - k)$ is the translation operator in either $\mathbb{R}^d$, but if $f$ is defined on $\mathbb{Z}^d$ then $\tau_k f = f$, for every $k$ in $\mathbb{Z}^d$. It is clear that some conditions on $f$ are necessary to ensure some type of convergence for the series $Pf$. Notable, there is a nice relation between these two ways of producing periodic functions.

**Theorem 5.9.** First, if $f$ belongs to $L^1(\mathbb{R}^d)$ then the series $Pf$ converges pointwise almost everywhere and in $L^1(\mathbb{Z}^d)$, i.e., the operator $P$ maps $L^1(\mathbb{R}^d)$ into $L^1(\mathbb{Z}^d)$, and $\|Pf\|_{L^1(\mathbb{Z}^d)} \leq \|f\|_{L^1(\mathbb{R}^d)}$. Moreover, the Fourier transform of $Pf$ on $\mathbb{Z}^d$, i.e.,

$$\hat{Pf}(k) = \int_{\mathbb{T}^d} f(x) e^{-2\pi i k \cdot x} dx, \quad \forall k \in \mathbb{Z}^d,$$

is equal to the Fourier transform $f$ on $\mathbb{R}^d$ restricted to $\mathbb{Z}^d$, i.e.

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \xi \cdot x} dx, \quad \forall \xi = k \in \mathbb{Z}^d.$$

Furthermore, if $f$ belongs to $C(\mathbb{R}^d)$ satisfies: there exist constants $C, \varepsilon > 0$ such that

$$|f(x)| \leq C(1 + |x|)^{-d-\varepsilon} \quad \text{and} \quad |\hat{f}(\xi)| \leq C(1 + |\xi|)^{-d-\varepsilon},$$

for every $x$ and $\xi$ in $\mathbb{R}^d$, then

$$\sum_{k \in \mathbb{Z}^d} f(x + k) = \sum_{k \in \mathbb{Z}^d} \hat{f}(k) e^{2\pi i k \cdot x}, \quad \forall x \in \mathbb{T}^d,$$

(5.7)

where both series converge absolutely and uniformly on $\mathbb{T}^d$, and in particular, for $x = 0$ this yields Poisson summation formula.

**Proof.** The key argument is based on the equality $\mathbb{R}^d = \sum_{k \in \mathbb{Z}^d} (Q + k)$, which implies

$$\int_{\mathbb{R}^d} |f(x)| dx = \sum_{k \in \mathbb{Z}^d} \int_Q |f(x)| dx = \int_Q \sum_{k \in \mathbb{Z}^d} |f(x - k)| dx.$$
This shows that the series $Pf$ converges pointwise almost everywhere and in $L^1(\mathbb{Z}^d)$-norm, and the estimate $\|Pf\|_{L^1(\mathbb{Z}^d)} \leq \|f\|_{L^1(\mathbb{R}^d)}$, with equality when $f \geq 0$. Similarly, in view of the periodicity of $x \mapsto e^{-2\pi ik \cdot x}$,

$$\hat{f}(k) = \int_{\mathbb{R}^d} f(x) e^{-2\pi ik \cdot x} \, dx = \sum_{k \in \mathbb{Z}^d} \int_{Q+k} f(x) e^{-2\pi ik \cdot x} \, dx$$

can be rewritten as

$$= \sum_{k \in \mathbb{Z}^d} \int_{Q+k} f(x) e^{-2\pi ik \cdot (x+k)} \, dx = \sum_{k \in \mathbb{Z}^d} \int_{Q} f(x-k) e^{-2\pi ik \cdot x} \, dx,$$

which yields $\hat{f}(k) = \hat{Pf}(k)$.

Finally, if $f$ is continuous then the convergence of the numerical series $\sum_{k \in \mathbb{Z}^d} (1 + |k|)^{-d-\varepsilon} < \infty$ implies the absolute and uniform convergence of both series (5.7). Hence, the series $Pf$ belongs to $C(\mathbb{T}^d) \subset L^2(\mathbb{T}^d)$, and therefore the equality (5.7) holds almost everywhere (because $\{e_k\}$ is a complete orthonormal system or orthonormal basis), and then by continuity, the equality remains true for every $x$ in $\mathbb{T}^d$.

The space $C^\infty(\mathbb{T}^d)$ of smooth periodic functions is a Fréchet space with the uniform convergence, i.e., given by the sequence of seminorms

$$p_n(\varphi) = \sup \{|\partial^\alpha \varphi(x)| : x \in \mathbb{T}^d, |\alpha| \leq n \}, \quad n = 0,1, \ldots$$

This space can also be regarded as the space $C^\infty(\overline{Q})$ with the periodic conditions on the boundary $\partial Q$, which can be written as $\partial^\alpha \varphi(x) = \partial^\alpha \varphi(x + k)$ for every $x, x+k$ in $\partial Q$ and $k$ in $\mathbb{Z}^d$, and any multi-index $\alpha$. In this sense, the Fréchet space of periodic test functions could be denoted by either $\mathcal{D}(\mathbb{T}^d)$ or $\mathcal{D}_p(\overline{Q})$, with $\overline{Q} = [-1/2,1/2]^d$ being the closed unit cube in $\mathbb{R}^d$. Moreover, the elements in its dual space, denoted by either $\mathcal{D}'(\mathbb{T}^d)$ or $\mathcal{D}'_p(\overline{Q})$, are called periodic distributions. Note that an element in $\varphi$ in $\mathcal{D}_p(\overline{Q})$ can be extended by periodicity to the whole space $\mathbb{R}^d$ as an element of the Fréchet space $C^\infty_b(\mathbb{R}^d)$, infinite differentiable bounded functions with the uniform convergence. Thus, if $f$ is a periodic integrable function, i.e., $f$ is an element in $L^1(\mathbb{T}^d)$ then $f$ can be considered as a periodic distribution

$$\langle Tf, \varphi \rangle = \int_{\mathbb{T}^d} f(x) \varphi(x) \, dx, \quad \forall \varphi \in \mathcal{D}(\mathbb{T}^d).$$

The integration by part formula

$$\int_{\mathbb{T}^d} (\partial^\alpha f(x)) \varphi(x) \, dx = (-1)^{|\alpha|} \int_{\mathbb{T}^d} f(x) (\partial^\alpha \varphi(x)) \, dx$$

let us define the weak (or distribution sense) derivative of any element in $\mathcal{D}'(\mathbb{T}^d)$.
Early, the Fourier transform was defined for any function in $L^2(\mathbb{T}^d)$, so in particular, for any periodic test function $\varphi$, an element in $\mathcal{D}(\mathbb{T}^d)$, the expression

$$k \mapsto (\mathfrak{F}\varphi)(k) = \hat{\varphi}(k) = \int_{\mathbb{T}^d} \varphi(x)e^{-2\pi i k \cdot x}dx = \int_{Q} \varphi(x)e^{-2\pi i k \cdot x}dx,$$

defines a function from $\mathbb{Z}^d$ into the complex numbers $\mathbb{C}$, which produces an element in the dual Fréchet space $\mathfrak{s}$, i.e., an element in the dual space $\mathfrak{F}$

which implies that $\hat{\varphi}$ belongs to $\ell^p(\mathbb{Z}^d)$ for every $p \geq 1$. Actually, the Fourier transform $\mathfrak{F}$ is a linear and continuous operator from $\mathcal{D}(\mathbb{T}^d)$ into $\mathfrak{s}(\mathbb{Z}^d)$, the Fréchet space of rapidly decreasing sequences with index in $\mathbb{Z}^d$. Moreover, the series $\sum_{k \in \mathbb{Z}^d} \hat{\varphi}(k)e_k$, with $e_k(x) = e^{-2\pi i k \cdot x}$, converges in $\mathcal{D}(\mathbb{T}^d)$ to $\varphi$, and the equality $\mathfrak{F}(\partial^a \varphi)(k) = (-2\pi i)^{a_k}k^a\mathfrak{F}(\varphi)(k)$ holds true.

Therefore, by duality, the Fourier transform $\mathfrak{F}$ of a periodic distribution $T$, i.e., an element in the dual space $\mathcal{D}'(\mathbb{T}^d)$, is defined by

$$k \mapsto (\mathfrak{F}T)(k) = \hat{T}(k) = \langle T, e_{-k} \rangle,$$

which produces an element in the dual Fréchet space $\mathfrak{s}'(\mathbb{Z}^d)$ of the slowly decreasing sequences with index in $\mathbb{Z}^d$, i.e., sequences $\{\hat{T}(k) : k \in \mathbb{Z}^d\}$ such that there exist constants $n = n(T)$ and $C = C(T)$ satisfying

$$|\hat{T}(k)| \leq C(T)(1 + |k|^2)^{n/2}, \quad \forall k \in \mathbb{Z}^d,$$

(5.8)
i.e., the Fourier transform $\mathfrak{F}$ is a linear and continuous operator from $\mathcal{D}'(\mathbb{T}^d)$ into $\mathfrak{s}'(\mathbb{Z}^d)$. Moreover, the series $\sum_{k \in \mathbb{Z}^d} \hat{T}(k)e_k$ converges in $\mathcal{D}'(\mathbb{T}^d)$ to $T$.

As implicitly mentioned early, a periodic distribution can be thought as a continuous linear functional on either the Fréchet space $\mathcal{D}(\mathbb{T}^d)$ or on the lcvs $\mathcal{D}(\mathbb{R}^d)$ that are invariant under translations, i.e., an element in

$$\mathcal{D}'_p(\mathbb{R}^d) = \{ T \in \mathcal{D}'(\mathbb{R}^d) : \tau_k T = T, \quad \forall k \in \mathbb{Z}^d \},$$

where $\langle \tau_k T, \varphi \rangle = \langle T, \tau_{-k} \varphi \rangle$ and $\tau_{-k} \varphi(x) = \varphi(x + k)$, for every $x$ in $\mathbb{R}^d$. As in Theorem 5.9, the periodization mappings $P$ given by (5.6) is a linear continuous operator from $\mathcal{D}(\mathbb{R}^d)$ into $\mathcal{D}(\mathbb{T}^d)$ and so its dual operator $P'$ from $\mathcal{D}'(\mathbb{T}^d)$ into $\mathcal{D}'(\mathbb{R}^d)$ is defined by $\langle P'T, \varphi \rangle = \langle T, P\varphi \rangle$, for every $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$. Since $P \circ \tau_k = P$ for every $k$ in $\mathbb{Z}^d$, the range of $P'$ lies in $\mathcal{D}'_p(\mathbb{R}^d)$. In fact, let us check that

$$P' : \mathcal{D}(\mathbb{T}^d) \longrightarrow \mathcal{D}'_p(\mathbb{R}^d) \quad \text{is a bijection.}$$

(5.9)

Indeed, if $f$ has a compact support then the series defining the periodization operator $Pf = \sum_{k \in \mathbb{Z}^d} \tau_k f$ is a finite sum fore every $x$ within a compact set, so that it is clear that $P$ maps continuously $\mathcal{D}(\mathbb{R}^d)$ into $\mathcal{D}(\mathbb{T}^d)$. Now, chose a smooth kernel with compact support, i.e., an element

$$\eta \in \mathcal{D}(\mathbb{R}^d) \quad \text{such that} \quad \int_{\mathbb{R}^d} \eta(x)dx = 1,$$
and define
\[ \chi(x) = (\eta \ast 1_{[0,1]^d})(x) = \int_{[0,1]^d} \eta(x-y) dy. \]

It is clear that \( \chi \) belongs to \( \mathcal{D}(\mathbb{R}^d) \), and the equality
\[ \sum_{k \in \mathbb{Z}^d} \chi(x+k) = \sum_{k \in \mathbb{Z}^d} \int_{\mathbb{R}^d} 1_{[0,1]^d}(x+k-y) \eta(y) dy = \int_{\mathbb{R}^d} \eta(y) dy = 1, \]
shows that \( P \chi = 1 \). Now, any element \( \varphi \) in \( \mathcal{D}(\mathbb{T}^d) \) can be regarded also as a smooth function in \( \mathbb{R}^d \), and the equality
\[ \sum_{k \in \mathbb{Z}^d} \chi(x+k) \varphi(x+k) = \sum_{k \in \mathbb{Z}^d} \varphi(x+k) \int_{\mathbb{R}^d} 1_{[0,1]^d}(x+k-y) \eta(y) dy = \varphi(x) \sum_{k \in \mathbb{Z}^d} \chi(x+k), \]
implies that \( \varphi = P(\varphi \chi) \), where \( \varphi \) is regarded a function on the torus \( \mathbb{T}^d \) on the left and as a periodic function on \( \mathbb{R}^d \) on the right. Therefore, \( P : \mathcal{D}(\mathbb{R}^d) \rightarrow \mathcal{D}(\mathbb{T}^d) \) is surjective and its dual map \( P' : \mathcal{D}'(\mathbb{T}^d) \rightarrow \mathcal{D}'(\mathbb{R}^d) \) is injective. Similarly, for any \( T \) in \( \mathcal{D}'(\mathbb{R}^d) \) define the element \( S \) in \( \mathcal{D}'(\mathbb{T}^d) \) by \( \langle S, \varphi \rangle = \langle T, \chi \varphi \rangle \), where again, \( \varphi \) is regarded a smooth function on the torus \( \mathbb{T}^d \) on the left and as a periodic smooth function on \( \mathbb{R}^d \) on the right. This shows that \( P'S = T \), which proves that \( P' \) maps \( \mathcal{D}'(\mathbb{T}^d) \) onto \( \mathcal{D}'(\mathbb{R}^d) \), and that \( P \) is injective, i.e., (5.9) holds true. Furthermore, if \( f \) belongs to \( L^1(\mathbb{T}^d) \) then \( f \) and \( P'f \) coincide as periodic distributions, i.e.,
\[ \langle P'f, \varphi \rangle = \langle f, P\varphi \rangle = \int_{[0,1]^d} f(x) \sum_{k \in \mathbb{Z}^d} \varphi(x-k) dx = \sum_{k \in \mathbb{Z}^d} \int_{[0,1]^d+k} f(x) \varphi(x) dx = \int_{\mathbb{R}^d} f(x) \varphi(x) dx = \langle f, \varphi \rangle. \]

This means that both descriptions of ‘integrable’ periodic functions are equivalent, i.e., as a function in \( L^1(\mathbb{T}^d) \) regarded as being defined on either the torus \( \mathbb{T}^d \) or initially defined on the semi-closed unit cube \([-1/2, 1/2]^d \) and extended by periodicity to the whole space \( \mathbb{R}^d \), which is then regarded as a periodic distribution on \( \mathbb{R}^d \).

For any given \( T \) in \( \mathcal{D}'(\mathbb{T}^d) \) the Fourier series converges to \( T \), i.e.,
\[ \sum_{k \in \mathbb{Z}^d} \hat{T}(k) e_k = T \quad \text{in} \quad \mathcal{D}'(\mathbb{T}^d), \quad \text{with} \quad e_k(x) = e^{-2\pi ik \cdot x}, \]
and by means of (5.8), it is also clear that Fourier series converges as a tempered distribution, i.e.,
\[ P'T = \sum_{k \in \mathbb{Z}^d} \hat{T}(k) e_k \quad \text{in} \quad \mathcal{S}'(\mathbb{R}^d), \quad \text{with} \quad e_k(x) = e^{-2\pi ik \cdot x}. \]
Now, recall that
\[ \mathcal{F}(x^\alpha) = (-2\pi i)^{|\alpha|} \partial^\alpha \delta \quad \text{and} \quad \mathcal{F}(e_k) = \tau_k \delta, \]
where \( \delta \) is the Dirac measure, \( \langle \delta, \varphi \rangle = \varphi(0) \), for every \( \varphi \) in \( \mathcal{E}(\mathbb{R}^d) \), and since \( \mathcal{D}'_p(\mathbb{R}^d) \subset \mathcal{S}'(\mathbb{R}^d) \), it is then clear that
\[ \mathcal{F}(P'T) = \sum_{k \in \mathbb{Z}^d} \hat{T}(k) \mathcal{F}(e_k) = \sum_{k \in \mathbb{Z}^d} \hat{T}(k) \tau_k \delta \quad \text{in} \quad \mathcal{S}'(\mathbb{R}^d), \]
which is a clean relation between the \( \mathbb{R}^d \)- and the \( \mathbb{T}^d \)-Fourier transform for periodic distributions. In particular, if \( T = \delta_{T^d} \), the point mass at the origin in \( \mathbb{T}^d \), then \( \hat{T}(k) = 1 \) for every \( k \). Hence \( P'T \) and \( \mathcal{F}(P'T) \) are both equal to \( \sum_{k \in \mathbb{Z}^d} \tau_k \delta \), and in particular
\[ \sum_{k \in \mathbb{Z}^d} \mathcal{F}(e_k) = \sum_{k \in \mathbb{Z}^d} \tau_k \delta \quad \text{in} \quad \mathcal{S}'(\mathbb{R}^d), \]
which is a restatement of the Poisson summation formula.

For instance, the reader is referred to Folland [44, Chapters 8 and 9, pp. 235–311], where also a list of exercises can be found.

### 5.5 Fourier Multiplier

This is mainly an informative section taken from Bergh and Löfström [18, Section 6.1, pp. 131–139], Duoandikoetxea [40], Grafakos [57, Sections 2.5 and 5.2, pp. 135–146 and 359–371], and other. A multiplier operator or multiplier refers to a continuous linear operator between \( L^p(\mathbb{R}^d) \) that commute with translations. Perhaps the fist point it the following

**Theorem 5.10.** Let \( T \) be a continuous linear operator from \( L^p(\mathbb{R}^d) \) into \( L^q(\mathbb{R}^d) \), \( 1 \leq p \leq q \leq \infty \), that commutes with translations. If \( f \) is a smooth function rapidly decreasing at infinity, i.e., belonging to \( \mathcal{S}(\mathbb{R}^d) \), and \( Tf \) is regarded as a distribution then \( \partial^\alpha(Tf) = T(\partial^\alpha f) \), for every multi-index \( \alpha \). Moreover, there exists a unique tempered distribution \( T \) such that \( Tf = f \ast T \), for every \( f \) in \( \mathcal{S} \). Furthermore, if \( 1 < p \leq q < \infty \) then the restriction of \( T \) to the subspace \( L^d(\mathbb{R}^d) \cap L^p(\mathbb{R}^d) \), \( 1/q + 1/p' = 1 \), can be extended to a continuous linear operator from \( L^q(\mathbb{R}^d) \) into \( L^p(\mathbb{R}^d) \), \( 1/p + 1/p' = 1 \), and the operator norm coincide, i.e., \( \|T\|_{q,p'} = \|T\|_{p',q'} \).

**Proof.** To check the first assertion, take the partial derivative in some variable, e.g., in \( x_1 \), to see that \( (\tau_t \varphi - \varphi) t^{-1} \rightarrow \partial_1 \varphi \) in \( L^p(\mathbb{R}^d) \) for any \( \varphi \) in \( \mathcal{S} = \mathcal{S}(\mathbb{R}^d) \),
where $\tau_t^1$ is a translation operator, namely, $\tau_t^1 \varphi(x) = \varphi(x_1 + t, x_2, \ldots, x_d)$. Hence

$$\langle \partial_1(Tf), \varphi \rangle = -\int_{\mathbb{R}^d} Tf(x) \partial_1 \varphi(x) \, dx =$$

$$= -\lim_{t \to 0} \int_{\mathbb{R}^d} Tf(x) (\tau_t^1 \varphi(x) - \varphi(x))^t^{-1} \, dx =$$

$$= \lim_{t \to 0} \int_{\mathbb{R}^d} (Tf(x) - \tau_t^1 \varphi(x))^t^{-1} \varphi(x) \, dx =$$

$$= \lim_{t \to 0} \int_{\mathbb{R}^d} T[(f(x) - \tau_t^1 \varphi(x))^t] \varphi(x) \, dx = \langle T(\partial_1 f), \varphi \rangle,$$

as desired.

The second assertion is similar to Proposition 3.37. In view of the first assertion, if $\varphi$ is smooth then $T \varphi$ is also smooth, and the definition below

$$\varphi \mapsto \langle T, \varphi \rangle = (T \varphi)(0), \quad \forall \varphi \in S, \quad \text{with} \quad \tilde{\varphi}(x) = \varphi(-x),$$

makes sense. Next, the arguments in Proposition 3.40 and Remark 3.41 show that

$$|(T\varphi)(0)| \leq C \sup_{|\alpha| \leq d} \|\partial^\alpha(T\varphi)\|_q = C \sup_{|\alpha| \leq d} \|T(\partial^\alpha \varphi)\|_q,$$

for some constant $C$. Since $T$ is continuous from $L^p$ into $L^q$, this yields

$$|\langle T, \varphi \rangle| \leq C \|T\|_{q,p} \sup_{|\alpha| \leq d} \|\partial^\alpha \varphi\|_p \leq$$

$$\leq C \|T\|_{q,p} \|(1 + |\cdot|^2)^{-k/2}\|_p' \sup_{|\alpha| \leq d} \sup_{x \in \mathbb{R}^d} \{|(1 + |x|^2)^{k/2} \partial^\alpha \varphi(x)|\},$$

with $kp' > d$, $1/p + 1/p' = 1$. Hence $T$ is a tempered distribution. Now, if $\varphi$ belongs to $S$ then the equalities

$$\langle T \ast \varphi \rangle(x) = \langle T, \varphi(x - \cdot) \rangle = (T \varphi(x + \cdot))(0) = (T \varphi)(x)$$

show that $T \varphi = T \ast \varphi$ holds true.

To check the last part, consider the adjoint operator $T^*$, i.e.,

$$\langle Tf, g \rangle = \int_{\mathbb{R}^d} (Tf)(x) \overline{g(x)} \, dx = \int_{\mathbb{R}^d} f(x) \overline{(T^*g)(x)} \, dx = (f, T^*g),$$

for any $f$ in $L^p(\mathbb{R}^d)$ and $g$ in a dense subspace of $L^{q'}(\mathbb{R}^d)$. Since $T$ is given as a convolution with the tempered distribution $\mathcal{T}$, we can check that

$$(f, T^*g) = (Tf, g) = (T \ast f, g) = (T \ast f \ast \overline{g})(0) = (f \ast \overline{T} \ast g)(0) =$$

$$(f, \overline{T} \ast g) = (f, \overline{T} \ast g),$$

i.e., the adjoint operator $T^*$ is given by a convolution with the tempered distribution $\overline{T}$ or equivalently by

$$\varphi \mapsto \langle T^*, \varphi \rangle = (T \varphi)(0), \quad \forall \varphi \in S, \quad \text{with} \quad \overline{\varphi} = \Re(\varphi) - i \Im(\varphi),$$
using directly the complex-conjugate $\overline{\varphi}$ of $\varphi$. Thus $T^*$ is a continuous linear operator from $L^q(R^d)$ into $L^p(R^d)$ and $\|T^*\|_{p',q'} = \|T\|_{q,p}$, due to the restriction $1 < p, q < \infty$.

Therefore, remark that the expression $S : f \mapsto \overline{Tf}$ yields a continuous linear operator from $L^p(R^d)$ into $L^q(R^d)$, with operator norm $\|S\|_{q,p} = \|T\|_{q,p}$, to apply all the above arguments and to deduce that $S^* = T$ is a continuous linear operator from $L^q(R^d)$ into $L^p(R^d)$ with operator norm $\|T\|_{p',q'} = \|T\|_{q,p}$. □

• **Remark 5.11.** If $f$ and $g$ are two measurable functions with disjoint support then $\|f + g\|_p^p = \|f\|_p^p + \|g\|_p^p$, for $0 \leq p < \infty$. Thus, if $f$ is a measurable function vanishing outside of a compact support then

$$
\lim_{|h| \to \infty} \int_{R^d} |f(x + h) + f(x)|^p dx = 2 \int_{R^d} |f(x)|^p dx,
$$

and by density, this equality remains true any $f$ in $L^p$, i.e.,

$$
\lim_{|h| \to \infty} \|f(\cdot + h) + f\|_p = 2^{1/p} \|f\|_p, \quad \forall f \in L^p(R^d).
$$

(5.10)

Hence, if $T$ is a continuous linear operator from $L^p(R^d)$ into $L^q(R^d)$, $1 \leq q < p < \infty$, that commutes with translations then

$$
\|(Tf)(\cdot + h) - Tf\|_q = \|T(f(\cdot + h) - f)\|_q \leq \|T\|_{q,p} \|f(\cdot + h) - f\|_p.
$$

Now, apply (5.10) to get

$$
2^{1/q} \|Tf\|_q \leq 2^{1/p} \|T\|_{q,p} \|f\|_p,
$$

which implies that $Tf = 0$ since $2^{1/p-1/q} < 1$. This means that $T = 0$ is the only continuous linear operator from $L^p(R^d)$ into $L^q(R^d)$ when $p > q$. □

The vector space of all continuous linear operator from $L^p(R^d)$ into $L^q(R^d)$, $1 \leq p \leq q \leq \infty$ that commute with the translations is denoted by $M^{p,q} = M^{p,q}(R^d)$, which is endowed with the operator norm $\|\cdot\|_{q,p}$ to become a Banach space.

In view of the convolution representation $Tf = T \ast f$, the Fourier transform yields the expression

$$
T\varphi = \mathcal{F}^{-1}(\mathcal{F}(T)\mathcal{F}(\varphi)), \quad \forall \varphi \in \mathcal{S},
$$

For $1 \leq p < \infty$ the space $\mathcal{S}$ is dense in $L^p$, and therefore, the tempered distribution $T$ completely describes the operator $T$. This effectively explains the name **Fourier Multipliers** for the tempered distribution $\mathcal{F}(T)$.

**Definition 5.12.** A tempered distribution $m$ is called a **Fourier Multiplier** if $m = \mathcal{F}(T)$ for some $T$ in $M^{p,p}$, with $1 \leq p < \infty$ and $T\varphi = T \ast \varphi$, for every $\varphi$ in $\mathcal{S}$, i.e., if and only if $\varphi \mapsto \mathcal{F}^{-1}(m) \ast \varphi$ defines a continuous linear operator from $L^p(R^d)$ into itself. This space of Fourier Multipliers (or $L^p$-Multipliers) is denoted by $M^p = M^{p,p}(R^d)$ and endowed with the norm

$$
\|m\|_{M^p} = \sup \{ \|\mathcal{F}^{-1}(m) \ast \varphi\|_{L^p} : \|\varphi\|_{L^p} \leq 1 \},
$$

i.e., the operator norm of $T$. □
For $p = \infty$, if $M^\infty$ is defined as the tempered distributions $m$ such that $T_m : \varphi \mapsto \mathcal{F}^{-1}(m) \ast \varphi$ can be extended to a continuous linear operator from $L^\infty$ into itself, then the Riesz representation shows that $M^\infty$ is indeed the space of all finite Borel complex-valued measure $m$, and by definition, the norm of $m$ in $M^\infty$ is the operator norm $\|T_m\|_{\infty, \infty}$, which is equal to the total variation of $m$ (this argument is also used in Theorem 5.14 below). Note that since $\mathcal{S}(\mathbb{R}^d)$ is not dense in $L^\infty(\mathbb{R}^d)$, the inclusion $M^\infty \subset M^{\infty, \infty}$ is necessarily strict. In general, the elements in $M^\infty$ are not referred to as Fourier multipliers, even if technically, the $L^\infty$-multipliers are indeed the finite Borel complex-valued measure.

Moreover, the last part of Theorem 5.10 shows that $M^p = M^p'$ if $1 < p < \infty$. Thus, the only Fourier multipliers $M^p$, $1 \leq p \leq 2$ needs discussion. Indeed

$$M^1 \subset M^p \subset M^q \subset M^2, \quad \forall 1 \leq p \leq q \leq 2.$$  

As shown below in Theorem 5.13, the $L^2$-multipliers are the essentially bounded functions, and the inclusion above says that all Fourier multipliers are tempered distribution identified with essentially bounded functions. Note that $M^{p, q}$ refers to the continuous linear operator, while $M^p$ refers to the Fourier transform of the tempered distribution defining the operator via Theorem 5.10.

A characterization of $M^{p, q}$ in term of the tempered distribution appearing in the convolution is highly desired, but beside $M^\infty$, only the extreme cases $M^1$ and $M^2$ are actually well understood, while only sufficient conditions for being a $L^p$-multiplier are known.

**Theorem 5.13.** A tempered distribution $m$ is an $L^2$-multiplier, i.e., $m$ is a Fourier multiplier in $M^2(\mathbb{R}^d)$, if and only if $m$ is identified with an essentially bounded function, i.e., $m$ belongs to $L^\infty(\mathbb{R}^d)$. In this case, $\|m\|_{M^2} = \|m\|_{L^\infty}$.

**Proof.** If $m = \mathcal{F}(T)$ belongs to $L^\infty$ then Parseval-Plancherel equality (see Corollary 5.2) gives

$$\|T \ast f\|_2 = \|m \mathcal{F}(f)\|_2 \leq \|m\|_\infty \|\mathcal{F}(f)\|_2 = \|m\|_\infty \|f\|_2, \quad \forall f \in L^2,$$

i.e., $\|T\|_{2, 2} \leq \|m\|_\infty$ and $T$ belongs to $M^{2, 2}$, or equivalently $m$ belongs to $M^2$.

Let $m$ be a Fourier multiplier in $M^2$, with $m = \mathcal{F}(T)$ and $Tf = T \ast f$. To show that $\mathcal{F}(T)$ is in $L^\infty$, consider a smooth cutting function $\chi(x) = 1$ if $|x| \leq r$ and $\chi(x) = 0$ if $|x| > 2r$. The product of the smooth function $\chi$ with the distribution $\mathcal{F}(T)$ is

$$\chi \mathcal{F}(T) = \mathcal{F}(T \ast \mathcal{F}^{-1}(\chi)) = \mathcal{F}(T \mathcal{F}^{-1}(\chi)),$$

which implies that $\chi \mathcal{F}(T)$ belongs to $L^2$, i.e., the tempered distribution $\mathcal{F}(T)$ can be identified with a function in $L^2_{\text{loc}}$. Therefore, $f \mathcal{F}(T)$ belongs to $L^2$, for any function $f$ in $L^\infty$ with a compact support. Moreover, Parseval-Plancherel yields

$$\|f \mathcal{F}(T)\|_2 = \|T(\mathcal{F}^{-1}f)\|_2 \leq \|T\|_{2, 2} \|\mathcal{F}^{-1}f\|_2 = \|T\|_{2, 2} \|f\|_2,$$

and hence

$$\int_{\mathbb{R}^d} (\|T\|_{2, 2} - \mathcal{F}(T)(x)) |f(x)| \, dx \geq 0,$$
Chapter 5. Basic Fourier Transform

for every bounded function \( f \) with compact support. This implies \( \mathcal{F}(T)(x) \leq \|T\|_{2,2} \) almost everywhere, i.e., \( \mathcal{F}(T) \) belongs to \( L^\infty \) and \( \|\mathcal{F}(T)\|_\infty \leq \|T\|_{2,2} \).

Finally, as above, Parseval-Plancherel equality yields the reverse inequality and the proof is completed.

**Theorem 5.14.** An essentially bounded function is an \( L^1 \)-multiplier, i.e., \( m \) is a Fourier multiplier in \( M^1(\mathbb{R}^d) \), if and only if \( m \) is the Fourier transform of a finite Borel complex-valued measure \( T \), i.e., \( m = \mathcal{F}(T) \). In this case, the \( \|m\|_{M^1} \) is equal to the total variation of \( T \).

**Proof.** It is clear that if \( T \) is finite Borel complex-valued measure with variation measure \( |T| \) then operator \( T : f \mapsto T * f \) is a Fourier multiplier from \( L^1 \) into itself, and \( \|T\|_{1,1} \leq |T|((\mathbb{R}^d)) \). Conversely, if \( T \) is a given element in \( M^1 \) then Theorem 5.10 yields \( T f = T * f \), for every \( f \) in \( L^1 \) and for some tempered distribution \( T \). Choose \( f_\varepsilon(x) = \varepsilon^{-d}e^{-|x/\varepsilon|^2} \), with \( \varepsilon > 0 \) to see that

\[
\int_{\mathbb{R}^d} |T * f_\varepsilon(x)| \, dx \leq \|T\|_{1,1}, \quad \forall \varepsilon > 0.
\]

Hence, the total variation of the complex valued Borel (or Radon) measures

\[
\nu_\varepsilon(A) = \int_A T * f_\varepsilon(x) \, dx
\]

are equi-bounded.

Recall that Riesz representation Theorem B.90 (see also Remark B.91) affirms that the space of complex-valued Borel (or Radon) measures can be identified with the dual of the space \( C_*(\mathbb{R}^d) \) of continuous functions vanishing at infinity, i.e., any continuous linear functional on \( C_*(\mathbb{R}^d) \) has the form

\[
g \mapsto \int_{\mathbb{R}^d} g(x) \nu(dx),
\]

for some complex-valued Borel (or Radon) measure \( \nu \), and its operator norm is the total variation of \( \nu \).

Therefore, Alaoglu Theorem 2.7 can be applied to obtain a sequence \( \{\varepsilon_k\} \) of positive numbers and a complex-valued Borel (or Radon) measure \( \nu \) such that

\[
\lim_k \int_{\mathbb{R}^d} g(x) T * f_{\varepsilon_k}(x) \, dx = \int_{\mathbb{R}^d} g(x) \nu(dx), \quad \forall g \in C_*(\mathbb{R}^d).
\]

In particular, if \( g = \varphi \) with \( \varphi \) in \( S \) then

\[
\langle T * f_{\varepsilon_k}, \varphi \rangle = \langle (T * f_{\varepsilon_k}) * \tilde{\varphi} \rangle(0) = \langle T, \tilde{f}_{\varepsilon_k} * \varphi \rangle = \langle T, f_{\varepsilon_k} * \varphi \rangle \to \langle T, \varphi \rangle.
\]

This implies

\[
\langle T, \varphi \rangle = \int_{\mathbb{R}^d} \varphi(x) \nu(dx), \quad \forall \varphi \in S,
\]
which means that $T$ is the complex valued Borel (or Radon) measure $\nu$.

For any $g$ in the space $C_\ast(R^d)$,

$$\left| \int_{R^d} g(x) \nu(dx) \right| \leq (\sup_{R^d} |g|) \left( \sup_k \| T \ast f_{\varepsilon_k} \|_1 \right) \leq (\sup_{R^d} |g|) \| T \|_{1,1},$$

which show that the total variation of $\nu$ is no larger than the operator norm $\| T \|_{1,1}$, and then, the proof is completed. \qed

It is clear that operator given by a convolution with a complex valued Borel (or Radon) measure maps $L^\infty$ into itself, thus $M^1 = M^{1,1}$ is a subspace of $M^{\infty,\infty}$, but not necessarily the whole space. An expression of the form

$$\lim_{r \to \infty} \frac{1}{r} \int_0^r f(x) \, dx$$

can be used to show that indeed, $M^{\infty,\infty}$ is strictly larger than $M^1$, i.e., there are elements in $M^{\infty,\infty}$ which are convolution with tempered distributions that are not complex-valued Borel (or Radon) measures.

A translation operator $T_h : f \mapsto f(\cdot + h)$, $h \in R^d$, corresponds to the convolution with the tempered distribution $\tilde{\delta}_h$,

$$\tilde{\delta}_h \ast \varphi = \langle \tilde{\delta}_h, \varphi \rangle = \langle \delta_h, \tilde{\varphi} \rangle = \varphi(\cdot + h),$$

which yields the $L^1$-multiplier $\mathfrak{F}(\tilde{\delta}_h)(\xi) = m(\xi) = e^{2\pi \xi \cdot h}$.

It is not hard to show that the space of multipliers $M^p(R^d)$, $1 \leq p < \infty$ is a Banach space. Moreover, remaking that the composition of two operators $T_1 \circ T_2$ corresponds to the convolution of their associated tempered distributions $T_1 \ast T_2$, which is turn corresponds to the pointwise multiplications of their Fourier multipliers $m_1 m_2$, the space of multipliers $M^p$ becomes a Banach algebra with the pointwise multiplication. Furthermore, besides the inclusion

$$M^1 \subset M^p \subset M^q \subset M^2, \quad \forall 1 \leq p \leq q \leq 2,$$

mentioned early, for every $1 \leq p < q \leq \infty$, the interpolation estimate on the norm in $M^r$,

$$\| m \|_r \leq \| m \|_p^\theta \| m \|_q^{1-\theta}, \quad \frac{1}{r} = \frac{\theta}{p} + \frac{1-\theta}{q}, \quad 0 < \theta < 1,$$

holds true, see Riesz-Thorin Theorem 2.46.

Another simple result concerning an affine surjective transformation $a : R^n \to R^d$ establishes that the mapping $a^*$ given by $a^* m(\xi) = m(a(\xi))$, for any $\xi$ in $R^n$, defines an isometry between $M^p(R^d)$ and $M^p(R^n)$.

In dealing with the $L^p$-multipliers some singular integrals appear, particularly, the Littlewood-Paley theory is involved. Only two key results are stated below, the full proof can be found in the references mentioned at the beginning of this section.
Notation: First, dyadic $n$-intervals of the form

$$ I^{(n)}_j = I(j_1) \times \cdots \times I(j_n), \quad I(i) = (-2^{i+1}, -2^i) \cup [2^i, 2^{i+1}), $$

for multi-integers $j = (j_1, \ldots, j_n)$, $j_i = 0, \pm 1, \pm 2, \ldots$, $i = 1, \ldots, n$,
give a partition of $\mathbb{R}_n$, the complement of coordinates axes, for $n = 1, \ldots, d$,
i.e., $\mathbb{R}^n_\ast \mathbb{R}^n_\ast = \bigcup_j I^{(n)}_j$. Second, the variable $\xi = (\xi_1, \ldots, \xi_n)$ in $\mathbb{R}^d$ is partitioned into $\xi^{(n)} = (\xi_1, \ldots, \xi_n)$, $1 \leq n \leq d$, with $\xi^{(d-n)} = (\xi_{n+1}, \ldots, \xi_d)$ if $n < d$, to write $\xi = (\xi^{(n)}, \xi^{(d-n)})$. Third, write $\xi_{\pi} = (\xi_{\pi(1)}, \ldots, \xi_{\pi(d)})$, for any of the $d!$ permutations $\pi$ of $\{1, 2, \ldots, d\}$ to define $m_{\pi}(\xi) = m(\xi_{\pi})$. Fourth, recall that $\partial_i$ denotes the derivative with respect to the $i$-variable, while $\partial^\alpha$ denotes the derivative $\partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d}$, for any multi-indices $\alpha = (\alpha_1, \ldots, \alpha_d)$ of order $\alpha_1 + \cdots + \alpha_d = |\alpha|$.

With the previous dyadic notation, a complex-valued function $\xi \mapsto m(\xi)$ in $C^d(\mathbb{R}^d_\ast)$ satisfies the Marcinkiewicz multipliers condition if

\begin{align}
\text{(a)} \ & |m(\xi)| \leq C, \quad \forall \xi \in \mathbb{R}^d_\ast, \\
\text{(b)} \ & \int_{I^{(n)}_j} |\partial_1 \partial_2 \cdots \partial_n m_{\pi}(\xi^{(n)}, \xi^{(d-n)})| d\xi^{(n)} \leq C, \quad \forall \xi^{(d-n)}, \ j, \ n, \ \pi, \tag{5.11}
\end{align}

for some constant $C = C_m > 0$. Note that the bound

$$ |\partial_1 \partial_2 \cdots \partial_n m_{\pi}(\xi^{(n)}, \xi^{(d-n)})| \leq C|\xi_1|^{-1} \cdots |\xi_n|^{-1}, \quad \forall \xi^{(d-n)}, \ n, \ \pi, $$

implies condition (b) of (5.11).

If $d$ is even then set $n = d/2 + 1$ and if $d$ is odd then set $n = (d + 1)/2$, i.e., $n = [d/2] + 1$. Now a complex-valued function $\xi \mapsto m(\xi)$ in $C^n(\mathbb{R}^d_\ast)$, $\mathbb{R}^d = \mathbb{R}^d_\ast \{0\}$, satisfies the Mihlin multipliers condition if there exists a constant $B = B_m > 0$ such that

\begin{align}
\text{(a)} \ & |m(\xi)| \leq B, \quad \forall \xi \in \mathbb{R}^d, \\
\text{(b)} \ & |\partial^\alpha m(\xi)| \leq B|\xi|^{-|\alpha|}, \quad \forall \xi \in \mathbb{R}^d_\ast, \tag{5.12}
\end{align}

for every multi-indices $\alpha = (\alpha_1, \ldots, \alpha_d)$ of order $|\alpha| \leq n$. If assuming (a), the condition (b) is replaced by the following weaker condition

\begin{align}
\text{(b)} \ & \sup_{r > 0} r^{-d+2|\alpha|} \int_{r < |\xi| < 2r} |\partial^\alpha m(\xi)|^2 d\xi \leq B^2, \quad \forall \xi \in \mathbb{R}^d, \tag{5.13}
\end{align}

then this is referred to as the Hörmander multipliers condition.

**Theorem 5.15.** If $m(x)$ is a complex-valued function satisfying Marcinkiewicz multipliers condition (5.11) or Mihlin multipliers condition (5.12) or Hörmander multipliers condition (5.13) then $m$ is an $L^p$-multiplier for any $1 < p < \infty$, i.e., $m$ belongs to $M^p(\mathbb{R}^d)$ and the norm $\|m\|_{M^p}$ is dominated by a constant depending only on the dimension $d$, the exponent $p$ and the bound $C_m$ or $B_m$ appearing in the assumptions. \hfill $\square$
Note that under Mihlin or Hörmander multipliers conditions (5.12), (5.13),
is of weak-type \((1,1)\), i.e., the operator \(f \mapsto \mathcal{F}^{-1}(\mathcal{F}(\varphi)m)\) is continuous from \(L^1\) into weak-\(L^1\).

Examples of \(L^p\)-multipliers satisfying Marcinkiewicz multipliers condition (5.11) are the expressions of the form

\[
\frac{\xi_1}{\xi_1 + i(\xi_2^2 + \cdots + \xi_d^2)} \quad \text{or} \quad \frac{|\xi_1|^{\alpha_1} \cdots |\xi_d|^{\alpha_d}}{(\xi_1^2 + \cdots + |\xi_d|^2)^{(\alpha_1+\cdots+\alpha_d)/2}},
\]

for any multi-index \(\alpha\), with \(\alpha_i > 0\), for every \(j = 1, \ldots, d\). Similarly, the function \(|\xi|^{i_s}\) with \(s\) real satisfies Mihlin and Hörmander multipliers conditions (5.12), (5.13). Remark that a limitation of these sufficient conditions can be observed when some of the techniques used to prove Theorem 5.15 is adapted to show that the characteristic function of an arbitrary polyhedron in \(\mathbb{R}^d\) is an \(L^p\)-multiplier, even none of the sufficient assumption is satisfied. For instance, besides the references quoted at the beginning of this section, the interested reader may check, Stein [113, Chapter IV, pp. 81–115].
Chapter 6

Besov and Sobolev Spaces

Sobolev and Besov spaces are tools to deal with partial differential equations, and depending on the objectives in mind, alternative definitions are given. In the whole Euclidean space \( \mathbb{R}^d \), by means of the Fourier transform \( \mathcal{F} \), the Sobolev space of order \( s \) (any real number), and then the restrictions to \( \Omega \subset \mathbb{R}^d \) of functions in the whole space \( \mathbb{R}^d \) completes the definition. Alternatively, after setting-up the right norm and restricting the domain of integration from \( \mathbb{R}^d \) to \( \Omega \), the completion of the space of smooth functions under the right norm provides the desired spaces. Each method has its advantages and all ways lead to the same space when \( \Omega \) is sufficiently smooth.

Interpolation techniques work fine to extend the definition of Sobolev spaces from integer exponents \( k \) like in \( W^{k,p}(\Omega) \) to real exponents \( s \) like in \( W^{s,p}(\Omega) \), e.g., see Adams and Fournier [3, Chapter 7, pp. 205–260] or in general Bergh and L"ofstr"om [18]. For instance, a more “distribution” approach is found in Maz’ya [88, Chapter 1, pp. 1–121], while at the introductory level, Leoni [79, Chapters 14 and 15, pp. 415–476] consider only the whole space \( \Omega = \mathbb{R}^d \) and in Haroske and Triebel [66, Chapter 4, pp. 87–116] the local coordinates argument is detailed. A comprehensive discussion can be found in the Triebel[128]. Also, checking Taira [120, Chapter 6] and Wheeden and Zygmund [133, Chapter 14] may prove interesting for the reader.

In what follows and with the Fourier transform as a tool, Chapter 4 is reconsidered to complete certain aspects and viewpoints of interest. First, the neat case of Sobolev space with \( p = 2 \) (i.e., under a Hilbert structure) is discussed in some details. Next, Riesz and Bessel potentials are briefly presented as a tool to study later the general case \( p \neq 2 \) and to introduce the Besov spaces, which complete the discussion on the trace. All these spaces are very useful to study time-independent partial differential equations, particularly, elliptic equations. The equivalent type of spaces used for parabolic equations are developed on \( \mathbb{R}^d \times [0, \infty) \) (instead of \( \mathbb{R}^d \)) and they could be defined by means of a combination of the Fourier and the Laplace transforms, but those spaces are not

\[ \text{note that chapter 6 was added in the second edition!} \]
considered in this chapter (see references above as well as Ladyzhenskaya [77] and Lieberman [81], among others).

6.1 Hilbertian Sobolev Spaces

At the price of perhaps repeating some concepts, this subsection is kept as independent as possible. In this way, this section could be presented early in the book, e.g., in Chapter 5, right before Fourier Multiplier 5.5, and without covering the Introduction to Sobolev space Chapter 4. This approach can be found in several books, e.g., Chazarain and Piriou [26], Tartar [121], among others. Sometimes, for \( s = k \) an integer, these spaces are called Beppo Levi spaces, see Nečas [93].

6.1.1 In the Whole Space

All begin with the definition of a Sobolev space as a linear subspace of the space of the tempered distributions \( S' = S'(\mathbb{R}^d) \),

\[
    f \in H^s(\mathbb{R}^d) \text{ if and only if } \mathfrak{F}^{-1} \left( (1 + |\cdot|^2)^{s/2} \mathfrak{F}(f)(\cdot) \right) \in L^2(\mathbb{R}^d), \quad (6.1)
\]

or equivalently (via Parseval-Plancherel’s equality),

\[
    f \in H^s(\mathbb{R}^d) \text{ if and only if } (1 + |\cdot|^2)^{s/2} \hat{f} \in L^2(\mathbb{R}^d), \quad (6.2)
\]

for any real number \( s \). Basic properties of the Fourier transform (denoted either by \( \mathfrak{F}(\cdot) \) or by \( \hat{\cdot} \) the hat) are used to justify that the natural inner product

\[
    (f, g)_{s} = \int_{\mathbb{R}^d} (1 + |\xi|^2)^{s/2} \hat{f}(\xi) \overline{g}(\xi) \, d\xi, \quad \forall f, g \in H^s(\mathbb{R}^d), \quad (6.3)
\]

with the induced norm \( \|f\|_{H^s} = \sqrt{(f, f)_{s}} \) make \( H^s = H^s(\mathbb{R}^d) \) a (real or complex) Hilbert space, where the bar means complex conjugate.

Clearly, (a) \( H^0 = L^2 \), (b) if \( s \geq t \) then \( H^s \subset H^t \), which implies that \( H^s \subset L^2 \) for \( s > 0 \), and (c) \( S \subset H^s \), for every \( s \). The Sobolev space \( H^s \) can be identify to its dual space \( (H^s)' \) via the usual Riesz representation using the inner product \( \langle \cdot, \cdot \rangle_s \). However, it is more convenient to use the real \( L^2 \)-parity \( (\cdot, \cdot)_{L^2} \) or the distribution-evaluation parity \( \langle \cdot, \cdot \rangle \) to have

\[
    S \subset H^s \subset L^2 = (L^2)' \subset (H^s)' \subset S', \quad s > 0.
\]

With this understanding, \( H^{-s} = (H^s)' \), for \( s > 0 \), in the sense that \( f \) belongs to \( (H^s)' \) if and only if the linear mapping

\[
    \varphi \mapsto \langle (1 + |\cdot|^2)^{-s/2} \hat{f}, (1 + |\cdot|^2)^{s/2} \varphi \rangle,
\]

initially defined for \( \varphi \) in \( S \), can be extended to a continuous functional on \( H^s \), i.e., \( (1 + |\cdot|^2)^{-s/2} \hat{f} \) can be identify to a square-integrable function (an element in \( L^2 \)). Certainly, this relation goes both ways, i.e., \( (H^{-s})' = H^s \) for \( s > 0 \).
Therefore, an alternative definition is consider first $H^s$, for $s > 0$, as the linear subspace of $L^2$ satisfying the condition (6.1), and next define $H^s$ as the dual space $(H^{-s})'$ for $s < 0$. The point is to recall that the Fourier transform is an isomorphism from $S'$ into itself and to regard $H^s$ as a subspace of $L^2_{\text{loc}}(\mathbb{R}^d)$, namely, the pre-image $\mathcal{F}^{-1}(L^2_s)$, where $L^2_s$ is the $L^2$-space with the weighted Lebesgue measure $(1 + |\xi|^2)^{s/2}d\xi$ in $\mathbb{R}^d$.

Recall the particular case of $\Omega = \mathbb{R}^d$ and $p = 2$ in Chapter 4 used to define the Sobolev space $W^{s,2}(\mathbb{R}^d)$, with $s = m + s'$ with $m = [s]$ a nonnegative integer (the integer part of $s$) and $0 \leq s' < 1$ (the fractional part), i.e.,

\[
\begin{cases}
f \in W^{s,2}(\mathbb{R}^d) \text{ if and only if } & \partial^\alpha f \in L^2(\mathbb{R}^d), \quad \forall |\alpha| \leq [s], \\
\text{and also, if } & 0 < s' = s - [s] < 1 \text{ then } \sum_{|\alpha| = [s]} |\partial^\alpha f|_{s',2} < \infty, 
\end{cases}
\]

(6.4)

with the norm

\[
\|f\|_{s,2} = \left( \sum_{|\alpha| \leq [s]} \int_{\mathbb{R}^d} |\partial^\alpha f(x)|^2 dx + \sum_{[s] \leq |\alpha| < s} |\partial^\alpha f|_{s-[s],2}^2 \right)^{1/2},
\]

(6.5)

\[
|f|_{s',2} = \left( \int_{|x-y| \leq 1} |f(x) - f(y)|^2 |x-y|^{-d-2s'}dxdy \right)^{1/2},
\]

where is clear that the seminorm $|\cdot|_{s',2}$ plays a role only when $s$ is not integer, i.e.,

$0 < s' = s - [s] < 1$, see (4.1) with $m$ integer and the (4.23) with $s = s'$. Recall that the above definition is valid only for $s > 0$, while for $s < 0$, the Sobolev spaces $W^{s,2}(\mathbb{R}^d)$ are defined by duality, as the dual space of $W^{-s,2}(\mathbb{R}^d)$.

**Theorem 6.1.** With above notation, the norm $\| \cdot \|_{s,2}$ given by (6.5) and the norm $\| \cdot \|_{H^s}$ induced by the inner product (6.3) are equivalent in $S(\mathbb{R}^d)$, i.e.,

$$c\|\varphi\|_{H^s} \leq \|\varphi\|_{s,2} \leq C\|\varphi\|_{H^s}, \quad \forall \varphi \in S,$$

and therefore, the space $W^{s,2}(\mathbb{R}^d)$ coincides with $H^s(\mathbb{R}^d)$, for any real $s$. Moreover, a function $f$ belongs to $H^s(\mathbb{R}^d)$, $s > 0$, if and only if there exists a sequence $\{f_k\}$ of test functions, i.e., $\{f_k\} \in D(\mathbb{R}^d)$, such that (a) $\partial^\alpha f_k \to \partial^\alpha f$ in $L^2(\mathbb{R}^d)$ as $k \to \infty$, for every multi-index $\alpha$ of order $|\alpha| \leq [s]$, and if $s' = s - [s] > 0$ then (b) $|\partial^\alpha f_k - \partial^\alpha f|_{s',2} \to 0$ $k \to \infty$, for every multi-index $\alpha$ of order $|\alpha| = [s]$ and the seminorm (6.5).

**Proof.** First consider the case when $s$ is a positive integer, say $s = m$. Based on the property

$$\mathcal{F}(\partial^\alpha f)(\xi) = (2\pi i \xi)^\alpha \mathcal{F}(f)(\xi), \quad \text{with} \quad (2\pi i \xi)^\alpha = (2\pi i)^{|\alpha|} \xi_1^{\alpha_1} \ldots \xi_d^{\alpha_d},$$

for any multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$, with $|\alpha| = \alpha_1 + \ldots + \alpha_d$, and Parseval-Plancherel’s equality $\|\mathcal{F}(\varphi)\|_2 = \|\varphi\|_2$, (see Section 5.1), the norm $\| \cdot \|_{m,2}$ given by (6.5) or (4.1) for $W^{m,2}(\mathbb{R}^d)$ can be written as

$$\|f\|_{m,2} = \left( \sum_{|\alpha| \leq m} \int_{\mathbb{R}^d} |(2\pi \xi)^\alpha|^2 |\mathcal{F}(f)(\xi)|^2 d\xi \right)^{1/2}.$$
Hence, the estimate
\[
    c \leq (1 + |\xi|^2)^{-m} \sum_{|\alpha| \leq m} |(2\pi i \xi)^\alpha|^2 \leq C, \quad \forall \xi \in \mathbb{R}^d,
\]
for constants \( C \geq c > 0 \), yields equality \( H^m(\mathbb{R}^d) = W^{m,2}(\mathbb{R}^d) \).

Now, it is clear that only the case \( 0 < s = s' < 1 \) needs further consideration. The inequality
\[
    |f(x) - f(y)|^2 \leq 2(|f(x)|^2 + |f(y)|^2),
\]
the change of variable \( x = y + z \) yields the estimate
\[
    \iint_{|x-y|>1} |f(x) - f(y)|^2 |x-y|^{-d-2s'} \mathbb{I}_{|x-y|>1} \, dx \, dy \leq
\]
\[
    \leq 4 \left( \int_{\mathbb{R}^d} |f(y)|^2 \, dy \right) \left( \int_{\mathbb{R}^d} z^{-d-2s'} \mathbb{I}_{|z|>1} \, dz \right),
\]
which proves that the seminorm \( |f|_{s',2} \) could be defined by
\[
    |f|_{s',2} = \left( \iint_{\mathbb{R}^d \times \mathbb{R}^d} |f(x) - f(y)|^2 |x-y|^{-d-2s'} \, dx \, dy \right)^{1/2},
\]
and the resulting norms \( \| \cdot \|_{s,2} \) are equivalent. Retaining this notation, use Parseval-Plancherel’s equality to obtain
\[
    |f|_{s',2}^2 = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x) - f(z)|^2 \, dx \, dz =
\]
\[
    = \int_{\mathbb{R}^d} |\hat{f}(\xi)|^2 \, d\xi \int_{\mathbb{R}^d} |e^{i(z \cdot \xi)} - 1|^2 |z|^{-d-2s'} \, dz.
\]
Using the invariance under orthogonal transformations of the last integral in \( z \), for a fixed \( \xi \) in \( \mathbb{R}^d \) find an orthogonal transformation \( S \) that maps \( \xi/|\xi| \) into \( e_d = (0, \ldots, 1) \) to have \( Sz \cdot e_d = z \cdot S^{-1}e_d = z \cdot \xi/|\xi| \) and \( |Sz| = |z| \). Hence the change of variable \( y = |\xi|Sz \) with \( y = (y', y_d) \) yields
\[
    \int_{\mathbb{R}^d} \frac{|e^{i(z \cdot \xi)} - 1|^2}{|z|^{d+2s'}} \, dz = |\xi|^{2s'} \int_{\mathbb{R}} dy_d \int_{\mathbb{R}^{d-1}} \frac{|e^{iy_d} - 1|^2}{(|y'|^2 + y_d^2)^{(d+2s')/2}} \, dy',
\]
and furthermore, the change of variables \( y' = y_d z' \) in the integral over \( \mathbb{R}^{d-1} \) implies
\[
    \int_{\mathbb{R}} dy_d \int_{\mathbb{R}^{d-1}} \frac{|e^{iy_d} - 1|^2}{(|y'|^2 + y_d^2)^{(d+2s')/2}} \, dy' =
\]
\[
    = \int_{\mathbb{R}} \frac{|e^{iy_d} - 1|^2}{y_d^{2s'+1}} \, dy_d \int_{\mathbb{R}^{d-1}} \frac{dz'}{(|z'|^2 + 1)^{(d+2s')/2}}.
\]
Combine this calculation and the equality
\[ |e^{iy_{d}} - 1|^{2} = (\cos y_{d} - 1)^{2} + \sin^{2} y_{d} = 2 - 2 \cos y_{d} = 4 \sin^{2}(y_{d}/2) \]
to deduce
\[
\int_{\mathbb{R}} \frac{|e^{iy_{d}} - 1|^{2}}{y_{d}^{2s+1}} dy_{d} = 2^{3-2s'} \int_{0}^{\infty} \frac{\sin^{2} \lambda}{\lambda^{2s'+1}} d\lambda < \infty, \quad \text{if} \quad 0 < s' < 1.
\]
In any way, the point is that
\[
\int_{\mathbb{R}^{d}} |\xi|^{2s'} |\hat{f}(\xi)|^{2} d\xi, \quad (6.6)
\]
for a constant \( c_{d,s'} \) depending only on the dimension \( d \) and the exponent \( 0 < s' < 1 \). Actually, this constant can be expressed as
\[
c_{d,s'} = \frac{2^{1-2s'} \pi^{1+d/2}}{\Gamma(s+1) \Gamma(s+d/2) \sin(\pi s)} \Gamma(r) = \int_{0}^{\infty} t^{r-1} e^{-t} dt,
\]
where \( \Gamma(r), \ r > 0, \) is the Gamma function, i.e., \( c(d,s') \) behaves like \( 1/s' \) as \( s' \to 0 \) and \( 1/(1-s') \) as \( s' \to 1 \).

Therefore, invoke Parseval-Plancherel’s equality and the inequality \( c \leq (1 + |\xi|^{2})^{-s'} (1 + |\xi|^{2s'}) \leq C, \) for any \( \xi \) in \( \mathbb{R}^{d} \) and some constants \( C \geq c > 0 \), to deduce that the norm \( \| \cdot \|_{H^{2}} \) is equivalent to \( \| \cdot \|_{s',2} = (\| \cdot \|_{L^{2}}^{2} + \| \cdot \|_{s',2}^{2})^{1/2} \).

Certainly, the general case \( s > 0 \) follows from the inequality
\[
c \leq (1 + |\xi|^{2})^{-s} \left( \sum_{|\alpha| \leq [s]} |(2\pi i \xi)^{\alpha}|^{2} + |\xi|^{2s'} \sum_{|\alpha| = [s]} |(2\pi i \xi)^{\alpha}|^{2} \right) \leq C,
\]
for every \( \xi \) in \( \mathbb{R}^{d} \) and some constants \( C \geq c > 0 \), and a duality argument is used for the case \( s < 0 \).

The second part, regarding the density of the test functions \( \mathcal{D}(\mathbb{R}^{d}) \) in \( H^{s}(\mathbb{R}^{d}) \) is relatively simple, since a variate of tools are available. First, by convolution with a smooth kernel \( k_{\epsilon}(x) = \epsilon^{-d} k(x/\epsilon) \), as \( \epsilon \to 0 \), deduce that smooth functions are dense in \( H^{s}(\mathbb{R}^{d}) \). Next, for the compact support part, take a test function such that \( \chi(0) = 1 \) and consider the pointwise multiplication with the cutting function \( \chi_{s}(x) = \chi(x\eta) \), as \( \eta \to 0 \). Therefore, the expression \( f_{\epsilon,n} = (\chi_{n})(f \ast k_{\epsilon}) \) provides a suitable approximation. Perhaps, the only point to verify is the case \( 0 < s' = s < 1 \).

To this purpose, note that \( \widehat{k_{\epsilon}}(\xi) = \widehat{k}(\epsilon x), \ \widehat{\chi_{\eta}}(\xi) = \eta^{-d} \widehat{\chi}(x/\eta) \),
\[
\widehat{k_{\epsilon}}(0) = \int_{\mathbb{R}^{d}} k_{\epsilon}(x) dx = 1, \quad \text{and} \quad \int_{\mathbb{R}^{d}} \widehat{\chi_{\eta}}(\xi) d\xi = \chi(0) = 1.
\]
Thus, with the previous notation, the equality 6.6 implies
\[
|f - f \ast k_{\epsilon}|_{s',2}^{2} = c_{d,s'} \int_{\mathbb{R}^{d}} |\xi|^{2s'} |1 - \widehat{k_{\epsilon}}(\xi)|^{2} |\hat{f}(\xi)|^{2} d\xi,
\]
which shows that if \(|f|_{s',2} < \infty\) then \(|f - f \ast k_\varepsilon|_{s',2} \to 0\) as \(\varepsilon \to 0\). Similarly, from the definition of the seminorm,
\[
|f - \chi_\eta f|_{s',2}^2 \leq 2 \int_{\mathbb{R}^d \times \mathbb{R}^d} |\chi_\eta(x) - \chi_\eta(x + z)|^2 |f(x)|^2 |z|^{-d - 2s'} \, dx \, dz + 2 \int_{\mathbb{R}^d \times \mathbb{R}^d} |\chi_\eta(x + z)|^2 |f(x) - f(x + z)|^2 |z|^{-d - 2s'} \, dx \, dz,
\]
which implies that if \(|f|_{s',2} < \infty\) then \(|f - \chi_\eta f|_{s',2} \to 0\) as \(\eta \to 0\). \(\square\)

Let us mention a couple of points with short proofs:

0.- **Duality pairing and representation.** As mentioned early, the distribution-evaluation parity \(\langle \cdot, \cdot \rangle\) yields
\[
\mathcal{D} \subset \mathcal{S} \subset H^s \subset L^2 \subset H^{-s} \subset \mathcal{S}' \subset \mathcal{D}', \quad s > 0.
\]
where all spaces are separable and all inclusions are continuous and dense. The inner product in \(W^{s,2}(\mathbb{R}^d)\), \(s \geq 0\), is given by
\[
\begin{align*}
\langle f, g \rangle_s &= \sum_{|\alpha| \leq [s]} \int_{\mathbb{R}^d} \partial^\alpha f(x) \partial^\alpha g(x) \, dx + \sum_{|\alpha| < s} [\partial^\alpha f, \partial^\alpha g]_{s - [\alpha]}, \\
[f, g]_{s'} &= \int \int_{\{|x - y| \leq 1\}} \frac{(f(x) - f(y)) (g(x) - g(y))}{|x - y|^{d + 2s'}} \, dx \, dy, \quad s' > 0.
\end{align*}
\]
Remark that for any function \(h\) in \(L^2(\mathbb{R}^{2d})\) the operation
\[
\langle h_{s'}, \varphi \rangle = \int_{\mathbb{R}^d} \, dx \int_{|z| \leq 1} h(x, z) |z|^{-d/2 - s'} (\varphi(x + z) + \varphi(x)) \, dz
\]
defines a continuous linear functional on \(H^{s'}(\mathbb{R}^d)\), i.e., \(h_{s'}\) belongs to the dual space \((H^{s'}(\mathbb{R}^d))'\), namely \(H^{-s'}(\mathbb{R}^d)\), to deduce that a tempered distribution \(g\) belongs to \(H^{-s}(\mathbb{R}^d)\), \(m = [s]\), \(s' = s - [s]\), if and only if \(g\) can be represented as the distribution \(g = h_{s'} + \sum_{|\alpha| \leq m} \partial^\alpha g_{\alpha}\) for some functions \(h\) in \(L^2(\mathbb{R}^{2d})\) and \(g_{\alpha}\) in \(L^2(\mathbb{R}^d)\), and the norm
\[
\|g\|_s = \inf \left\{ \|h_{s'}\|_{L^2(\mathbb{R}^{2d})} + \sum_{|\alpha| \leq m} \|g_{\alpha}\|_{L^2(\mathbb{R}^d)} \right\},
\]
where the infimum is taken over all possible representation of \(g\), is equivalent to \(\| \cdot \|_{H^{-s}}\). Indeed, invoke Riesz representation of functionals for a Hilbert space to affirm that any element \(g\) belonging to \(H^{-s}\), \(s > 0\), can be represented as \(\langle g, f \rangle = \langle f, g \rangle_s\), for a unique \(g\) in \(H^s\). Hence, if \(h_{s'}(x, z) = [g(x + z) - g(x)] |z|^{-d/2 - s'}\) and \(g_{\alpha} = (-1)^{|\alpha|} \partial^\alpha g\) then \(f = h_{s'} + \sum_{|\alpha| \leq m} \partial^\alpha g_{\alpha}\) as a distribution with \(h_{s'}\) in \(L^2(\mathbb{R}^{2d})\) and \(g_{\alpha}\) in \(L^2(\mathbb{R}^d)\).

1.- **A function \(f\) in \(L^2(\mathbb{R}^d)\) belongs to \(H^1(\mathbb{R}^d)\) if and only if \(\|f - f(\cdot + z)\|_2 = O(|z|)\) as \(|z| \to 0\).** Indeed, for a function \(f\) in \(L^2(\mathbb{R}^d)\), the \(L^2\)-modulus of continuity is given by
\[
\|f - f(\cdot + z)\|_2 = \left( \int_{\mathbb{R}^d} |f(x + z) - f(x)|^2 \, dx \right)^{1/2}, \quad z \in \mathbb{R}^d,
\]
and it vanishes as $|z| \to 0$. If $f$ belongs to $H^1(\mathbb{R}^d)$ then the equality

$$f(x + z) - f(x) = \int_0^1 [z \cdot \nabla f(x + tz)] dt$$

shows that $|z|^{-1} \omega_2(f, z)$ is bounded as $|z| \to 0$, i.e., $\|f - f(\cdot + z)\|_2 = O(|z|)$ as $|z| \to 0$. Conversely, if $f$ belongs to $L^2(\mathbb{R}^d)$ and $\|f - f(\cdot + z)\|_2 = O(|z|)$ as $|z| \to 0$ then the family of functions $x \mapsto [f(x + e_i h) - f(x)]/h$ remains bounded in $L^2$ as $h \to 0$, which implies that the family is a weakly pre-compact in $L^2$, and so, the weak derivative $\partial_i f$ belongs to $L^2$.

2.- Case $0 < s < 1$: A function $f$ in $L^2(\mathbb{R}^d)$ belongs to $H^s(\mathbb{R}^d)$ if and only if

$$\int_{\mathbb{R}^d} \left( \|f - f(\cdot + z)\|_2 \right)^2 |z|^{-d-2s'}dz < \infty, \quad 0 < s < s' < 1.$$ 

Indeed, equality 6.6, in the proof of Theorem 6.1, shows that a function $f$ in $L^2(\mathbb{R}^d)$ belongs to $H^s(\mathbb{R}^d)$, with $0 < s = s' < 1$, if and only if $|f|_{s',2} < \infty$, as given by (6.5), or equivalently the above condition.

3.- Case $0 < s < 2$: A function $f$ in $L^2(\mathbb{R}^d)$ belongs to $H^s(\mathbb{R}^d)$ if and only if

$$|f|_{s,2}^2 = \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|f(x + z) - 2f(x) + f(x - z)|^2}{|z|^{d+2s}}dx dz < \infty, \quad 0 < s < 2,$$

and $f \mapsto (\|f\|_{L^2}^2 + |f|_{s,2}^2)^{1/2}$ provides an equivalent Hilbertian norm. Indeed, the use of the symmetrically modified $L^2$-modulus of continuity

$$\|f(\cdot + z) - 2f + f(\cdot - z)\|_2 = \left( \int_{\mathbb{R}^d} |f(x + z) - 2f(x) + f(x - z)|^2 dx \right)^{1/2},$$

for any $z$ in $\mathbb{R}^d$, allows an exponent $0 < s < 2$. As in obtaining (6.6), the equality

$$|e^{iz} - e^{-iz}|^2 = (2 \cos yd - 2)^2 = 16 \sin^4(yd/2)$$

yields

$$\int_{\mathbb{R}} \frac{|e^{iz} - e^{-iz}|^2}{|yd|^{2s+1}} dyd = 2^{5-2s} \int_0^\infty \frac{\sin^4 \lambda}{\lambda^{2s+1}} d\lambda, \quad \text{if} \quad 0 < s < 2,$$

which means that

$$\int_{\mathbb{R}^d} \frac{(\|f(\cdot + z) - 2f + f(\cdot - z)\|_2)^2}{|z|^{d+2s}}dz = c_{s,2}' \int_{\mathbb{R}^d} |\xi|^{2s} |\hat{f}(\xi)|^2 d\xi, \quad (6.10)$$

where the constant $c_{s,2}'$ depends only on the dimension $d$ and the exponent $0 < s < 2$. This establishes the desired assertion. In particular, the conditions $f$ in $L^2(\mathbb{R}^d)$ and

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|f(x + z) - 2f(x) + f(x - z)|^2}{|z|^{d+2}}dx dz < \infty$$

could be used as the definition of the Sobolev space $H^1(\mathbb{R}^d)$.  

4.- For a real number \( s \), if
\[
J_s : T \mapsto \mathfrak{F}^{-1}\left((1 + |\cdot|^2)^{s/2}\mathfrak{F}(T)(\cdot)\right)
\]
is considered as a continuous linear operator from the tempered distributions \( S' \) into itself, then the Sobolev spaces \( H^s(\mathbb{R}^d) \) is definite as the image of \( L^2(\mathbb{R}^d) \) via \( J_s \), i.e., \( H^s = J_s(L^2) \). Actually, if \( H^t(\mathbb{R}^d) \) is previously defined for some real number \( t \), then \( H^{t+s} = J_s(H^t) \) also define \( H^s(\mathbb{R}^d) \), for any real number \( s \). Indeed, this is the definition (6.1) of \( H^s \) for \( t = 0 \), i.e., \( H^s = J_s(L^2) \). For the general assertion, just note the composition equality \( J_s \circ J_t = J_{s+t} \) in \( S' \).

5.- The derivative \( \partial^\alpha \) is a continuous operator from \( H^s(\mathbb{R}^d) \) into \( H^{s-|\alpha|}(\mathbb{R}^d) \), for any \( s \), where \( |\alpha| = \alpha_1 + \cdots + \alpha_d \) is the order of the multi-index \( \alpha \). Indeed, this is the definition (6.1) of \( J \). In general, directly from the definition of \( H^s \), the inequality
\[
|\xi^\alpha|^2(1 + |\xi|^2)^{(s-|\alpha|)} \leq (1 + |\xi|^2)^s, \quad \forall \xi \in \mathbb{R}^d,
\]
yields the estimate \( \|\partial^\alpha f\|_{H^s} \leq \|f\|_{H^{s-|\alpha|}} \), which completes the arguments.

6.- If \( k \) is an integer such that \( s > d/2+k \) then \( H^s(\mathbb{R}^d) \) is continuously embedded in \( C^k(\mathbb{R}^d) \). Indeed, if \( f \) belongs to \( H^s \) and \( g = \partial^\alpha f \), with a multi-index \( \alpha \) of order \( |\alpha| \leq k \), then \( g \) belongs to \( H^{s-|\alpha|} \) and \( s - |\alpha| > d/2 \). Now, consider \( \hat{u}(\xi) = (1 + |\xi|^2)^{(s-|\alpha|)/2} \) and \( \hat{v}(\xi) = (1 + |\xi|^2)^{-(s-|\alpha|)/2}\hat{g}(\xi) \). Both functions \( \hat{u} \) and \( \hat{v} \) belong to \( L^2 \), \( \hat{g} = \hat{u}\hat{v} \) and \( \|\hat{v}\|_{L^2} = \|g\|_{H^{s-|\alpha|}} \). Hence Schwartz’s inequality implies
\[
\|g\|_{L^\infty} \leq \int_{\mathbb{R}^d} |\hat{g}(\xi)| d\xi \leq C_s \|g\|_{H^{s-|\alpha|}}, \quad C_s = \left( \int_{\mathbb{R}^d} (1 + |\xi|^2)^{-s-|\alpha|} d\xi \right)^{1/2},
\]
and \( g \) is continuous.

7.- The multiplication \( (\varphi, f) \mapsto \varphi f \) is a continuous operation from \( S(\mathbb{R}^d) \times H^s(\mathbb{R}^d) \) into \( H^s(\mathbb{R}^d) \), actually, \( \|\varphi f\|_{H^s} \leq \|\varphi\|_{L^\infty} \|f\|_{H^s} \). Indeed, the relation \( \hat{\varphi f} = \hat{\varphi} \hat{f} \) and Parseval-Plancherel’s equality yield
\[
\|\varphi f\|_{H^s}^2 = \int_{\mathbb{R}^d} (1 + |\xi|^2)^s \left| \int_{\mathbb{R}^d} \hat{\varphi}(\eta) \hat{f}(\xi - \eta) d\eta \right|^2 d\xi =
\int_{\mathbb{R}^d} (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 \left| \int_{\mathbb{R}^d} \hat{\varphi}(\eta) e^{-2\pi i \xi \cdot \eta} d\eta \right|^2 d\xi =
\int_{\mathbb{R}^d} (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 |\varphi(\xi)|^2 d\xi \leq \|\varphi\|_{L^\infty}^2 \|f\|_{H^s}^2.
\]
Alternatively, first note that
\[
(1 + |a + b|^2) \leq 1 + (|a| + |b|)^2 \leq 1 + 2|a|^2 + 2|b|^2 \leq 2(1 + |a|^2)(1 + |b|^2)
\]
yields
\[
(1 + |a + b|^2)^s \leq 2^s(1 + |a|^2)^s(1 + |b|^2)^s, \quad s \geq 0,
\]
which proves Peetre’s inequality, i.e.,

$$(1 + |\xi|^2)^s(1 + |\eta|^2)^{-s} \leq 2^{|s|}(1 + |\xi - \eta|^2)^{|s|}, \quad \forall \xi, \eta \in \mathbb{R}^d, \ s \in \mathbb{R},$$

(6.11)
after taking $a + b = \xi$, $a = \eta$ when $s \geq 0$ and exchange the role of $\eta$ and $\xi$ when $s < 0$. Now, Peetre’s inequality and the equality

$$(1 + |\xi|^2)^{s/2} \widehat{\varphi f}(\xi) =$$

$$= \int_{\mathbb{R}^d} (1 + |\xi|^2)^{s/2}(1 + |\eta|^2)^{-s/2} \widehat{\varphi}(\xi - \eta)(1 + |\eta|^2)^{s/2} \widehat{f}(\eta) d\eta$$
yield

$$(1 + |\xi|^2)^{s/2} |\widehat{\varphi f}(\xi)| \leq (u \ast v)(\xi), \quad \forall \xi \in \mathbb{R}^d,$$ with

$$u(\xi) = (1 + |\xi|^2)^{s/2} |\widehat{f}(\xi)| \quad \text{and} \quad v(\xi) = 2^{|s|}(1 + |\xi|^2)^{|s|/2} |\widehat{\varphi}(\xi)|.$$

Hence, use Young inequality (see Proposition B.65) and set

$$C_s(\varphi) = 2^{|s|} \int_{\mathbb{R}^d} (1 + |\xi|^2)^{|s|/2} |\widehat{\varphi}(\xi)| d\xi$$
to deduce the estimate $\|\varphi f\|_{H^s} \leq C_s(\varphi)\|f\|_{H^s}$.

**8.- Truncation and regularization:** If $k$ and $\chi$ belong to $\mathcal{S}(\mathbb{R}^d)$ and satisfy $\hat{k}(0) = \chi(0) = 1$ then $\|k_\varepsilon \ast f - f\|_{H^s} \to 0$ and $\|\chi_\varepsilon f - f\|_{H^s} \to 0$ as $\varepsilon \to 0$, where $k_\varepsilon(x) = \varepsilon^{-d}k(x/\varepsilon)$ and $\chi_\varepsilon(x) = \chi(\varepsilon x)$. Indeed, because $\hat{k}$ is bounded and continuous in $\mathbb{R}^d$, the equality

$$\|k_\varepsilon \ast f - f\|_{H^s} = \int_{\mathbb{R}^d} (1 + |\xi|^2)^s |\hat{k}(\varepsilon \xi) - 1|^2 |\hat{f}(\xi)|^2 d\xi$$
and shows the first part. Similarly, the first equality

$$\|\chi_\varepsilon f - f\|_{H^s}^2 = \int_{\mathbb{R}^d} (1 + |\xi|^2)^s \left| \int_{\mathbb{R}^d} \varepsilon^{-d} \hat{\chi}((\xi - \eta)/\varepsilon) [\hat{f}(\eta) - \hat{f}(\xi)] d\eta \right|^2 d\xi =$$

$$= \int_{\mathbb{R}^d} (1 + |\xi|^2)^s \left| \int_{\mathbb{R}^d} \hat{\chi}(\eta) [\hat{f}(\xi - \varepsilon \eta) - \hat{f}(\xi)] d\eta \right|^2 d\xi =$$

$$= \int_{\mathbb{R}^d} (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 \left| \int_{\mathbb{R}^d} \hat{\chi}(\eta) \left[ e^{-2\pi i \varepsilon \eta \cdot \xi} - 1 \right] d\eta \right|^2 d\xi$$
completes the arguments. Alternatively, use Peetre’s inequality (6.11) to get $\|\chi_\varepsilon f\|_{H^s} \leq C_s(\chi)\|f\|_{H^s}$, with a constant $C_s(\chi)$ independent of $0 < \varepsilon < 1$ and then a density argument to conclude.

**9.-** If $r < s < t$, $\varepsilon > 0$ and $C_\varepsilon = \varepsilon^{-(s-r)/(t-s)}$ then

$$\|f\|_{H^s}^2 \leq \varepsilon \|f\|_{H^t}^2 + C_\varepsilon \|f\|_{H^r}^2, \quad \forall f \in H^t.$$  

Indeed, this follows from the inequality

$$(1 + |\xi|^2)^s \leq \varepsilon(1 + |\xi|^2)^r + C_\varepsilon(1 + |\xi|^2)^t, \quad \forall \xi \in \mathbb{R}^d,$$

which can easily verified.
It could be convenient to define
\[
H^{+\infty} = \bigcap_{s \in \mathbb{R}} H^s = \bigcap_{m=0}^{\infty} H^m \quad \text{and} \quad H^{-\infty} = \bigcup_{s \in \mathbb{R}} H^s = \bigcup_{m=0}^{\infty} H^{-m}, \tag{6.12}
\]
and to note that (a) \( \mathcal{S}(\mathbb{R}^d) \subset H^{+\infty}(\mathbb{R}^d) \) and (b) \( \mathcal{E}'(\mathbb{R}^d) \subset H^{-\infty}(\mathbb{R}^d) \), i.e., a distribution with compact support belongs to some \( H^s \). Moreover, since an element of \( H^s(\mathbb{R}^d) \) is either a function in \( L^2(\mathbb{R}^d) \) or a tempered distribution belonging to the dual space of \( W^{s,2}(\mathbb{R}^d) \), we conclude that \( H^{-\infty}(\mathbb{R}^d) \) contains only tempered distributions of finite order.

### 6.1.2 In Continuous Domains

In most of what follows in this subsection, it is implicitly assume that \( \Omega \) is a domain (connected open sets which are equal to the interior of its closure) with at least a continuous boundary \( \partial \Omega \), i.e., for every point \( y \) on the boundary \( \partial \Omega \) there exists \( r = r(y) > 0 \), an orthogonal system of coordinates \((x', x_d)\), \( x' = (x_1, \ldots, x_{d-1}) \) and a continuous function \( \phi \) of \( x' \) such that \( \{ x \in \Omega : \| x - y \| < r \} = \{ x \in \mathbb{R}^d : \| x - y \| < r, \ x_d > \phi(x') \} \). As a simple example, the open set \( \{(x, y) \in \mathbb{R}^2 : x > 0, ax^2 < y < bx^2 \} \) has a continuous boundary if and only if \( ab < 0 \). Typical domains with a continuous boundary have the form \( \Omega_{\phi} = \{ x \in \mathbb{R}^d : x_d > \phi(x') \} \) for a uniformly continuous function \( \phi \). Note that the 'uniformly' is needed to deal with unbounded domain. The use of truncation and regularization with domains \( \Omega_{\phi} \) is a good example, as an extension of \( \mathbb{R}^d_0 \). If \( f \) is function defined on \( \Omega_{\phi} \), then the translation-down \( f_\varepsilon(x) = f(x', x_d + \varepsilon) \) plus a convolution with a test kernel is the expected approximation. The presence of \( \phi \) imposes intermediate truncation-regularization steps, namely, (1) use the uniform continuity and convolution to replace \( \phi \) by a smooth function \( \phi_\varepsilon \) such that \( |\phi(x') - \phi_\varepsilon(x')| \leq \varepsilon/6 \) for any \( x' \) in \( \mathbb{R}^{d-1} \), and (2) choose a smooth (cutting) function \( \eta \) satisfying \( \eta(t) = 0 \) for \( t \leq -2/3 \) and \( \eta(t) = 1 \) for \( t > -1/3 \), to define \( g_\varepsilon(x) = f_\varepsilon(x)\eta((x_d - \phi_\varepsilon(x')))/\varepsilon \), which satisfies \( g_\varepsilon(x) = f_\varepsilon(x) \) if \( x_d > \phi(x') - \varepsilon/6 \) and \( \partial^\alpha g_\varepsilon = \partial^\alpha f_\varepsilon \) as distributions on \( \Omega_{\phi} \), for any multi-index \( \alpha \). This cutting-smoothing technique is used to show that functions in Sobolev space \( W^{m,p}(\Omega) \) defined in Section 4 can be approximate by \( C_0^\infty(\mathbb{R}^d)|\Omega \), i.e., restrictions to \( \Omega \) of test functions in \( \mathbb{R}^d \).

In any case the focus is not on the type of 'smoothness' necessary on the domains \( \Omega \) to validate all results below. However, several assertions holds true for more general domains. Moreover, the main interest is on bounded domains, even if most of the times this is not mentioned. Begin with the following

**Definition 6.2.** If \( \Omega \) is an open subset of \( \mathbb{R}^d \) then \( H^s_{\text{loc}}(\Omega) \) is the linear space of all distributions \( f \) such that \( \chi f \) belongs to \( H^s(\mathbb{R}^d) \) for any \( \chi \) in \( \mathcal{D}(\Omega) \). The expression \( p_d(f) = \| \chi f \|_{H^s} \) defines a family of seminorms on \( H^s_{\text{loc}}(\Omega) \), which becomes a locally convex topological vector space (in short, lctvs). Similarly, if \( K \) is a compact subset of \( \Omega \) then \( H^s_{K}(\Omega) \) is the linear space of all (tempered) distributions \( f \) in \( H^s(\mathbb{R}^d) \) with support in \( K \), while \( H^s_{c}(\Omega) \) the linear space of all...
(tempered) distributions $f$ in $H^s(\mathbb{R}^d)$ with compact support in $\Omega$. Since $H^s_K(\Omega)$ is a closed subspace of $H^s(\mathbb{R}^d)$, it is also a Hilbert space with the $H^s(\mathbb{R}^d)$-inner product. However, $H^s_c(\Omega)$ becomes a lcvs with the inductive topology via $H^s_K(\Omega)$, with $K$ a compact of $\Omega$. Remark the notation $H^s_{loc} = H^s_{loc}(\mathbb{R}^d)$, $H^s_K = H^s_K(\mathbb{R}^d)$ and $H^s_c = H^s_c(\mathbb{R}^d)$.

The space $H^s_{loc}(\Omega)$ is a separable Fréchet space, i.e., complete metrizable lcvs, while $H^s_c(\Omega)$ is a complete lcvs not metrizable. The convergence $f_n \to 0$ in $H^s_c(\Omega)$ means: (a) there is a compact $K$ of $\Omega$ such that the support of $f_n$ is in $K$, i.e., $f_n$ belongs to $H^s_K(\Omega)$, and (2) $\|f_n\|_{H^s} \to 0$. Certainly, because $K$ is compact, the space $H^s_K(\Omega)$ contains only distributions with compact support, and $H^s_c(\Omega) = H^s(\mathbb{R}^d) \cap \mathcal{E}'(\Omega)$.

**Proposition 6.3.** The space $H^{-s}_c(\Omega)$ is the dual of the space $H^s_{loc}(\Omega)$ and the space $H^{-s}_{loc}(\Omega)$ is the dual of $H^s_c(\Omega)$, and therefore, both spaces are complete, separable and reflexive. Moreover,

$$\bigcap_{s \in \mathbb{R}} H^s_{loc}(\Omega) = C^\infty(\Omega) \quad \text{and} \quad H^s_{loc}(\Omega) \subset C^k(\Omega), \quad \text{if} \ s > d/2 + k,$$

and also, $H^s_{loc}(\Omega) \subset \mathcal{D}'(\Omega)$, and the both inclusions are continuous. Furthermore, $\mathcal{D}(\Omega)$ is dense in $H^s_{loc}(\Omega)$ and $H^s_c(\Omega)$, and the inclusions $\mathcal{E}(\Omega) \subset H^s_{loc}(\Omega)$ and $\mathcal{D}(\Omega) \subset H^s_c(\Omega)$ are continuous.

**Proof.** Indeed, for any element $g$ in $H^{-s}_c(\Omega)$ there is a cutting function $\chi$ in $\mathcal{D}(\Omega)$ such $\chi g = g$, this yields

$$f \mapsto \langle g, f \rangle = \langle g, \chi f \rangle = (\hat{\chi} f, \hat{g}) = \int_{\mathbb{R}^d} \hat{\chi} f(\xi) \overline{g}(\xi) d\xi,$$

i.e., $g$ is a continuous linear functional on $H^s_{loc}(\Omega)$, in other words, $H^{-s}_c(\Omega)$ is a subspace of the dual of $H^s_{loc}(\Omega)$, denoted by $(H^s_{loc}(\Omega))'$. Conversely, if $g$ is a continuous linear functional on $H^s_{loc}(\Omega)$ then there exists a constant $C > 0$ and seminorm $p_\chi$ such that

$$|\langle g, f \rangle| \leq C p_\chi(f) = C \|\chi f\|_{H^s}, \quad \forall f \in H^s_{loc}(\Omega).$$

Hence, the support of $g$ is contained into the support of $\chi$, i.e., $g$ belongs to $\mathcal{E}'(\Omega)$. As above, take $\chi$ in $\mathcal{D}(\mathbb{R}^d)$ such that $\chi = 1$ on the support of $g$ to deduce that $g = \chi g$ is a continuous linear functional on $H^s(\mathbb{R}^d)$, i.e., an element of $H^{-s}(\mathbb{R}^d)$, which proves that $H^{-s}_c(\Omega)$ is the dual space of $H^s_{loc}(\Omega)$.

On the other hand, linear functional $g$ on $H^s_c(\Omega)$ is continuous if and only if its restriction to any $H^s_K(\Omega)$ is continuous, and because $H^s_K(\Omega)$ is a closed linear subspace of $H^s(\mathbb{R}^d)$, it can be extended to the whole space, i.e., it belongs to $H^{-s}(\mathbb{R}^d)$. This proves that for any cutting function $\chi$ in $\mathcal{D}(\Omega)$, the linear functional $\chi g$ belongs to $H^{-s}(\mathbb{R}^d)$, in other words, $g$ belongs to $H^s_{loc}(\Omega)$. Similarly, any element in $H^{-s}_{loc}(\Omega)$ yields a continuous linear functional on $H^s_c(\Omega)$, i.e., the dual space of $H^s_c(\Omega)$ is $H^s_{loc}(\Omega)$.

We conclude the second part by recalling that $H^s(\mathbb{R}^d)$ is continuously embedded into $C^k(\mathbb{R}^d)$ if $s > d/2 + k$. \[\square\]
Regarding compactness, we have

**Theorem 6.4 (Reillich).** If \( s \) and \( t \) are tow real number such that \( s > t \) and \( K \) is a compact subset of \( \mathbb{R}^d \) then the identity maps closed bounded sets in \( H^s_K(\mathbb{R}^d) \) into compact sets in \( H^t_K(\mathbb{R}^d) \).

**Proof.** Let \( \{f_n\} \) be a sequence in \( H^s_K \) such that \( \|f_n\|_{H^s} \leq 1 \) for every \( n \). We need to show that there exists a subsequence which converges in \( H^t \).

To this end, choose a smooth cutting \( \chi \) for \( K \) (i.e., \( \chi \in \mathcal{D}(\mathbb{R}^d) \) with \( \chi = 1 \) in a neighborhood of \( K \)) to have: \( f = \chi f \), \( \hat{f} = \hat{\chi} \ast \hat{f} \), and \( \partial^\alpha \hat{f} = (\partial^\alpha \hat{\chi}) \ast \hat{f} \), for every \( f \in H^s_K \). Thus, begin with

\[
((\partial^\alpha \hat{\chi}) \ast \hat{f})(\xi) = \int_{\mathbb{R}^d} (1 + |\eta|^2)^{-s/2}(\partial^\alpha \hat{\chi})(\xi - \eta)(1 + |\eta|^2)^{s/2} \hat{f}(\eta) \leq \|f\|_{H^s} \left( \int_{\mathbb{R}^d} (1 + |\eta|^2)^{-s} |(\partial^\alpha \hat{\chi})(\xi - \eta)|^2 d\eta \right)^{1/2},
\]

and use Peetre’s inequality (6.11)

\[
(1 + |\eta|^2)^{-s} \leq 2^{-s}(1 + |\xi|^2)^{|s|}(1 + |\xi - \eta|^2)^{|s|},
\]

to obtain

\[
\int_{\mathbb{R}^d} (1 + |\eta|^2)^{-s} |(\partial^\alpha \hat{\chi})(\xi - \eta)|^2 d\eta \leq 2^{|s|}(1 + |\xi|^2)^{-s} \int_{\mathbb{R}^d} (1 + |\xi - \eta|^2)^{|s|} |(\partial^\alpha \hat{\chi})(\xi - \eta)|^2 d\eta = 2^{|s|}(1 + |\xi|^2)^{-s} \|\chi\|_{H^{|s|}},
\]

i.e.,

\[
(1 + |\xi|^2)^{s/2} |(\partial^\alpha \hat{f})(\xi)| \leq 2^{|s|} \|\chi\|_{H^{|s|}} \|f\|_{H^s}, \quad \forall f \in H^s_K(\mathbb{R}^d), \tag{6.13}
\]

for every multi-index \( \alpha \).

In view of \( \{f_n\} \subset H^s_K \) and \( \|f_n\|_{H^s} \leq 1 \), estimate (6.13) shows that \( \{\hat{f}_n\} \) is sequence equi-continuous and bounded on any compact set of \( \mathbb{R}^d \). Hence, use Arzela-Ascoli Theorem 2.9 to deduce that there exists a subsequence \( \{f_{n_k}\} \) such that \( \{\hat{f}_{n_k}\} \) converges uniformly on any compact set of \( \mathbb{R}^d \).

Now, to show that the subsequence \( \{f_{n_k}\} \) converges in \( H^t \), with \( t < s \), it suffice to show that it is a Cauchy sequence in the \( H^t \)-norm. To this end, decompose the integral in two pieces \( \mathbb{R}^d = \{|\xi| \leq r\} \cup \{|\xi| > r\} \) to get

\[
\|f_{n_k} - f_{n_t}\|_{H^t}^2 \leq \int_{|\xi| \leq r} (1 + |\xi|^2)^t |\hat{f}_{n_k} - \hat{f}_{n_t}|^2 d\xi + \int_{|\xi| > r} (1 + |\xi|^2)^{t-s}(1 + |\xi|^2)^s |\hat{f}_{n_k} - \hat{f}_{n_t}|^2 d\xi,
\]
and
\[
\int_{|\xi| > r} (1 + |\xi|^2)^{t-s}(1 + |\xi|^2)^s |\hat{f}_{nk} - \hat{f}_{n\ell}|^2 d\xi \leq (1 + r^2)^{t-s}\|f_{nk} - f_{n\ell}\|_{H^s}^2 \leq 4(1 + r^2)^{t-s}.
\]

This proves that \(\|f_{nk} - f_{n\ell}\|_{H^s} \to 0\) as \(k\) and \(\ell\) goes to \(\infty\), as desired. \(\square\)

Since a set is bounded in the lctvs \(H^s_H(\Omega)\) if and only if it is bounded in \(H^s_K(\Omega)\) for some compact \(K\) of \(\Omega\), Theorem 6.4 implies that the identity maps closed bounded sets in \(H^s_\varepsilon(\Omega)\) into compact sets in \(H^t_c(\Omega)\). Recall that \(\Omega\) is any open subset of \(\mathbb{R}^d\).

**Definition 6.5.** First, if \(F\) is a closed subset of \(\mathbb{R}^d\) then \(H^s_F = H^s_F(\mathbb{R}^d)\) denotes the closed linear space of all elements in \(H^s(\mathbb{R}^d)\) with support in \(F\). Next, if \(\Omega\) is an open subset of \(\mathbb{R}^d\) then \(H^s_0(\Omega)\) is the closure of the test functions \(\mathcal{D}(\Omega)\) in \(H^s(\mathbb{R}^d)\), while \(H^s(\Omega)\) denotes the linear space of all restriction to \(\Omega\) of tempered distributions in \(H^s(\mathbb{R}^d)\), i.e., an element \(f\) in \(H^s(\Omega)\) is an element in \(\mathcal{D}'(\Omega)\) which can be extended to become an element \(f^e\) in \(H^s(\mathbb{R}^d)\). In other words, if \(\big|\big|\big|_\Omega\) denotes the restriction to \(\Omega\) and it is regarded as an operator from \(H^s(\mathbb{R}^d)\) into \(\mathcal{D}'(\Omega)\) then its kernel is the closed linear space \(H^s_{\mathbb{R}^d, \Omega}(\mathbb{R}^d)\) and \(H^s(\Omega)\) is the quotient space \(H^s(\mathbb{R}^d)/H^s_{\mathbb{R}^d, \Omega}(\mathbb{R}^d)\), which is a Hilbert space with the quotient norm

\[
\|f\|_{H^s(\Omega)} = \inf\{\|f^e\|_{H^s} : f^e \in H^s, f^e|_\Omega = f\}.
\]

If the orthogonal complement is used to write \(H^s = H^s_{\mathbb{R}^d, \Omega} + (H^s_{\mathbb{R}^d, \Omega})^\perp\) then for each element \(f\) in \(H^s(\Omega)\) there exists a unique extension \(f^e\) orthogonal to \(H^s_{\mathbb{R}^d, \Omega}\) such that \(\|f\|_{H^s(\Omega)} = \|f^e\|_{H^s}\). \(\square\)

Note that a priori, the whole space \(H^s(\Omega)\) can not be identify as distribution in \(\Omega\), i.e., as a sub-space of \(\mathcal{D}'(\Omega)\), in the sense that an element in \(\mathcal{D}'(\Omega)\) may correspond to several elements in \(H^s(\Omega)\). Remark that \(H^s_F\) is a Hilbert space with the \(H^s\)-inner product, and it may contain distributions with non compact support. However, this is similar to \(H^s_K\) as in Definition 6.2, when \(K\) is a compact subset of \(\Omega \subset \mathbb{R}^d\). Also \(H^s_0(\Omega)\) is a Hilbert space with the \(H^s\)-inner product, and it can be regarded as a closed linear sub-space of \(H^s(\mathbb{R}^d)\) or \(H^s(\Omega) \cap \mathcal{D}'(\Omega)\), thus \(H^s_0(\Omega) = W^{s,2}_0(\Omega)\), as defined in Section 4.4. Moreover, since the derivative commutes with the restriction operation, it is clear that the differentiation operator \(\partial^\alpha\) of order \(|\alpha|\) maps continuously \(H^s_F(\mathbb{R}^d)\) or \(H^s_0(\Omega)\) or \(H^s(\Omega)\) into \(H^{s-|\alpha|}(\mathbb{R}^d)\) or \(H^{s-|\alpha|}(\Omega)\) or \(H^{s-|\alpha|}(\Omega)\), respectively. It is clear that if any distribution in \(H^s(\mathbb{R}^d)\) with support in the open set \(\Omega\) belongs to the space \(H^s_0(\Omega)\), but the converse fails.

Recall that in Chapter 4, the Sobolev space \(W^{s,2}(\Omega)\), with \(s = m + s'\) with \(m = [s]\) a nonnegative integer (the integer part of \(s\)) and \(0 \leq s' < 1\) (the
fractional part), were defined as

\[
\begin{cases}
  f \in W^{s,2}(\Omega) & \text{if and only if} \quad \partial^\alpha f \in L^2(\Omega), \quad \forall |\alpha| \leq \lfloor s \rfloor, \\
  \text{and also, if } 0 < s' = s - \lfloor s \rfloor < 1 & \text{then } \sum_{|\alpha| = \lfloor s \rfloor} |\partial^\alpha f|_{s',2}^2 < \infty, 
\end{cases}
\tag{6.14}
\]

with the inner product

\[
\begin{cases}
  \langle f, g \rangle_s = \sum_{|\alpha| \leq \lfloor s \rfloor} \int_\Omega \partial^\alpha f(x) \overline{\partial^\alpha g(x)} \, dx + \sum_{\lfloor s \rfloor \leq |\alpha| < s} \langle \partial^\alpha f, \partial^\alpha g \rangle_{s' - \lfloor s \rfloor}, \\
  [f, g]_{s'} = \iint_{\{x, y \in \Omega : |x - y| \leq 1\}} \frac{(f(x) - f(y))(g(x) - g(y))}{|x - y|^{d + 2s'}} \, dx \, dy,
\end{cases}
\tag{6.15}
\]

where is clear that the seminorm $| \cdot |_{s',2}$ induced by the semi-inner product $[\cdot, \cdot]_{s'}$ plays a role only when $s$ is not integer, i.e., $0 < s' = s - \lfloor s \rfloor < 1$, see (4.1) with $m$ integer and the (4.23) with $s = s'$. Recall that the above definition is valid only for $s > 0$, while for $s < 0$, the Sobolev spaces $W^{s,2}(\Omega)$ are defined by duality, as the dual space of $W_0^{-s,2}(\Omega)$, which is the closure in $W^{-s,2}(\Omega)$, $s < 0$, of the text functions $D(\Omega)$.

Referring to Definition 4.4, for a domain $\Omega$ of class $C^{m,\alpha}$, there is a locally finite open cover $\{\mathcal{O}_i\}$ of $\Omega$ and construct a regular partition of unity subordinate to this covering (i.e., $\sum_i \chi_i(x) = 1$, $\chi_i$ is $C^\infty$ with compact support in $\mathcal{O}_i$) with the following properties:

(a) For every $i$, we have either $d(\mathcal{O}_i, \partial \Omega) > 0$ or $\mathcal{O}_i \cap \partial \Omega \neq \emptyset$;

(b) There exists one-to-one transformations $y = Y_i(x)$ of class $C^{m,\alpha}$ mapping $\mathcal{O}_i$ into either the open ball $B = \{y \in \mathbb{R}^d : |y| < 1\}$ or the open half-ball $B_+ = \{y \in \mathbb{R}^d_+ : |y| < 1\}$, where the image of $\mathcal{O}_i \cap \partial \Omega$ is a flat part of $\partial B_+$.

If $\Omega$ is bounded then the open cover $\{\mathcal{O}_i\}$ is finite and if only the boundary $\partial \Omega$ is bounded then $\partial \mathcal{O}_i \cap \partial \Omega \neq \emptyset$ only for finite many $i$. Usually, on the case of bounded domain is of our interest.

**Definition 6.6.** A domain $\Omega$ has the extension property $(s, p)$ if there exists a continuous linear extension operator from $W^{t,p}(\Omega)$ into $W^{t,p}(\mathbb{R}^d)$, for every $0 < t \leq s$. In view of Theorem 4.6, a domain of class $C^{m-1,1}$ has the extension property $(m, p)$, e.g., see Nečas [93, Theorem 3.9, pp. 70–71]. Also, the special domain $\mathbb{R}^+$ has the extension property $(s, p)$ for every $s$ and $1 \leq p \leq \infty$. □

**Theorem 6.7.** For any $s \geq 0$, the spaces $W^{s,2}(\Omega)$ and $H^s(\Omega)$ are the same if $\Omega$ has the extension property $(s,2)$, with $m = \lfloor s \rfloor$, the integer part of $s$, i.e., the quotient norm $\| \cdot \|_{H^s(\Omega)}$ and the norm $\| \cdot \|_{s,2}$ induced by the inner product (6.14) are equivalent.

**Proof.** First, the case $s = 0$ is trivial since $W^{0,2}(\Omega) = L^2(\Omega) = H^0(\Omega)$, and everything else reduces $\Omega = \mathbb{R}^d$ as in Theorem 6.1. Indeed, if $s > 0$ and a continuous linear extension operator $E$ exists then the inclusion $W^{s,2}(\Omega)$ into
Proposition 6.8. If \( s \geq 0 \) then \( H^{-s}(\Omega) = H_{0}^{-s}(\Omega) \) and this is the space of all distributions of the form

\[
\langle f, \varphi \rangle = \sum_{|\alpha| \leq s} \int_{\Omega} dx \int_{|z| \leq 1} \frac{f_{\alpha}'(x,z)(\partial^{\alpha}\varphi(x+z) - \partial^{\alpha}\varphi(x))}{|z|^{d/2+s-[\alpha]}} \, dz + \sum_{|\alpha| \leq s} (-1)^{|\alpha|} \int_{\Omega} f_{\alpha}(x) \partial^{\alpha}\varphi(x) \, dx, \quad \forall \varphi \in \mathcal{D}(\Omega),
\]

for some functions \( f_{\alpha} \) in \( L^{2}(\Omega) \) and \( f_{\alpha}' \) in \( L^{2}(\Omega \times \mathbb{R}^{d}) \), and the norm \( \| \cdot \|_{H^{-s}(\Omega)} \) is equivalent to

\[
\| f \|_{-s} = \inf \left\{ \sum_{|\alpha| \leq |\alpha| < s} \| f_{\alpha}' \|_{L^{2}(\Omega \times \mathbb{R}^{d})} + \sum_{|\alpha| \leq s} \| f_{\alpha} \|_{L^{2}(\Omega)} \right\},
\]

where the infimum is taken over all possible decompositions of \( f \) as above.

Proof. As shown in the previous subsection \( H^{s}(\mathbb{R}^{d}) = W^{s,2}(\mathbb{R}^{d}) \), see Theorem 6.1, which implies that the above statements are true if \( \Omega = \mathbb{R}^{d} \).

First consider the case \( s = m \) integer. Now, if \( f \) is an element in \( \mathcal{D}'(\Omega) \) of the form \( f = \sum_{|\alpha| \leq m} \partial^{\alpha} f_{\alpha} \), with \( f_{\alpha} \) in \( L^{2}(\Omega) \) then it suffices to extend the function \( f_{\alpha} \) by zero to the whole space \( \mathbb{R}^{d} \), denoted by \( f_{\alpha}^{0} \), and to define \( f^{0} = \sum_{|\alpha| \leq m} \partial^{\alpha} f_{\alpha}^{0} \). Since \( f^{0} \) belongs to \( H^{-m}(\mathbb{R}^{d}) \), we deduce that \( f \) is in \( H^{-m}(\Omega) \).

Conversely, if \( f \) belongs to \( H^{-m}(\Omega) \) then find \( f^{e} \) in \( H^{-m}(\mathbb{R}^{d}) \) such that \( f = f^{e}|_{\Omega} \) and \( \| f \|_{H^{-m}(\Omega)} = \| f^{e} \|_{H^{-m}} \). For this \( f^{e} \), there exists \( f^{e}_{\alpha} \) in \( L^{2}(\mathbb{R}^{d}) \), \( |\alpha| \leq m \), satisfying

\[
f^{e} = \sum_{|\alpha| \leq m} \partial^{\alpha} f^{e}_{\alpha} \quad \text{and} \quad \| f^{e} \|_{H^{-m}} = \sum_{|\alpha| \leq m} \| f^{e}_{\alpha} \|_{L^{2}}.
\]

Therefore, consider the restrictions \( f_{\alpha} = f^{e}_{\alpha}|_{\Omega} \) to deduce that \( f = \sum_{|\alpha| \leq m} \partial^{\alpha} f_{\alpha} \) and \( \| f \|_{-m} \leq \| f^{e} \|_{H^{-m}} = \| f \|_{H^{-m}(\Omega)} \).

Similarly, the elements of \( H^{-s'}(\Omega), 0 < s' < 1 \), have the form

\[
\langle f, \varphi \rangle = \int_{\Omega} dx \int_{|z| \leq 1} \frac{f_{s'}(x,z)(\varphi(x+z) - \varphi(x))}{|z|^{d/2+s'}} \, dz + \int_{\Omega} f_{0}(x) \varphi(x) \, dx, \quad \forall \varphi \in \mathcal{D}(\Omega),
\]

for some functions \( f_{0} \) in \( L^{2}(\Omega) \) and \( f_{s'} \) in \( L^{2}(\Omega \times \mathbb{R}^{d}) \).

In view of the above, the equality \( H^{-s}(\Omega) = H_{0}^{-s}(\Omega) \) follows from the density of the test functions \( \mathcal{D}(\Omega) \) in \( L^{2}(\Omega) \). □
Proposition 6.9. With the previous notation: (a) the space $\mathcal{C}_0^\infty(\mathbb{R}^d)|_\Omega$ is dense in $H^s(\Omega)$, (b) the spaces $H^s(\Omega)$ and $H_{-s}^\infty(\mathbb{R}^d)$ are identified one with each other by the pairing

$$\langle f, g \rangle = \langle f^e, g \rangle, \quad \forall f \in H^s(\Omega), \ g \in H_{-s}^\infty(\mathbb{R}^d), \ f^e \in H^s(\mathbb{R}^d),$$

where $f^e$ is an arbitrary extension of $f$, (c) the dual space of $H_0^s(\Omega)$ is identified with $H_{-s}^0(\mathbb{R}^d)$, and (d) if $s \geq 0$ then $H_{-s}^\infty(\mathbb{R}^d) = H_0^s(\Omega)$.

Proof. Since $\mathcal{C}_0^\infty(\mathbb{R}^d)$ is dense in $H^s(\mathbb{R}^d)$ and the restriction operator is continuous, we verify part (a).

To check part (b), first let us prove that:

(*) If $f$ belongs to $H^s(\mathbb{R}^d)$ with support in $\bar{\Omega}$ and $g$ belongs to $H_{-s}^\infty(\mathbb{R}^d)$ with support in $\mathbb{R}^d \setminus \Omega$ then $\langle f, g \rangle = 0$.

Indeed, if $s \geq 0$ then the distribution $f$ can be identified with an element in $L^2(\mathbb{R}^d)$ and therefore $f(x) = 0$ a.e. $x$ in $\mathbb{R}^d \setminus \bar{\Omega}$, while the distribution $g$ has the form

$$\langle g, \varphi \rangle = \sum_{|\alpha| \leq |s|} \int_{\mathbb{R}^d} dx \int_{\{|z| \leq 1\}} \frac{g'_\alpha(x, z)(\partial^\alpha \varphi(x + z) - \partial^\alpha \varphi(x))}{|z|^{d/2 + s - |\alpha|}} \, dz + \sum_{|\alpha| \leq |s|} (-1)^{|\alpha|} \int_{\mathbb{R}^d} g_\alpha(x) \partial^\alpha \varphi(x) \, dx, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d),$$

for some functions $g_\alpha$ in $L^2(\mathbb{R}^d)$ and $g'_\alpha$ in $L^2(\mathbb{R}^{2d})$. Thus, it is clear that the distribution $g$ does not change if the functions $g_\alpha$ and $h'_\alpha$ are set equal to zero outside the support of $g$, i.e., $h_\alpha(x) = 0$ and $h'_\alpha(x, z) = 0$ a.e. $x$ in $\Omega$. Hence, use this representation and the fact that the boundary $\partial \Omega$ has Lebesgue measure zero, to deduce that $\langle f, g \rangle = 0$. Reversing the roles of $f$ and $g$, the assertion is also valid for $s < 0$. Alternatively, assume a smooth domain and use local coordinates to reduce to the case $\Omega = \mathbb{R}^d_+$. Hence, a convolution with a smooth function with support in the open set $\{x \in \mathbb{R}^d : x_d < 0\}$ yields $\langle f, g \rangle = 0$.

Now, apply (*) above to deduce that the pairing $\langle f, g \rangle = \langle f^e, g \rangle$ does not depend on the particular extension $f^e$ used, i.e., the pairing is well defined, and $|\langle f, g \rangle| \leq \|f^e\|_{H^s}\|g\|_{H_{-s}}$ yields $|\langle f, g \rangle| \leq \|f\|_{H^s(\Omega)}\|g\|_{H_{-s}}$, after taking infimum over all possible extensions.

Conversely, if $T$ is a continuous linear functional on $H^s(\Omega)$, i.e., an element in its dual space with norm $\|T\|$, then the linear functional $T_\Omega : f \mapsto \langle T, f \rangle_{\Omega}$ belongs to the dual space of $H^s(\mathbb{R}^d)$, i.e, $T$ is a distribution belonging to $H_{-s}^\infty(\mathbb{R}^d)$, satisfying $\langle T_\Omega, f \rangle = 0$ for every $f$ with support in $\mathbb{R}^d \setminus \Omega$. Hence $T_\Omega$ belongs to $H_{-s}^\infty(\mathbb{R}^d)$.

Similarly, if $T$ is a continuous linear functional on $H_{-s}^\infty(\mathbb{R}^d)$, i.e., an element in its dual space, then it can be extended (Hahn-Banach Theorem 2.26) to a continuous linear functional on the whole $H_{-s}^\infty(\mathbb{R}^d)$ (with the same norm),
which is identify with a distribution (still denoted by $T$) in $H^s(\mathbb{R}^d)$. Hence, the restriction $T|_\Omega$ belongs to $H^s(\Omega)$ and

$$
\|T|_\Omega\|_{H^s(\Omega)} \leq \|T\|_{H^s} = \|T\|_{H^{-s}} = \|T\|,
$$

which concludes part (b).

Since $H^s_0(\Omega)$ is a reflexive (Hilbert) space, to verify part (c), it suffices to show that the dual space of $H^s_0(\Omega)$, $s > 0$, is the space $H^{-s}_0(\Omega)$. To this end, proceed similar to Proposition 6.8 and recall that $H^s_0(\Omega)$ is a closed linear subspace of $H^s(\mathbb{R}^d)$ and the inner product initially given to $H^s(\mathbb{R}^d)$ is equivalent to $(\cdot, \cdot)_s$ given by (6.8).

Thus, invoke Riesz representation of functionals for a Hilbert space to affirm that any element $g$ in $H^s_0(\Omega)$ can be represented as $f \mapsto (\langle f, g \rangle)_s$, for a unique $g$ in $H^s_0(\Omega)$. Since $f$ and $g$ vanish on $\mathbb{R}^d \setminus \Omega$, the inner product in the whole space $\mathbb{R}^d$ (6.8) is reduced to inner product on $\Omega$ given by (6.15). The density of the test functions $\mathcal{D}(\Omega)$ in $H^s_0(\Omega)$ implies that the unique element $g$ in $H^s_0(\Omega)$ obtained can be regarded as a distribution on $\Omega$, i.e., an element in $\mathcal{D}'(\Omega)$.

This means that the given continuous linear functional on $H^s_0(\Omega)$ can be identified with the distribution

$$
T_g = \sum_{|\alpha| \leq [s]} \partial^\alpha g_\alpha + \sum_{s \leq |\alpha| < [s]} g'_\alpha, \quad \text{with}
$$

$$
g_\alpha = (-1)^{|\alpha|} \partial^\alpha g, \quad g'_\alpha(x, z) = [\partial^\alpha g(x + z) - \partial^\alpha g(x)]|x|^{-d-s+[s]},
$$

where $g_\alpha$ belongs to $L^2(\Omega)$, $g'_\alpha$ belongs to $L^2(\Omega \times \mathbb{R}^d)$, and the distribution $g'_\alpha$ is given by

$$
\langle g'_\alpha, \varphi \rangle = \int_{\Omega} dx \int_{|z| \leq 1} g'_\alpha(x, z) \frac{[\varphi(x + z) - \varphi(x)]}{|z|^{d+s-[s]}} dz, \quad \forall \varphi \in \mathcal{D}(\Omega).
$$

Moreover, $\|g\|_{H^s_0(\Omega)}$ is the norm $T_g$ as a linear functional on $H^s_0(\Omega)$, which is bounded by the sum of $L^2$-norms of the functions $g_\alpha$ and $g'_\alpha$.

Hence, use the assertion that the test function $\mathcal{D}(\Omega)$ are dense in $L^2(\Omega)$ to approximate each of the functions $g_\alpha$ and $g'_\alpha$, and to deduce that $\mathcal{D}(\Omega)$ is dense in the dual space $(H^s_0(\Omega))'$. This proves that indeed, the dual space of $H^s_0(\Omega)$ is the space $H^{-s}_0(\Omega)$.

Finally, to prove the last part (d), note that Proposition 6.8 shows that if $s \geq 0$ then $H^{-s}(\Omega) = H^s_0(\Omega)$, which implies that the dual of the space $H^{-s}(\Omega)$ is dual of the space $H^s_0(\Omega)$. However, the dual of the space $H^{-s}_0(\Omega)$ is $H^s_0(\Omega)$. In view of part (b), this means that $H^s_{\Omega}(\mathbb{R}^d) = H^s_0(\Omega)$.

6.1.3 Trace Operator

For a given domain $\Omega$, our interest is now on the restriction on the boundary $\partial \Omega$ of functions in $H^s(\Omega)$. This includes a suitable definition of the spaces on the boundary, namely, $H^s(\partial \Omega)$. A minimum of regularity of the domain $\Omega$ is necessary. For instance, the space $L^2(\partial \Omega)$ could be define using the $(d-$
1)-dimensional Hausdorff measure, but the Lebesgue surface measure gives a proper interpretation. This requires the normal direction to be defined almost everywhere, i.e., a Lipschitz domain Ω.

It is convenient to rephrase Definition 4.4 on smooth domain of class $C^{m,\alpha}$, for a nonnegative integer $m$ and $0 \leq \alpha \leq 1$ by saying that for every point $y$ on the boundary $\partial \Omega$ there exists $r = r(y) > 0$, an orthogonal system of coordinates $(x', x_d)$, $x' = (x_1, \ldots, x_{d-1})$ and a $C^{m,\alpha}$ function $\psi$ of $x'$ such that

$$\left\{ x \in \Omega : |x - y| < r \right\} = \{ x \in \mathbb{R}^d : |x - y| < r, x_d > \phi(x') \}.$$ 

This means that if the boundary $\partial \Omega$ is bounded then there exists a smooth partition of the unity $\chi_1, \ldots, \chi_n$, $\chi_k$ with support in $O_k$, $\partial \Omega \subset \bigcup_k O_k$, and local hypographs $\phi_1, \ldots, \phi_n$ of class $C^{m,\alpha}$ such that $O_k \cap \Omega = \{ x = (x', x_d) \in O_k : x_d > \phi_k(x') \}$, under a suitable orthogonal system of coordinates, $k = 1, \ldots, n$. The outward unit normal to $\partial \Omega$ at the point $x = (y', \phi_k(y'))$ is (locally) defined by

$$n(y', \phi_k(y')) = \frac{\left( \partial_1 \phi_k(y'), \ldots, \partial_{d-1} \phi_k(y'), -1 \right)}{\left[ 1 + (\partial_1 \phi_k(y'))^2 + \cdots + (\partial_{d-1} \phi_k(y'))^2 \right]^{1/2}},$$

which is a function of class $C^{m-1,\alpha}$. Certainly, this yields $d - 1$ independent tangential unit vectors $t_i$, for $i = 1, \ldots, d - 1$, e.g.,

$$t_1(y', \phi_k(y')) = \frac{(1, 0, \ldots, 0, \partial_1 \phi_k(y'))}{\left[ 1 + (\partial_1 \phi_k(y'))^2 \right]^{1/2}},$$

which are orthogonal to $n$ as expected. The surface Lebesgue measure $\ell_{d-1}(dx)$ or $d\sigma$ on the boundary $\partial \Omega$ is locally defined as

$$d\sigma = \sum_{k=1}^{n} \left[ 1 + (\partial_1 \phi_k(y'))^2 + \cdots + (\partial_{d-1} \phi_k(y'))^2 \right]^{1/2} \chi_k(y', \phi_k(y')) dy'.$$

Certainly, these definitions are independent of the particular smooth partition and hypogras used. If the boundary $\partial \Omega$ is not bounded then some extra conditions are involved, but we are not concerned with this situation.

In a Lipschitz domain, the hypogras are of class $C^{0,1}$ and the outward unit normal is defined almost everywhere with respect to the surface measure. The mapping $y' \mapsto x = (y', \phi_k(y'))$ is an homeomorphism of class $C^{m,\alpha}$ from an open subset of $\mathbb{R}^{d-1}$ onto $O_k \cap \partial \Omega$. The method of local coordinates consists in transporting whatever elements are studied or defined in $\partial \mathbb{R}^{d-1} \simeq \mathbb{R}^{d-1}$ into elements on the boundary $\partial \Omega$, e.g., the typical examples are the differentiation operations and the integration with respect to the surface Lebesgue measure as above, where a weighted Lebesgue measure in $\mathbb{R}^{d-1}$ is transported into the boundary $\partial \Omega$. This means that typically, a smooth domain is locally expressible as $\Omega_{\phi} = \{ x = (x', x_d) \in \mathbb{R}^d : x' \in \mathbb{R}^{d-1}, x_d > \phi(x') \}$, where $\phi$ is a real-valued function of class $C^{m,\alpha}(\mathbb{R}^{d-1})$ uniformly, i.e., $\phi$ of class $C^{m,\alpha}(\mathbb{R}^{d-1} \cup \{ \infty \})$. Therefore, it is convenient to write $\mathbb{R}^d = \{ (x', x_d) : x' \in \mathbb{R}^{d-1}, x_d \in \mathbb{R} \}$, $\mathbb{R}^d_+ = \{ x = (x', x_d) \in \mathbb{R}^d : x_d > 0 \}$ and $\partial \mathbb{R}^d_+ = \{ (x', 0) : x' \in \mathbb{R}^{d-1} \}$, i.e., the hyperplane $\{ x_d = 0 \}$. 

Proposition 6.10. If \( \delta \) is the Dirac measure at the origin in \( \mathbb{R} \), \( \partial_k \delta \) denotes the \( k \)-derivative, and \( g_0, g_1, \ldots \) are non-zero tempered distributions in \( \mathcal{S}'(\mathbb{R}^{d-1}) \) then the tensor-product distribution \( g_k \otimes \partial^k \delta \) belongs to \( H^s(\mathbb{R}^d) \) if and only if \( s + k < -1/2 \) and \( g_k \) belongs to \( H^{s+k+1/2}(\mathbb{R}^{d-1}) \). Moreover, a tempered distribution \( f \) with support in \( \partial \mathbb{R}^d_+ \) belongs to \( H^s(\mathbb{R}^d) \) if and only if \( f \) has the form \( f = \sum_{0 \leq k < -s - 1/2} g_k \otimes \partial^k \delta \). In particular, for any \( s \geq -1/2 \), only the zero element \( f = 0 \) in \( H^s(\mathbb{R}^d) \) could have support in \( \partial \mathbb{R}^d_+ \), and by duality, the test functions vanishing in a neighborhood of the hyperplane \( \partial \mathbb{R}^d_+ \) are dense in \( H^s(\mathbb{R}^d) \) if \( s \leq 1/2 \).

Proof. For the first part, note that the Fourier transform
\[
\hat{f}[g_k \otimes \partial^k \delta](\xi', \xi_d) = \hat{g}_k(\xi')(2\pi i \xi_d)^k.
\]
Thus, the change of variable \( \xi_d = (1 + |\xi'|^2)^{1/2} \lambda \) and the equality
\[
(1 + |\xi'|^2 + \xi_d^2)^{s/2} = (1 + |\xi'|^2)^{s/2}(1 + \lambda^2)^{s/2}
\]
show that
\[
\|g_k \otimes \partial^k \delta\|_{H^s(\mathbb{R}^d)}^2 = (2\pi)^k \int d\xi_d \int_{\mathbb{R}^{d-1}} |\hat{g}_k(\xi')(\xi_d)|^2 (1 + |\xi'|^2 + \xi_d^2)^{s/2} d\xi' = (2\pi)^k \int_R (1 + \lambda^2)^{s/2}\lambda^kd\lambda \int_{\mathbb{R}^{d-1}} |\hat{g}_k(\xi')(\xi_d)|^2 (1 + |\xi'|^2)^{s+k+1/2} d\xi',
\]
meaning
\[
\begin{align*}
\|g_k \otimes \partial^k \delta\|_{H^s(\mathbb{R}^d)} &= C_k \|g_k\|_{H^{s+k+1/2}(\mathbb{R}^{d-1})}, \\
\text{with} \quad C_k^2 &= (2\pi)^k \int_R (1 + \lambda^2)^{s/2}\lambda^kd\lambda,
\end{align*}
\]
which proves the first assertion.

Since \( H^s(\mathbb{R}^d) \) can only contain tempered distribution of finite order \( n \leq [-s] \), in view of Remark 3.32 and Proposition 3.31, any element \( f \) in \( H^s(\mathbb{R}^d) \) has the form \( f = \sum_{0 \leq k < n} g_k \otimes \partial^k \delta \) for some tempered distributions \( g_k \) in \( \mathcal{S}'(\mathbb{R}^{d-1}) \). Hence, the equality (6.16) proves the representation formula for elements in \( H^s(\mathbb{R}^d) \) with support in the hyperplane \( \partial \mathbb{R}^d_+ \).

If \( s \geq -1/2 \) then there is no \( k \) such that \( 0 \leq k < -s - 1/2 \), i.e., the representation means that \( f = 0 \) if \( f \) belongs \( H^s(\mathbb{R}^d) \) and has support in \( \partial \mathbb{R}^d_+ \). The duality argument regarding the density goes as follows.

Consider the bipolar space, i.e., the space \( X \) of all elements \( f \) in \( H^{-s}(\mathbb{R}^d) \), \( s \geq 1/2 \) such that \( \langle f, \phi \rangle = 0 \) for every test function \( \phi \) vanishing near the hyperplane \( \partial \mathbb{R}^d_+ \). The assertion translates into proving that \( X \) is actually the null space. However, a tempered distribution \( f \) belonging to \( X \) has necessarily support in \( \partial \mathbb{R}^d_+ \), and hence, \( f \) is zero.

Now, let us look at the restriction or trace operator
\[
\varphi \mapsto \varphi|_0 = \varphi|_{\partial \mathbb{R}^d_+} = \varphi|_{\mathbb{R}^{d-1}}, \quad \text{from} \quad \mathcal{S}(\mathbb{R}^d) \quad \text{onto} \quad \mathcal{S}(\mathbb{R}^{d-1}),
\]
i.e., \( \varphi(x', x_d) \mapsto \varphi(x', 0), (x', x_d) \) in \( \mathbb{R}^d, x' \) in \( \mathbb{R}^{d-1} \).
Theorem 6.11. If $s > 1/2$ then the trace operator $|_{\mathbb{R}^{d-1}}$ given by (6.17) can be extended as a continuous linear operator from $H^s(\mathbb{R}^d)$ onto $H^{s-1/2}(\mathbb{R}^{d-1})$. Moreover, if $m + 1/2 < s < m + 3/2$ with a positive integer $m$ then the normal derivatives and trace operator

$$f \mapsto T(f) = (f|_{\mathbb{R}^{d-1}}, (\partial_d f)|_{\mathbb{R}^{d-1}}, \ldots, (\partial_d^m f)|_{\mathbb{R}^{d-1}})$$

from $H^s(\mathbb{R}^d)$ onto $H^{s-1/2}(\mathbb{R}^{d-1}) \times H^{s-1/2}(\mathbb{R}^{d-1}) \times \ldots \times H^{s-1/2}(\mathbb{R}^{d-1})$, is also a continuous linear operator, and with the same meaning, the space $\mathbb{R}^d$ and $\mathbb{R}^{d-1}$ could be replaced by $\mathbb{R}^d_+$ and $\partial \mathbb{R}^d$, respectively. Furthermore, if $s < 1/2$ then the test functions $D(\mathbb{R}^d_+)$ are dense in $H^s(\mathbb{R}^d_+)$, i.e., $H^s(\mathbb{R}^d_+) = H^s_0(\mathbb{R}^d_+)$.

Proof. First, consider the trace over the hyperplane $\{(x', x_d) \in \mathbb{R}^d : x_d = r\}$, which is identified with the space $\mathbb{R}^{d-1}$. Thus, for any real number $r$ and any test function $\varphi$ in $\mathbb{R}^d$, the Fourier inversion formula yields

$$\varphi|_r(x') = \varphi(x', r) = \int_{\mathbb{R}^{d-1}} e^{2\pi i x' \cdot \xi} \hat{\varphi}|_r(\xi')d\xi',$$

with $\hat{\varphi}|_r(\xi') = \int_{\mathbb{R}} e^{2\pi ir \xi_d} \hat{\varphi}(\xi', \xi_d)d\xi_d$.

Now, if $s > 1/2$ then the change of variable $\xi_d = (1 + |\xi'|^2)^{1/2} \lambda$ shows that

$$\int_{\mathbb{R}} (1 + |\xi'|^2 + \xi_d^2)^{-s}d\xi_d = (1 + |\xi'|^2)^{-s+1/2} \int_{\mathbb{R}} (1 + \lambda^2)^{-s}d\lambda$$

and Cauchy-Schwarz inequality yields

$$|\hat{\varphi}|_0(\xi')|^2 \leq \left( \int_{\mathbb{R}} (1 + |\xi'|^2 + \xi_d^2)^{-s}d\xi_d \right) \left( \int_{\mathbb{R}} (1 + |\xi'|^2 + \xi_d^2)^s |\hat{\varphi}(\xi', \xi_d)|^2d\xi_d \right).$$

Hence, integrate

$$(1 + |\xi'|^2)^{s-1/2} |\hat{\varphi}|_0(\xi')|^2 \leq C_s^2 \left( \int_{\mathbb{R}} (1 + |\xi'|^2 + \xi_d^2)^s |\hat{\varphi}(\xi', \xi_d)|^2d\xi_d \right),$$

with $C_s^2 = \int_{\mathbb{R}} (1 + \lambda^2)^{-s}d\lambda$

to deduce

$$\|\hat{\varphi}|_r\|_{H^{s-1/2}(\mathbb{R}^{d-1})} \leq C_s \|\hat{\varphi}\|_{H^s(\mathbb{R}^d)}, \quad \forall \varphi \in D(\mathbb{R}^d),$$

which proves that the restriction (or trace) operator $(r, \varphi) \mapsto \varphi|_r$ can be considered as a continuous linear operator from $\mathbb{R} \times H^s(\mathbb{R}^d)$ into $H^{s-1/2}(\mathbb{R}^{d-1})$, $s > 1/2$, initially defined on $D(\mathbb{R}^d)$ and continuously extended by density. Of particular interest is the case $r = 0$, i.e., the trace $\varphi \mapsto \varphi|_0$ on $x_d = 0$. 

Alternatively, for a real number \( r \), if \( \delta_r \) is the Dirac measure concentrated at \( x_d = r \) and \( g \) is a distribution in \( H^{-s+1/2}(\mathbb{R}^{d-1}) \), \( s > 1/2 \), then consider the distribution \( g \otimes \delta_r \) defined as the tensor product

\[
\langle g \otimes \delta_r, \varphi \rangle = \langle g, \varphi(\cdot, r) \rangle.
\]

Calculate

\[
\widehat{g \otimes \delta_r}(\xi', \xi_d) = \hat{g}(\xi')e^{-2\pi i r \xi_d},
\]

\[
(1 + |\xi'|^2 + \xi_d^2)^{-s}|g \otimes \delta_r(\xi', \xi_d)|^2 = (1 + |\xi'|^2 + \xi_d^2)^{-s}|\hat{g}(\xi')|^2
\]
to check that the change of variable \( \xi_d = (1 + |\xi'|^2)^{1/2}\lambda \) shows

\[
\int_{\mathbb{R}^d} (1 + |\xi'|^2 + \xi_d^2)^{-s}|g \otimes \delta_r(\xi', \xi_d)|^2 d\xi =
\int_{\mathbb{R}} (1 + \lambda^2)^{-s}d\lambda \int_{\mathbb{R}^{d-1}} (1 + |\xi'|^2)^{-s+1/2}|\hat{g}(\xi')|^2 d\xi',
\]

i.e.,

\[
\|g \otimes \delta_r\|_{H^{-s}(\mathbb{R}^d)} = C_s \|g\|_{H^{-s+1/2}(\mathbb{R}^{d-1})}, \quad \text{with} \quad C_s^2 = \int_{\mathbb{R}} (1 + \lambda^2)^{-s}d\lambda.
\]

This proves that the \( \delta \)-lift \( g \mapsto g \otimes \delta_r \) is a continuous linear operator from \( H^{-s+1/2}(\mathbb{R}^{d-1}) \) into \( H^{-s}(\mathbb{R}^d) \). Hence, by transposition, the paring

\[
\langle f, g \otimes \delta_r \rangle = \langle f \mid_r, g \rangle, \quad \forall f \in H^s(\mathbb{R}^d), g \in H^{-s+1/2}(\mathbb{R}^{d-1}),
\]

shows afresh the continuity of the restriction operator \( \mid_r \). Conversely, if a priori, the restriction \( (\cdot) \mid_r \) is a continuous linear operator then, by transposition, the \( \delta \)-lift \( (\cdot) \otimes \delta_r \) is also a continuous linear operator.

To check that the restriction operator \( \mid_r \) is surjective, define the lift mapping for any \( f = \Upsilon_{r, \eta}g \) in \( H^{s-1/2}(\mathbb{R}^{d-1}) \) via the Fourier transform as

\[
\hat{f}(\xi', \xi_d) = \hat{g}(\xi')(1 + |\xi'|^2)^{-1/2}\eta(\xi_d(1 + |\xi'|^2)^{-1/2})e^{2\pi i r \xi_d},
\]

with \( \eta \in \mathcal{D}(\mathbb{R}) \) such that \( \int_{\mathbb{R}} \eta(\lambda)d\lambda = 1. \)

The Fourier inversion and the change of variable \( \xi_d = (1 + |\xi'|^2)^{1/2}\lambda \) shows that

\[
[\Upsilon_{r, \eta}g](x', r) = [\Upsilon_{0, \eta}g](x', 0) = \int_{\mathbb{R}} \hat{\Upsilon}_{0, \eta}g(\xi', \xi_d)d\xi_d = \hat{g}(\xi') \int_{\mathbb{R}} \eta(\lambda)d\lambda
\]

and

\[
\int_{\mathbb{R}^d} (1 + |\xi|^2)^s|\hat{f}(\xi', \xi_d)|^2 d\xi = C_{\eta, s}^2 \int_{\mathbb{R}^{d-1}} (1 + |\xi'|^2)^{s-1/2}|\hat{g}(\xi')|^2 d\xi',
\]

with \( C_{\eta, s}^2 = \int_{\mathbb{R}} (1 + \lambda^2)^s \eta(\lambda)d\lambda. \)
i.e., $\Upsilon_{r,\eta}g|_r = g$ and $\|\Upsilon_{r,\eta}g\|_{H^s(\mathbb{R}^d)} = C_{\eta,s}\|g\|_{H^{s-1/2}(\mathbb{R}^{d-1})}$, which proves that the lift $\Upsilon_{r,\eta}g$ is a continuous linear operator from $H^{s-1/2}(\mathbb{R}^{d-1})$ into $H^s(\mathbb{R}^d)$ (regardless of the assumption $s > 1/2$), and the restriction $|_r$ is surjective. In contract, note that $\delta$-lift $g \mapsto g \otimes \delta_r$ is a continuous linear operator from $H^{-s+1/2}(\mathbb{R}^{d-1})$, $s > 1/2$, into $H^{-s}(\mathbb{R}^d)$.

If $f$ is a function in $H^s(\mathbb{R}^d)$, $s > 1/2$ and $\phi(x',x_d)$ is a test function in $\mathbb{R}^d$ then the expression

$$
\langle T_f, \phi \rangle = \int_\mathbb{R} \langle f|_r, \phi(\cdot, r) \rangle dr
$$

defines a distribution $T_f$ on $\mathbb{R}^d$ which is identified with $f$. Indeed, this holds true if $f$ is a test function, and by density and continuity, this remains valid. In particular, this implies that if $f = 0$ almost everywhere on the region $\{(x', x_d) \in \mathbb{R}^d : 0 < x_d < \varepsilon\}$ then $f|_r = 0$ as a distribution for any $0 < r < \varepsilon$, and the continuity in $r$ ensures that $f|_r = 0$ is also true for $r = 0$ and $r = \varepsilon$. Thus, if $f$ and $g$ are two elements in $H^s(\mathbb{R}^d)$, $s > 1/2$ such that $f = g$ almost everywhere in $\mathbb{R}^{d-1} \times (r, \varepsilon)$ for some $\varepsilon > 0$ then $f|_r = g|_r$.

Therefore, if $\mathbb{R}_{d+}^d = \{(x', x_d) \in \mathbb{R}^d : x_d = 0\}$ with the boundary $\partial\mathbb{R}_{d+}^d = \mathbb{R}^{d-1} \times \{0\}$ identified to $\mathbb{R}^{d-1}$, $s > 1/2$, and $f$ belongs to $H^s(\mathbb{R}_{d+}^d)$, then the trace on boundary $\partial\mathbb{R}_{d+}^d$ of $f$ can be defined as $f^e|_o$, where $f^e$ in $H^s(\mathbb{R}^d)$ is any extension of $f$ and the definition does not depends on the choice of the extension $f^e$. This trace operator, denoted by either $f^e|_o$ or $f^e|_{\mathbb{R}^{d-1}}$, is continuous and linear from $H^s(\mathbb{R}_{d+}^d)$ onto $H^{s-1/2}(\mathbb{R}^{d-1})$. Similarly, the lift operator $\Upsilon_{r,\eta}$ is continuous and linear from $H^{s-1/2}(\mathbb{R}^{d-1})$ into $H^s(\mathbb{R}_{d+}^d)$.

Composing with the derivative operator $\partial^\alpha$, if $s > |\alpha| + 1/2$ then the restriction of the partial derivatives $f \mapsto (\partial^\alpha f)|_r$ is also a continuous and linear operator from $H^s(\mathbb{R}^d)$ (or $H^s(\mathbb{R}_{d+}^d)$ for $r = 0$) into $H^{s-1-2|\alpha|}(\mathbb{R}^{d-1})$. Of particular interest are the normal derivatives, i.e., the operator $f \mapsto (\partial^j_d f)|_{\mathbb{R}^{d-1}}$ for $s > k + 1/2$. Therefore, the normal derivative and

\[
\begin{aligned}
\text{Trace Operator} & \quad \text{from} \quad H^s(\mathbb{R}_{d+}^d), \quad m + 1/2 < s < m + 3/2, \quad \text{into} \\
H^{s-1/2}(\mathbb{R}^{d-1}) \times H^{s-1-1/2}(\mathbb{R}^{d-1}) \times \cdots \times H^{s-1-m}(\mathbb{R}^{d-1}),
\end{aligned}
\]

is defined by $T(f) = (f|_{\partial^q_+^d}, (\partial_d f)|_{\partial^q_+^d}, \ldots, (\partial_d^m f)|_{\partial^q_+^d})$. Initially, this makes sense for smooth functions in $S(\mathbb{R}^d)$ and then the definition is extended by continuity.

To verify that $T$ is a surjective operator, note that if $\varphi$ belongs to $S(\mathbb{R}^d)$ then

$$
\mathfrak{F}[(\partial^j_d \varphi)|_{\mathbb{R}^{d-1}}](\xi') = \int_\mathbb{R} (2\pi i \xi_d)^k \hat{\varphi}(\xi', \xi_d) d\xi_d.
$$

Thus, again, by means of the Fourier transform define the lift operators $\Upsilon_{\eta_0} g$, $\Upsilon_{\eta_1} g$, $\ldots$, $\Upsilon_{\eta_d} g$, where

\[
\begin{aligned}
\Upsilon_{\eta_k} g &= \mathfrak{F}^{-1}[\mathfrak{F}[(\partial^k_d \varphi)|_{\mathbb{R}^{d-1}}](\xi')(1 + |\xi'|^2)^{-1/2} \eta(\xi_d(1 + |\xi'|^2)^{-1/2})], \\
\eta_k &\in D(\mathbb{R}) \quad \text{such that} \quad \int_\mathbb{R} (2\pi i \lambda)^j \eta_k(\lambda)d\lambda = \delta_{jk}, \quad \forall j, k = 0, 1, \ldots, m,
\end{aligned}
\]
with $\delta_{kk} = 1$ and $\delta_{jk} = 0$ if $j \neq k$. This construction yields
\[
(\partial_k \Upsilon_{\eta_k} g)|_{x^d} = g \quad \text{and} \quad (\partial_j \Upsilon_{\eta_k} g)|_{x^d} = 0, \quad \text{if} \quad j \neq k,
\]
i.e., given any distributions $g_k$ in $H^{s-1/2-k}(\mathbb{R}^{d-1})$, for $k = 0, 1, \ldots, m$, the distribution $f = \sum_{k=0}^m \Upsilon_{\eta_k} g_k$ satisfies $T(f) = (g_0, g_1, \ldots, g_m)$. Note that as expected, the assumption $s > m + 1/2$ is not needed for the lift operators.

Combining with Proposition 6.10, the equality $H^s(\mathbb{R}^d_+) = H^s_0(\mathbb{R}^d_+)$ holds true if $s \leq 1/2$. Moreover, the trace operator $T$ (as defined by (6.18) with $s > 1/2$) is surjective and its kernel $\{f \in H^s(\mathbb{R}^d_+) : T(f) = 0\}$ is actually the space $H^s_0(\mathbb{R}^d_+)$. \hfill \Box

\textbf{Remark 6.12.} Essentially the same technique used in Theorem 6.11 can be adapted to define the trace operator on $\mathbb{R}^n$, with any dimension $n < d$, i.e., the trace operator $|_{\mathbb{R}^n}$ is a continuous linear operator from the space $H^s(\mathbb{R}^d_+)$ onto $H^{s-(d-n)/2}(\mathbb{R}^n)$, if $s > (d-n)/2$. \hfill \Box

Now, by means of local coordinates (or charts), the results of Theorem 6.11 are valid for smooth domain.

\textbf{Definition 6.13.} Let $\Omega$ be a smooth domain in $\mathbb{R}^d$ of class $C^{m,1}$, $m \geq 0$, with a bounded boundary and hypographs $\phi_1, \ldots, \phi_n$ and associated smooth partition of the unity $\chi_1, \ldots, \chi_n$. If $0 \leq s \leq 1 + m$ then a function $f$ belongs to the Sobolev space $H^s(\partial\Omega)$ if and only if the function $f_k : y' \mapsto f(y', \phi_k(y'))\chi_k(y', \phi_k(y'))$ is in $H^s(\mathbb{R}^{d-1})$, for every $k$. The hilbertian norm is defined through the expression
\[
\|f\|^2_{H^s(\partial\Omega)} = \sum_k \|f_k\|^2_{H^s(\mathbb{R}^{d-1})}.
\]

It should be clear that $H^s(\partial\Omega)$ is a Hilbert space, that $L^2(\partial\Omega) \simeq H^0(\partial\Omega)$, that $H^t(\partial\Omega) \subset H^s(\partial\Omega)$ for any $s \leq t$, and that its definition is independent of the particular hypographs and smooth partition of the unity used. Moreover, based on the previous results, if $s > 1/2$ then the trace operator $f \mapsto f|_{\partial\Omega}$, initially defined on smooth functions, can be extended to a continuous linear operator from $H^s(\Omega)$ onto $H^{s-1/2}(\partial\Omega)$, i.e., Theorem 6.11 holds true with $\mathbb{R}^d$ and $\partial\mathbb{R}^d_+$ replaced with $\Omega$ and $\partial\Omega$.

\section{6.2 Riesz and Bessel Potentials}

The negative Laplacian operator $-\Delta = -\sum_{i=1}^d \partial_i^2$ and its variant $I - \Delta$, with $I = 1$ the identity, yield typical elliptic partial differential equations (PDE), namely $-\Delta u = f$ and $u - \Delta u = f$. The Fourier transform in $\mathbb{R}^d$ allows to consider fractional power of them, i.e.,
\[
\tilde{\mathcal{F}}((-\Delta)^{\alpha/2} u)(\xi) = (2\pi|\xi|)^\alpha \tilde{\mathcal{F}}(u)(\xi) \quad \text{and} \quad \tilde{\mathcal{F}}((I - \Delta)^{\beta/2} u)(\xi) = (1 + 4\pi^2|\xi|^2)^{\beta/2} \tilde{\mathcal{F}}(u)(\xi),
\]
for any real numbers $\alpha$ and $\beta$. Certainly, $\alpha = \beta = -1$ means to solve the PDE or find the inverse operator. Because the multiplication by the function
\((\lambda + 4\pi^2|\xi|^2)^{\beta/2}\), with \(\lambda > 0\) is an homeomorphism from either the rapidly
decreasing smooth functions \(S(\mathbb{R}^d)\) or the tempered distributions \(S'(\mathbb{R}^d)\) onto itself, the expression

\[
((\lambda I - \Delta)^{\beta/2} u)(x) = \mathfrak{F}^{-1}\left( (\lambda + 4\pi^2|\xi|^2)^{\beta/2} \mathfrak{F}(u)(\xi) \right)(x)
\]
defines an homeomorphism from either \(S(\mathbb{R}^d)\) or \(S'(\mathbb{R}^d)\) onto itself. Certainly, the identities \((\lambda I - \Delta)^{\alpha/2}(\lambda I - \Delta)^{\beta/2} = (\lambda I - \Delta)^{(\alpha+\beta)/2}\) holds in either \(S(\mathbb{R}^d)\) or \(S'(\mathbb{R}^d)\), for any real number \(\alpha\) and \(\beta\). If the exponent \(\beta\) is such that the function \((\lambda + 4\pi^2|\xi|^2)^{\beta/2}\) with \(\lambda = 0\) is locally integrable (i.e., \(\beta > -d\)), the above expression can be taken as \(\lambda \to 0\) in the space of tempered distributions \(S'(\mathbb{R}^d)\).

### 6.2.1 Initial Discussion

Our interest is the Riesz potential \((-\Delta)^{-s/2}\) with \(0 < s < d\), and the Bessel potential \((I - \Delta)^{-s/2}\) with \(s > 0\). The Fourier inverses of the above multipliers are called the Riesz and the Bessel kernels, and are given by as follows:

**Lemma 6.14.** If \(\Gamma\) denotes the Gamma function,

\[
g_{s,0}(x) = \gamma(s)|x|^{s-d}, \quad \text{with} \quad \gamma(s) = \pi^{d/2}2^s\Gamma(s/2)/\Gamma(d/2-s/2),
\]

for \(0 < s < d\), and

\[
g_{s,\lambda}(x) = \frac{1}{(4\pi)^{s/2}\Gamma(s/2)} \int_0^\infty e^{-\lambda r/(4\pi)}e^{-\pi|x|^2/r}\,dr,
\]

for \(s > 0\), then

\[
g_{s,\lambda'}(x) \leq g_{s,\lambda}(x) \rightarrow g_{s,0}(x) \quad \text{as} \quad \lambda' \geq \lambda \to 0, \quad \forall x \in \mathbb{R}^d \setminus \{0\},
\]

and

\[
\mathfrak{F}^{-1}\left((\lambda + 4\pi^2|\xi|^2)^{-s/2}\right) = g_{s,\lambda}(x), \quad \forall \lambda \geq 0.
\]

Moreover the function \(g_{s,\lambda}\) belongs to \(L^\infty(\mathbb{R}^d) \cap C^0(\mathbb{R}^d)\) if \(s > d\) and to \(L^p(\mathbb{R}^d)\), for every \(1 \leq p < d/(d-s)\), and also \(\|g_{s,\lambda}\|_1 = \lambda^{-s/2}\).

**Proof.** It is clear that in the definition of \(g_{s,\lambda}(x)\) the integrand is integrable for any \(s, \lambda > 0\) and \(x \in \mathbb{R}^d \setminus \{0\}\). Moreover, \(\lambda' \geq \lambda > 0\) implies \(g_{s,\lambda'}(x) \leq g_{s,\lambda}(x)\), for every \(x \in \mathbb{R}^d \setminus \{0\}\) and \(s > 0\). Furthermore, the change of variable \(\pi|x|^2/r = t\) yields

\[
g_{s,\lambda}(x) = \frac{1}{2^{s/2}\pi^{d/2}\Gamma(s/2)}|x|^{s-d} \int_0^\infty e^{-\pi|x|^2/(4t)}e^{-t(s-d)/2-1}\,dt,
\]

and the convergence as \(\lambda \to 0\) follows for any \(0 < s < d\) and \(x \in \mathbb{R}^d \setminus \{0\}\).
Now, for any $a, s > 0$ the change of variable $t = ar$ in the expression
\[
\Gamma(s/2) = \int_0^\infty e^{-t} t^{s/2-1} dt
\]
of the Gamma function yields the equality
\[
\frac{a^{s/2}}{\Gamma(s/2)} \int_0^\infty e^{-ar} r^{s/2-1} dr = 1,
\]
and in particular, the choice $a = \lambda + 4\pi^2 |\xi|^2$ implies
\[
\Gamma(s/2)(\lambda + 4\pi^2 |\xi|^2)^{-s/2} = \int_0^\infty e^{-(\lambda+4\pi^2 |\xi|^2)r} r^{s/2-1} dr.
\]
Hence, calculate the inverse Fourier transform
\[
\mathfrak{F}^{-1}(\Gamma(s/2)(\lambda + 4\pi^2 |\xi|^2)^{-s/2}) = \int_{\mathbb{R}^d} e^{2\pi i x \cdot \xi} \Gamma(s/2)(\lambda + 4\pi^2 |\xi|^2)^{-s/2} d\xi = \int_{\mathbb{R}^d} e^{2\pi i x \cdot \xi} \int_0^\infty e^{-(\lambda+4\pi^2 |\xi|^2)r} r^{s/2-1} dr d\xi = \int_0^\infty e^{-\lambda r} r^{s/2-1} dr \int_{\mathbb{R}^d} e^{2\pi i x \cdot \xi} e^{-4\pi^2 |\xi|^2 r} d\xi,
\]
and use the equality
\[
\mathfrak{F}(\alpha^{d/2} e^{-\pi |x|^2/\alpha}) = e^{-\alpha \pi |\xi|^2}, \quad \alpha > 0
\]
with $\alpha = 1/(4\pi r)$, to deduce
\[
\mathfrak{F}^{-1}(\Gamma(s/2)(\lambda + 4\pi^2 |\xi|^2)^{-s/2}) = \int_0^\infty e^{-\lambda r/(4\pi)} r^{-d/2} r^{s/2-1} e^{-|x|^2/(4r)} dr = (4\pi)^{-s/2} \int_0^\infty e^{-\lambda r/(4\pi)} e^{-\pi |x|^2/r} r^{(s-d)/2-1} dr.
\]
Moreover, if $\lambda > 0$ (in particular $\lambda = 1$) then exchange the order of integrals,
\[
\int_{\mathbb{R}^d} dx \int_0^\infty e^{-\lambda r/(4\pi)} e^{-\pi |x|^2/r} r^{(s-d)/2-1} dr = \int_0^\infty e^{-\lambda r/(4\pi)} r^{(s-d)/2-1} dr \int_{\mathbb{R}^d} e^{-\pi |x|^2/r} dx = \int_0^\infty e^{-\lambda r/(4\pi)} r^{(s-d)/2-1} dr \int_{\mathbb{R}^d} e^{-\pi |x|^2/r} dx
\]
and implement the change of variables $r = 4\pi t/\lambda$ to obtain
\[
\int_{\mathbb{R}^d} dx \int_0^\infty e^{-\lambda r/(4\pi)} e^{-\pi |x|^2/r} r^{(s-d)/2-1} dr = \int_0^\infty e^{-\lambda r/(4\pi)} r^{s/2-1} dr = (4\pi/\lambda)^{s/2} \int_0^\infty e^{-t} t^{s/2-1} dt = (4\pi)^{s/2} \Gamma(s/2) \lambda^{-s/2}.
\]
On the other hand, if $\lambda = 0$ then the change of variable $\pi|x|^2/r = t$ yields
\[
\mathcal{F}^{-1}(\Gamma(s/2)(4\pi^2|\xi|^2)^{s/2}) = \int_0^\infty (4\pi r)^{-d/2}e^{-r (s-d)/2-1}e^{-|x|^2/(4r)}dr = \\
= (4\pi)^{-s/2} \int_0^\infty e^{-\pi|x|^2/r}r^{(s-d)/2-1}dr = \\
= (4\pi)^{-s/2} (\pi|x|^2)^{(s-d)/2} \int_0^\infty e^{-t(d-s)/2-1}dt = \\
= 2^{-s}\pi^{-d/2}|x|^{s-d}\Gamma((d-s)/2).
\]
The argument is completed by collection all pieces,
\[
\|g_{s,\lambda}\|_1 = \int_{\mathbb{R}^d} g_{s,\lambda}(x)dx = \lambda^{-s/2}, \quad \forall s, \lambda > 0,
\]
and the Fourier inverse transforms are established.

It is clear that $g_{s,\lambda}$ belongs to $L^\infty(\mathbb{R}^d)$ if $s > d$. Now, for $0 < s \leq d$, regarding the integral
\[
\int_0^\infty e^{-\lambda r/(4\pi)}e^{-\pi|x|^2/r}r^{(s-d)/2-1}dr
\]
as a function in $x$, Minkowski inequality for integrals (see Remark B.59) yields
\[
\left\| \int_0^\infty e^{-\lambda r/(4\pi)}e^{-\pi|x|^2/r}r^{(s-d)/2-1}dr \right\|_{L^p(\mathbb{R}^d)} \leq \\
\leq \int_0^\infty \left\| e^{-\lambda r/(4\pi)}e^{-\pi|x|^2/r}r^{(s-d)/2-1} \right\|_{L^p(\mathbb{R}^d)}dr = \\
= \int_0^\infty e^{-\lambda r/(4\pi)}r^{(s-d)/2-1} \left\| e^{-\pi|x|^2/r} \right\|_{L^p(\mathbb{R}^d)}dr
\]
which is equal to
\[
p^{-d/2} \int_0^\infty e^{-\lambda r/(4\pi)}r^{(s-d)/2+d/(2p)-1}dr = \\
= p^{-d/2} (4\pi/\lambda)^{(s-d)/2+d/(2p)}\Gamma((s-d)/2+d/(2p)),
\]
provided $(s-d)/2+d/(2p) > 0$, i.e., $p < d/(d-s)$ if $0 < s < d$ or $p < \infty$ if $s = d$. This complete the proof. \qed

Note that the Riesz and Bessel kernels, as well as their Fourier transforms are locally integrable functions with polynomial growth at infinity, and therefore, they can be interpreted as tempered distributions. Also, it may be convenient to include the dimension $d$, to set $g_{s,1} = b_{s,d}$ and to implement the change of variable $r = 4\pi t$ to obtain
\[
\begin{cases} 
    b_{s,d}(x) = \frac{1}{(4\pi)^{s/2}\Gamma(s/2)} \int_0^\infty e^{-r/(4\pi)}e^{-\pi|x|^2/r}r^{(s-d)/2-1}dr = \\
                   \frac{1}{(4\pi)^{d/2}\Gamma(s/2)} \int_0^\infty e^{-r}e^{-|x|^2/(4r)}r^{(s-d)/2-1}dr,
\end{cases}
\]
(6.19)
which is referred to as the Bessel kernel of order $s$ and dimension $d$, for any real number $s > 0$ and any non-zero $x$ in $\mathbb{R}^d$, even if our main interest is $0 < s < d$. Naturally, if the normalizing constants $\Gamma(s/2)$ is ignored then the integral expression is finite for any $s$ and $x \neq 0$.

Based on the above calculation, the Bessel potential is defined by

$$\left( (I - \Delta)^{-s/2} f \right)(x) = \int_{\mathbb{R}^d} f(y) b_s(x - y) dy, \quad s > 0,$$

and the Riesz potential is given by

$$\left( \Delta^{-s/2} f \right)(x) = \gamma(s, d) \int_{\mathbb{R}^d} f(y)|x - y|^{s - d} dy, \quad 0 < s < d,$$

with $\gamma(s, d) = \pi^{d/2} 2^s \Gamma(s/2)/\Gamma(1/2 - s/2)$.

**Remark 6.15.** In view of the identity $(I - \Delta)^{-s/2}(I - \Delta)^{-t/2} = (I - \Delta)^{-(s + t)/2}$ in $S(\mathbb{R}^d)$, the Fourier transform yields the convolution identity $b_{s,d} * b_{t,d} = b_{s+t,d}$, for any $s, t > 0$.

A priori, both potentials are kernel convolutions and defined on smooth functions, say $S(\mathbb{R}^d)$. In view of Lemma 6.14, $b_{s,d}$ belongs to $L^1(\mathbb{R}^d)$, and thus Young inequality (see Proposition B.65) implies that the Bessel potential maps $L^p(\mathbb{R}^d)$ into itself and

$$\| (I - \Delta)^{-s/2} f \|_p \leq \| b_{s,d} \|_1 \| f \|_p = \| f \|_p, \quad \forall f \in L^p(\mathbb{R}^d),$$

for any $1 \leq p \leq \infty$. Moreover, because $\| b_{s,d} \|_r < \infty$ for suitable $r \geq 1$, the Bessel operator $f \mapsto (I - \Delta)^{-s/2} f$ maps $L^p(\mathbb{R}^d)$ into $L^q(\mathbb{R}^d)$, for every $1 \leq p \leq q < pd/(d-s)$ if $0 < s \leq d$ or any $1 \leq p \leq q \leq \infty$ if $s > d$. Also note that from Lemma 6.14 follows that

$$\| (\lambda I - \Delta)^{-s/2} f \|_p \leq \lambda^{-s/2} \| f \|_p, \quad \forall f \in L^p(\mathbb{R}^d),$$

so that we cannot expect to take $\lambda \to 0$ to obtain a similar estimate for the Riesz potential.

If the Riesz kernel $k$ is split in two pieces, i.e., $k = k_1 + k_2$ with $k_1(x) = \gamma(s)|x|^{s-d} \mathbb{1}_{|x| \leq 1}$ and $k_2(x) = \gamma(s)|x|^{s-d} \mathbb{1}_{|x| > 1}$ then the Riesz potential can be expressed as the sum of two terms, i.e., $\left((-\Delta)^{-s/2} f \right) = k_1 * f + k_2 * f$, where each convolution is an absolutely convergent integral when $f$ belongs to $L^p(\mathbb{R}^d)$, $p > 1$. Indeed, $k_1$ belongs to $L^1$ so that $k_1 \ast f$ belongs to $L^p$, while $k_2$ belongs to $L^r(\mathbb{R}^d)$ if $(s-d)r < -d$, i.e., $r > d/(d-s)$ or $1/r < 1 - s/d$. Thus, for any $p > 1$ and $q = d/s$ there is $r > d/(d-s)$ such that $1/p + 1/r = 1 - 1/q = 1 - s/d$ and Young inequality implies that $k_2 \ast f$ belongs to $L^q(\mathbb{R}^d)$ with $q = s/d$. This also proves that the Riesz potential is continuous from $L^p(\mathbb{R}^d) \cap L^q(\mathbb{R}^d)$ into $L^q(\mathbb{R}^d)$ for any $p > 1$ and $q = d/s$. However, to shows that continuity from $L^p(\mathbb{R}^d)$ (with $1 < p < \infty$) into $L^q(\mathbb{R}^d)$ (with $1/q = 1/p - s/d$) of the Riesz potential, a
more detailed analysis is necessary, one proves that \( f \mapsto k \ast f \) is of weak type \((p,q)\), with \( 1 \leq p < q < \infty \), \( 1/q = 1/p - s/d \), and the estimate

\[
\|(-\Delta)^{-s/2} f\|_q \leq C_{p,q} \|f\|_p, \quad \forall f \in L^p(\mathbb{R}^d),
\]

follows from Marcinkiewicz interpolation inequality, e.g., see Hardy-Littlewood-Sobolev Theorem of fractional integration in Stein [113, Section V.1.2, Theorem 1, p. 119] or Grafakos [57, 58].

Identities like \((-\Delta)^{-s/2}(-\Delta)^{-t/2} = (-\Delta)^{-(s+t)/2}\) for any \( s,t > 0 \) with \( s + t < d \), and \( \Delta(-\Delta)^{-s/2} = (-\Delta)^{-s/2}\Delta = (-\Delta)^{-(s-2)/2}\), for any \( d \geq 3 \), \( 2 \leq s \leq d \), can be easily verified in the space of tempered distributions \( S'(\mathbb{R}^d) \).

A related hard point is to estimate the Riesz transforms

\[
\begin{aligned}
(R_j u)(x) &= c_d \lim_{\varepsilon \to 0} \int \frac{y_j}{|y|^{d+1}} u(x - y) dy, \quad j = 1, 2, \ldots, d, \\
c_d &= \pi^{-(d+1)/2} \Gamma(d/2 + 1/2),
\end{aligned}
\]

in \( L^p(\mathbb{R}^d) \), \( 1 \leq p < \infty \), which involves the so-called singular integrals. Actually, the Riesz transforms can be expressed in term of the Fourier transform as

\[
(R_j u)(x) = \hat{\mathfrak{s}}^{-1}\left(\frac{\xi_j}{|\xi|}\right)\mathfrak{s}(u)(\xi)(x),
\]

which make clear the relation \( \partial_{ij}u = -R_i R_j \Delta u \) with \( \partial_{ij} \) denoting the second partial derivative in the \( i \) and \( j \) variables. The \( L^p \)-bounds refers to the following key result (see Section 5.5 on Fourier Multipliers)

**Theorem 6.16.** Let \( k \) be a continuously differentiable function in \( \mathbb{R}^d \setminus \{0\} \) such that it belongs to \( L^2(\mathbb{R}^d) \), its Fourier transform \( \mathfrak{s}k \) is essentially bounded, and there exists a constant \( A_k > 0 \) satisfying

\[
\|\mathfrak{s}k\|_\infty \leq A_k \quad \text{and} \quad |\nabla k(x)| \leq A_k |x|^{-d-1}, \quad \forall x \neq 0.
\]

Then the convolution \( k \ast u \), which is defined for every \( u \) in \( L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d) \), can be extended by continuity to \( L^p(\mathbb{R}^d) \), i.e.,

\[
\|k \ast u\|_p \leq C_p \|u\|_p, \quad \forall u \in L^p(\mathbb{R}^d), \ 1 < p < \infty,
\]

where the constant \( C_p \) depends only on \( A_k \), \( p \) and the dimension \( d \). \( \square \)

An interesting point is that the arguments used in the proof can be extended to a measurable kernel \( k_\varepsilon(x) = \mathbb{1}_{\{|x| \geq \varepsilon\}} k(x) \), \( \varepsilon > 0 \), where \( k \) satisfies the cancelation property

\[
\int_{r < |x| < R} k(x) dx = 0, \quad \forall R > r > 0.
\]

and the bounds

\[
|k(x)| \leq A_k |x|^{-d} \quad \text{and} \quad \int_{|x| \geq 2|y|} |k(x + y) - k(x)| dx \leq A_k,
\]
for any \(x, y \in \mathbb{R}^d \setminus \{0\}\) and for some constant \(A_k > 0\). Moreover, if \(k(x) = x_j|x|^{-d-1}\) then the inequalities

\[
|k(x + y) - k(x)| \leq (d + 2)|y|(|x|^{-d-1} + |x + y|^{-d-1}),
\]

\[
\int_{|x| \geq r} \frac{1}{|x|^{d+1}} \, dx = \frac{2\pi^{d/2}}{r \Gamma(d/2)}, \quad \text{and} \quad |x + y| \geq |y| \quad \text{if} \quad |x| \geq 2|y|,
\]

and Theorem 6.16 yield the estimate

\[
\|R_j u\|_p \leq C_p \|u\|_p, \quad \forall u \in L^p(\mathbb{R}^d), \quad 1 < p < \infty, \quad j = 1, 2, \ldots, d,
\]

(6.23)

where the constant \(C_p\) depends only on \(p\) and the dimension \(d\). In particular, the equality \(\partial_{ij} u = -R_i R_j \Delta u\) shows that \(\|\partial_{ij} u\|_p \leq C_p \|\Delta u\|_p\), for any smooth function with compact support. For instance, the interested reader is referred to the detailed account in Stein [113, Chapters I–II, pp. 3–80].

### 6.2.2 Bessel Kernel and Potentials

Our interest is on the kernel \(b_{s,d}\) given by (6.20) and some properties of the potential \((I - \Delta)^{-s/2} f\) for functions in \(L^p(\mathbb{R}^d)\). Certainly, recalling that \(\| \cdot \|_p\) denotes the norm in the space \(L^p(\mathbb{R}^d)\), we have already mentioned that \(\|b_{s,d}\|_1 = 1\) and \(\|b_{s,d}\|_r < \infty\), for any \(1 \leq r < d/(d - s)\), which together with Young inequality (see Proposition B.65) yield the estimate

\[
\|(I - \Delta)^{-s/2} f\|_q \leq \|b_{s,d}\|_r \|f\|_p, \quad \forall f \in L^p(\mathbb{R}^d),
\]

i.e., the linear mapping \(f \mapsto (I - \Delta)^{-s/2} f\) is continuous from \(L^p(\mathbb{R}^d)\) into \(L^q(\mathbb{R}^d)\), for \(1/p - 1/q = 1 - 1/r, 1 \leq p \leq q < pd/(d - s)\), and for any \(s > 0\). Since \((I - \Delta)^{-s/2}\) is an isomorphism from \(S(\mathbb{R}^d)\) or \(S'(\mathbb{R}^d)\) onto itself, and \(L^p(\mathbb{R}^d) \subset S'(\mathbb{R}^d)\), the operator \((I - \Delta)^{-s/2}\) is one-to-one (injective) as defined on \(L^p(\mathbb{R}^d)\), \(1 \leq p \leq \infty\), for any \(s > 0\). Note that if \(s < 0\) then \((I - \Delta)^{-s/2}\) is not a continuous operator on \(L^p(\mathbb{R}^d)\).

- **Remark 6.17.** Consider the operator \((-\Delta)^{s/2}(I - \Delta)^{-s/2}\), with \(s > 0\), and defined via the Fourier transform as

\[
((-\Delta)^{s/2}(I - \Delta)^{-s/2} \varphi)(x) = \hat{\varphi}^{-1}\left((2\pi |\xi|)^{s} (1 + 4\pi^2 |\xi|^2)^{-s/2} \mathfrak{F}(\varphi)(\xi)\right)(x),
\]

initially for functions \(\varphi\) in \(S(\mathbb{R}^d)\). In this case, it becomes relevant to obtain its kernel, i.e., a tempered distribution \(\varpi_s\) such that

\[
\hat{\varpi}_s(\xi) = (2\pi |\xi|)^{s} (1 + 4\pi^2 |\xi|^2)^{-s/2}.
\]

To this end, as in Stein [113, Section 3.2, pp. 133–134], first note the expansion

\[
(1 - \lambda)^{s/2} = 1 + \sum_{k=1}^{\infty} a_{k,s} t^k, \quad \text{with} \quad \sum_{k=1}^{\infty} |a_{k,s}| < \infty,
\]
since \(a_{k,s}\) has eventually a constant sign and \((1 - \lambda)^{s/2}\) is bounded as \(\lambda \to 1\) and \(s \geq 0\). If \(\lambda = 1/(1 + 4\pi^2|\xi|^2)\) then
\[
\left(\frac{4\pi^2|\xi|^2}{1 + 4\pi^2|\xi|^2}\right)^{s/2} = 1 + \sum_{k=1}^{\infty} a_{k,s} (1 + 4\pi^2|\xi|^2)^{-k},
\]
and because the Bessel kernel \(b_{2k,d}(x)\), given by (6.19) with \(s = 2k\), satisfies \(\hat{b}_{k,d}(\xi) = (1 + 4\pi^2|\xi|^2)^{-k}\), we deduce
\[
\varpi = \delta_0 + \left(\sum_{k=1}^{\infty} a_{k,s} b_{2k,d}(x)\right)dx, \quad \text{with} \ \delta_0 \ \text{the Dirac measure}.
\]
Moreover, since the Bessel kernels are nonnegative and \(\|b_{2k,d}\|_1 = 1\), the tempered distribution \(\varpi\) is a finite signed measure on \(\mathbb{R}^d\), actually, with \(\varpi(A) \leq 1 + \sum_{k=1}^{\infty} |a_{k,s}|\), for any Borel subset \(A\) of \(\mathbb{R}^d\).

Beside the convolution property
\[
b_{s,d} \ast b_{t,d} = b_{s+t,d}, \quad \forall s, t > 0,
\]
and the fact that Bessel kernel is bounded in \(\mathbb{R}^d\) for \(s>d\), let us mention other preliminary bounds.

**Proposition 6.18.** If \(b_{s,d}(x)\) is the Bessel kernel defined by (6.19) with \(s>0\), \(x\) in \(\mathbb{R}^d\), \(x \neq 0\), then the asymptote as \(|x| \to \infty\) is the function
\[
b_{s,d}^\infty(x) = \frac{|x|^{(s-d-1)/2}e^{-|x|/2}}{2^{(s+d-1)/2\pi(d-1)/2}\Gamma(s/2)},
\]
while the asymptote as \(|x| \to 0\) is given by
\[
b_{s,d}^0(x) = \begin{cases} 
\frac{\Gamma(|d-s|/2)}{2^{s-d}\pi^{d/2}/2\Gamma(s/2)} |x|^{(s-d)\wedge 0} & \text{if } d \neq s, \\
-\ln|x| & \text{if } s = d,
\end{cases}
\]
where \((\cdot \wedge \cdot)\) denotes the minimum between numbers. Moreover
\[
b_{s,d}(x) = \frac{2^{d-s}\Gamma(d-s/2)}{\Gamma(s/2)} |x|^{s-d} b_{2d-s,d}(x), \quad \forall x \in \mathbb{R}^d, \ x \neq 0,
\]
if \(0 < s < 2d\).

**Proof.** To check the asymptote as \(|x| \to \infty\), let us show that
\[
0 \leq \frac{b_{s,d}(x)}{b_{s,d}^\infty(x)} - 1 \leq \int_0^{\infty} e^{-t} t^{-1/2} \rho_{s,d}(t/|x|) \, dt, \quad \forall x \in \mathbb{R}^d,
\]
(6.25)
where \( \rho_{s,d}(r) \) is a continuous function with polynomial growth as \( r \to \infty \) and satisfying \( \rho_{s,d}(r) \to 0 \) as \( r \to 0 \). Indeed, use the change of variables \( r + |x|^2/(4r) = |x|t \) within the \( r \)-interval \((0, |x|/2)\) with

\[
    r = (|x|/2)(t - \sqrt{t^2 - 1}) = \frac{|x|/2}{t + \sqrt{t^2 - 1}}, \quad 1 \leq t < \infty,
\]

and within the interval \((|x|/2, \infty)\) with

\[
    r = (|x|/2)(t + \sqrt{t^2 - 1}), \quad 1 \leq t < \infty,
\]

to deduce that

\[
    b_{s,d}(x) = \frac{(|x|/2)^{(s-d)/2}}{(4\pi)^{d/2}\Gamma(s/2)} \int_1^\infty e^{-|x|t} \left[ (t + \sqrt{t^2 - 1})^{(d-s)/2} + (t + \sqrt{t^2 - 1})^{(s-d)/2} \right] (t^2 - 1)^{-1/2} dt.
\]

Again, another simple change of variables \( t = 1 + r \) yields

\[
    b_{s,d}(x) = \frac{|x|^{(s-d)/2}e^{-|x|}}{2^{(s-d)/2}\pi^{d/2}\Gamma(s/2)} \int_0^\infty e^{-r|x|} q_{s,d}(r) dr, \quad \text{with}
\]

\[
    q_{s,d}(r) = \frac{1 + r + \sqrt{r^2 + 2r}^{(d-s)/2} + (1 + r + \sqrt{r^2 + 2r})^{(s-d)/2}}{\sqrt{r^2 + 2r}}.
\]

Since the change of variable \( |x|r = t \) yields

\[
    \int_0^\infty e^{-r|x|} q_{s,d}(r) dr = |x|^{-1/2} \int_0^\infty e^{-t^{1/2} - 1} \sqrt{t/|x|} q_{s,d}(t/|x|) dt,
\]

and \( \Gamma(1/2) = \sqrt{\pi} \), and \( \sqrt{r} q_{s,d}(r) \to \sqrt{2} \) as \( r \to 0 \), use the inequality

\[
    0 \leq \sqrt{r} q_{s,d}(r) \leq \frac{1 + [1 + r(r + 2)]^{d-s)/2}}{\sqrt{r + 2}} \leq 2 + C_{q,s,d} r^{(d-s)/2-1/2},
\]

for any \( r \geq 0 \) and some suitable constant \( C_{q,s,d} > 0 \), to obtain

\[
    0 \leq |x|^{1/2}(2\pi)^{-1/2} \int_0^\infty e^{-r|x|} q_{s,d}(r) dr - 1 \leq \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-t^{1/2} - 1} \left[ \sqrt{t/|x|} q_{s,d}(t/|x|) - \sqrt{2} \right] dt,
\]

which implies

\[
    c_{d,s}^{\infty}|x|^{(s-d-1)/2}e^{-|x|} \leq b_{s,d}(x) \leq C_{d,s}^{\infty}|x|^{(s-d-1)/2}e^{-|x|} \quad \text{if} \quad |x| \geq 1, \quad (6.26)
\]

with

\[
    c_{d,s}^{\infty} = \frac{1}{2^{(s+d-1)/2}\pi^{(d-1)/2}\Gamma(s/2)}, \quad C_{d,s}^{\infty} = C_{d,s}^{\infty}[2\sqrt{\pi} + C_{q,s,d}^{\infty}(d-s)/2],
\]
i.e., the estimate (6.25) holds true with \( \rho(r) = \sqrt{r} q_{s,d}(r) - \sqrt{2} \).

To verify the asymptote as \( |x| \to 0 \) for \( s \neq d \), let us show that

\[
0 \leq \frac{b_{s,d}(x)}{b_{s,d}^0(x)} - 1 \leq \int_0^\infty e^{-t|s-d|/2-1} \rho(|x|/t) \, dt, \quad \forall x \in \mathbb{R}^d,
\]

(6.27)

where \( \rho(|x|/t) = [1 - e^{-|x|/(4t)}]/\Gamma(|s - d|/2) \). Indeed, the change of variables \( |x|^2/(4r) = t \) yields

\[
b_{s,d}(x) = \frac{|x|^{s-d}2^{d-s}}{(4\pi)^{d/2}\Gamma(s/2)} \int_0^\infty e^{-|x|^2/(4t)}e^{-t(d-s)/2-1} dt
\]

and in particular

\[
b_{s,d}(x) = \frac{|x|^{s-d}2^{d-s}\Gamma(d-s/2)}{\Gamma(s/2)} b_{2d-s,d}(x), \quad \text{if } 0 < s < 2d.
\]

Hence

\[
b_{2d-s,d}(x) \uparrow \frac{\Gamma(d-s/2)}{(4\pi)^{d/2}\Gamma(s/2)}, \quad \text{as } |x| \to 0, s < d,
\]

\[
|x|^{d-s} b_{s,d}(x) \uparrow \frac{\Gamma(d/2-s/2)}{2^s \pi^{d/2}\Gamma(s/2)}, \quad \text{as } |x| \to 0, s < d,
\]

and

\[
c_{d,s}^0 |x|^{(s-d)\wedge 0} \leq b_{s,d}(x) \leq C_{d,s}^0 |x|^{(s-d)\wedge 0}, \quad \text{if } 0 < |x| \leq 2,
\]

(6.28)

with \( s \neq d \),

\[
c_{d,s}^0 = \frac{1}{2^s \pi^{d/2}\Gamma(s/2)} \int_0^\infty e^{-1/t}e^{-t(d-s)/2-1} dt, \quad C_{d,s}^0 = \frac{\Gamma(d/2-s/2)}{2^s \pi^{d/2}\Gamma(s/2)},
\]

i.e., the estimate (6.27) is proved.

Now, when \( s = d \), let us show that

\[
0 \leq b_{d,d}(x) - b_{d,d}^0(x) \leq \frac{c_0}{2^{d-1}\pi^{d/2}\Gamma(d/2)}, \quad \forall x \in \mathbb{R}^d, 0 < |x| \leq 1,
\]

(6.29)

for a suitable constant \( c_0 > 0 \). Indeed, begin with

\[
b_{d,d}(x) = \frac{2}{(4\pi)^{d/2}\Gamma(d/2)} \int_1^{\infty} e^{-|x|t}(t^2 - 1)^{-1/2} dt
\]

and use the change of variables \( |x|t = 1/r \) to obtain

\[
\int_1^{\infty} e^{-|x|t}(t^2 - 1)^{-1/2} dt = \int_0^{1/|x|} \frac{e^{-1/r}}{r \sqrt{1 - r^2|x|^2}} dr = \int_0^1 \frac{e^{-1/r}}{r \sqrt{1 - r^2|x|^2}} dr + \int_1^{1/|x|} \frac{e^{-1/r}}{r \sqrt{1 - r^2|x|^2}} dr.
\]
Hence, apply the inequalities
\[
\frac{1}{\sqrt{1 - r^2|x|^2}} \leq \frac{1}{\sqrt{1 - r}}, \quad \text{if} \quad |x| \leq 1 \quad \text{and} \quad 0 < r|x| < 1,
\]
\[
0 \leq \frac{1 - e^{-1/r}}{r} \leq \frac{1}{r^2}, \quad |\sqrt{x} - \sqrt{y}| \leq \sqrt{|x - y|},
\]
\[
0 \leq \frac{1}{\sqrt{1 - r^2|x|^2}} - 1 = \frac{1 - \sqrt{1 - r^2|x|^2}}{\sqrt{1 - r^2|x|^2}} \leq \frac{r|x|}{\sqrt{1 - r^2|x|^2}},
\]
and the decomposition
\[
e^{-1/r} \frac{1}{r \sqrt{(1 - r^2|x|^2)}} - \frac{1}{r} = e^{-1/r} \left[ \frac{1}{\sqrt{1 - r^2|x|^2}} - 1 \right] + \frac{e^{-1/r} - 1}{r}
\]
to find that
\[
\left| \ln |x| + \int_0^{1/|x|} \frac{e^{-1/r}} {r \sqrt{(1 - r^2|x|^2)}} \, dr \right| \leq \int_0^1 \frac{e^{-1/r}} {r \sqrt{1 - r^2|x|^2}} \, dr + \int_1^{1/|x|} \frac{e^{-1/r}|x|} {\sqrt{1 - r^2|x|^2}} \, dr + \int_1^{1/|x|} \frac{1}{r^2} \, dr \leq \int_0^1 \frac{e^{-1/r}} {r \sqrt{1 - r}} \, dr + \int_1^{|x|} \frac{1}{\sqrt{1 - t}} \, dt + \int_1^{\infty} \frac{1}{r^2} \, dr \leq c_0,
\]
for any $0 < |x| \leq 1$, i.e., the estimate (6.29) holds true.

- **Remark 6.19.** Actually, by means of the well known formula for the Fourier transform of a radial function $f(x) = f(|x|)$
\[
\hat{f}(\xi) = |\xi|^{(2-d)/2} \int_0^{\infty} f(\rho) \rho^{d/2} J_{(d-2)/2}(\rho|x|) \, d\rho, \quad x, \xi \in \mathbb{R}^d,
\]
with $J_\nu$ being the Bessel function of the first kind with order $\nu$, the Bessel kernel $b_{s,d}, 0 < s < d + 1,$ has the representation
\[
b_{s,d}(x) = \frac{|x|^{(s-d)/2}} {2^{(d+s)/2 - 1} \pi^{d/2} \Gamma(s/2)} K_{(d-s)/2}(|x|),
\]
where $K_{(d-s)/2}$ is the modified Bessel function of the third kind, e.g.,
\[
K_\nu(r) = \frac{(\pi/2)^{1/2} r^{\nu} e^{-r}} {\Gamma(\nu + 1/2)} \int_0^{\infty} e^{-rt} t^{\nu-1/2}(1 + t/2)^{\nu-1/2} \, dt,
\]
if $r > 0$, $\nu > -1/2$ and $K_\nu(r) = K_\nu(-r)$.

Certainly, from this representation the also the asymptotic limits follows, e.g., see Aronszajn and Smith [10].

Corollary 6.20. If $\partial_i$ denotes the partial derivative with respect to $x_i$ and $b_{s,d}(x)$ is the Bessel kernel defined by (6.19) then

$$\Gamma(s/2) \sum_{i=1}^{d} x_i \partial_i b_{s,d}(x) = -|x|^{s-d} \frac{2^d+2-s}{d-s} \Gamma(d+1-s/2) b_{2d+2-s,d}(x),$$

for $0 < s < 2d + 2$, and

$$|\partial_i b_{s,d}(x)| \leq C_{d,s} [b_{s,d}(x) + b_{s-1,d}(x)], \quad \forall x \in \mathbb{R}^d, \ x \neq 0, \ s > 1,$$

for a suitable constant $C_{d,s}$ depending only on the dimension $d$ and the order $s$. Moreover, $\partial_i b_{1,d}$ is an $L^p$-multiplier, i.e., for every $1 < p < \infty$ there exists a constant $C_{p,d}$ depending on $p$ and the dimension $d$ such that

$$\|\partial_i b_{1,d} \ast \varphi\|_p \leq C_{p,d} \|\varphi\|_p, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).$$

Furthermore, for $s > 0$ and any $R > 0$ there is a constant $C$ depending only on $s$, $R$ and the dimension $d$ such that the average estimate

$$r^{-d} \int_{|y-x|<r} b_{s,d}(z-y) \, dy \leq C b_{s,d}(z-x), \quad \forall x, z \in \mathbb{R}^d, \ z \neq x,$$

holds true.

Proof. First, remark that

$$\sum_{i=1}^{d} x_i \partial_i b_{s,d}(x) = -\frac{1}{(4\pi)^{d/2} \Gamma(s/2)} \int_0^\infty e^{-r} e^{-|x|^2/(4r)} \frac{|x|^2}{4r} r^{(s-d)/2-1} \, dr$$

and that the change of variables $|x|^2/(4r) = t$ yields

$$\sum_{i=1}^{d} x_i \partial_i b_{s,d}(x) = -\frac{|x|^{s-d} \Gamma(d+1-s/2)}{(4\pi)^{d/2} \Gamma(s/2)} \int_0^\infty e^{-t} e^{-t} t^{(s-d)/2-1} \, dt,$$

which yields the first equality.

Next, note that if $|x| \geq 1$ then the bound (6.26) yields

$$|x|^\alpha b_{s-2\alpha,d}(x) \leq |x|^\alpha C_{d,s-2\alpha}^\infty |x|^{(s-2\alpha-d-1)/2} e^{-|x|} \leq \frac{C_{d,s-2\alpha}^\infty}{C_{d,s}} b_{s,d}(x),$$

i.e., $|x|^\alpha b_{s-2\alpha,d}(x) \sim b_{s,d}(x)$ when $|x| \geq 1$. However, if $0 < |x| \leq 2$ the bound (6.28) implies

$$|x|^\alpha b_{s-2\alpha,d}(x) \leq C^0_{d,s-2\alpha} |x|^\alpha |x|^{s-2\alpha-d} \leq \frac{C^0_{d,s-2\alpha}}{C^0_{d,s}} b_{s-\alpha,d}(x),$$

i.e., $|x|^\alpha b_{s-2\alpha,d}(x) \sim b_{s-\alpha,d}(x)$ when $|x| \leq 2$, provided $0 < s - 2\alpha < d$ and $0 < s - \alpha < d$. This shows that

$$|x|^\alpha b_{s-2\alpha,d}(x) \leq C_{s,\alpha,d} [b_{s,d}(x) + b_{s-\alpha,d}(x)], \quad \forall x \in \mathbb{R}^d \setminus \{0\}. \quad (6.30)$$
Moreover, if \( s - 2\alpha = d \) then \(|x|\alpha b_{s-2\alpha,d}(x)\) is bounded as \(|x| \to 0\) for any \( \alpha > 0\), and because all Bessel kernels approach either a positive constant or infinite as \(|x| \to 0\), the estimate \((6.30)\) holds true for all possible values of \( \alpha \). Similarly, if \( s - 2\alpha = d \) and \( \alpha < 0 \) then \( s - d < s - \alpha - d = \alpha < 0 \), so that as \(|x| \to 0\), the \(|x|\alpha b_{s-2\alpha,d}(x)\) behaves as \(|x|\alpha \ln |x|\), while the right-hand side behaves as \(|x|^{s-d} + |x|^{\alpha}\), i.e., the estimate \((6.30)\) holds true for all possible values of \( \alpha \) and \( s \) (either \( s > \alpha \), \( s > 2\alpha \) and \( s > 0 \) or simply, just ignore the normalizing constants). Finally, use the equality

\[
\partial_i b_{s,d}(x) = \frac{x_i}{2(4\pi)^{d/2}\Gamma(s/2)} \int_0^\infty e^{-r|x|^2/(4r)} (s-2-d/2-1)dr
\]

and the previous argument with \( \alpha = 1 \) to obtain the estimate for the derivative.

To verify that \( \partial_i b_{1,d}(\xi) = -2\pi i \xi(1 + |\xi|^2)^{-1/2} \) and use Theorem 6.16. It is clear that this is not a self-contained proof, since Theorem 6.16 was only stated.

To show the validity of the average estimate, note that in view of the composition property \((6.24)\), i.e., \( b_{s,d} \ast b_{t,d} = b_{s+t,d} \) for \( s, t > 0 \), only the case \( 0 < s < d \) should be checked. Thus, the asymptotic estimate \((6.28)\) implies the assertion when \(|z - x| \leq 2\). Indeed, use the inequalities \(|z - y| \geq |z| - |y|\) and \(|z - y| \geq |y| - |z|\) to see that if \(|z - y| < |y|/2\) then \(|y|/2 < |z| < 3|y|/2\). Hence, take \( x = 0\), \( R = 1\), replace \( b_{s,d}\) with \(|\cdot|^{s-d}\), and for \( r \leq 1 \) and \(|z| \leq 2\) decompose the integral over \(|y| \leq r\) into two pieces, over \(|z - y| \leq |y|/2\) and over \(|z - y| > |y|/2\) to obtain

\[
r^{-d}|z|^{d-s} \int_{|y| \leq r, |z-y| \leq |z|/2} |z - y|^{s-d} dy \leq c_d,
\]

\[
r^{-d}|z|^{d-s} \int_{|y| \leq r, |z-y| > |z|/2} |z - y|^{s-d} dy \leq 2
\]

for a suitable constant \( c_d > 0 \) depending only on the dimension \( d\), e.g., \( c_d = 3^4\omega_d\), \( \omega_d = \pi^{-d/2}/\Gamma(d/2 + 1)\).

Next, if \(|z - x| \geq 2\) and \( x = 0\) then use the asymptotic as \(|x| \to \infty\) and note that the Bessel kernel \( b_{s,d}(|x|) = b_{s,d}(x)\) is a decreasing function of \(|x|\) to deduce

\[
\frac{1}{r^d b_{s,d}(z)} \int_{|y| \leq r} b_{s,d}(z - y) dy \leq \sup_{\rho \geq 2} \frac{b_{s,d}(\rho - 1)}{b_{s,d}(\rho)},
\]

where the supremum is finite in view of the continuity and the asymptotic \( b_{s,d}(\rho - 1)/b_{s,d}(\rho) \to e \) as \( \rho \to \infty\).

\*Remark 6.21. The last estimate on the Bessel kernel yields several interesting consequences. For instance, the convolution \( b_{s,d} \ast \varphi \) satisfies

\[
r^{-d} \int_{|y-x| < r} (b_{s,d} \ast \varphi)(y) dy \leq C(b_{s,d} \ast \varphi)(x), \quad \forall x \in \mathbb{R}^d, \, \varphi \geq 0,
\]

\[
\lim_{r \to 0} \frac{\Gamma(d/2 + 1)}{\pi^{d/2} r^d} \int_{|y-x| < r} (b_{s,d} \ast \varphi)(y) dy = (b_{s,d} \ast \varphi)(x), \quad \forall x \in \mathbb{R}^d,
\]

[Preliminary]
In particular if \( \chi_\varepsilon(x) = \varepsilon^{-d} \mathbb{1}_{|x/\varepsilon| \leq 1} \) then \( (b_{s,d} \star \chi_\varepsilon)(x) \leq C b_{s,d}(x) \) for every \( x \in \mathbb{R}^d \), and for any regularizing functions (continuous with compact support) \( \varphi_\varepsilon(x) = \varepsilon^{-d} \varphi(x/\varepsilon) \) replacing \( \chi_\varepsilon \).

\[ \square \]

- **Remark 6.22.** Based on the equality \( \partial_i b_{1,d}(\xi) = -2\pi i \xi (1+|\xi|^2)^{-1/2} \) we deduce in Corollary 6.20 that \( \partial_i b_{1,d} \) is an \( L^p \)-multiplier, i.e., for every \( 1 < p < \infty \) there exists a constant \( C_{p,d} \) depending on \( p \) and the dimension \( d \) such that

\[
\| \partial_i b_{1,d} \star \varphi \|_p \leq C_{p,d} \| \varphi \|_p, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).
\]

Similarly, the kernel \( k_z(x) = [b_{s,d}(x + z) - b_{s,d}(x)]|z|^{-s} \), \( 0 < s \leq 1 \), is an \( L^p \)-multiplier uniformly in \( z \) with \( |z| \leq 1 \), i.e.,

\[
\| k_z \star \varphi \|_p \leq C_{p,d} \| \varphi \|_p, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d),
\]

where the constant \( C_{p,d} \) is independent of \( z \) in \( \mathbb{R}^d \), \( |z| \leq 1 \). Indeed, the asymptotic estimates as \( |x| \to 0 \) yields

\[
|b_{s,d}(x + z) - b_{s,d}(x)| \leq C |z| [\| |x|^{-d-s+1} + |x|^{-d+s-1}],
\]

for a suitable constant \( C > 0 \), and because \( |z + x| < |x|/2 \) implies \( |x|/2 < |z| < 3|x|/2 \), we deduce \( |k_z(x)| \leq C |x|^{-d} \), for every \( x \in \mathbb{R}^d \) with \( |x| \leq 2 \) and \( |z| \leq 1 \). If \( |x| > 2 \) then the asymptotic as \( |x| \to \infty \) take care of the estimate, which allows us to applied Theorem 6.16. Actually, the above estimate on \( b_{s,d}(x + z) - b_{s,d}(x) \) as well as the relation \( \hat{\chi}[b_{s,d}(x + z) - b_{s,d}(x)](\xi) = [e^{-2\pi i z \cdot \xi} - 1](1 + |\xi|^2)^{-s/2} \), also prove that if \( 0 < s < 1 \) then

\[
\int_{|z| \leq 1} |b_{s,d}(x + z) - b_{s,d}(x)| |z|^{-d-s} \, dz \leq C |x|^{-d},
\]

\[
\int_{|x| \geq 2|y|} dx \int_{|z| \leq 1} |b_{s,d}(x + y + z) - b_{s,d}(x + y)| |z|^{-d-s} \, dz \leq C.
\]

Hence, tracking the dependency of the bounds \( A_k \) in Theorem 6.16, the estimate

\[
\left[ \int_{|z| \leq 1} \| k_z \star \varphi \|_p^p |z|^{-d} \, dz \right]^{1/p} \leq C_{p,d} \| \varphi \|_p, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d)
\]

follows. \[ \square \]

In any case, it is clear that the arguments given in the previous Remark 6.22 are not self-contained, since Theorem 6.16 was only stated. However, the following estimate can be obtained in a self-contained way.

**Proposition 6.23.** If \( b_{s,d}(x) \) is the Bessel kernel defined by (6.19) with \( x = (x', x_d) \), \( x' \) in \( \mathbb{R}^{d-1} \) and \( x_d \) in \( \mathbb{R} \), and \( 0 < s < 1 \) then

\[
\int_{\mathbb{R}^{d-1}} b_{s,d}(y' - x', x_d) \, dx' \leq C_{d,s} |x_d|^{s-1}, \quad \forall x', y' \in \mathbb{R}^{d-1}, \, x_d \in \mathbb{R} \setminus \{0\},
\]
and
\[ \int_{\mathbb{R}^{d-1}} |b_{s,d}(z' - x', x_d) - b_{s,d}(y' - x', x_d)| dx' \leq \]
\[ \leq C_{d,s}|x_d|^{8-2}|z' - y'|, \quad \forall x', y', z' \in \mathbb{R}^{d-1}, x_d \in \mathbb{R} \setminus \{0\}, \]
for some constant \(C_{d,s}\) depending only on the dimension \(d\).

Proof. This estimate can be obtained as a consequence of the asymptotic behavior of the Bessel kernel \(b_{s,d}(x)\) or by means of the heat-kernel

\[
\begin{align*}
\left\{ \begin{array}{l}
\mathbf{b}_{s,n}(x) = \frac{1}{\Gamma(s/2)} \int_0^\infty e^{-t s/2 - 1} h_n(x, t) dt, \quad \forall x \in \mathbb{R}^n, \\
\text{with} \quad h_n(x, t) = (4\pi t)^{-n/2} e^{-|x|^2/(4t)}, \quad \forall t > 0, \ x \in \mathbb{R}^n,
\end{array} \right.
\end{align*}
\]

(6.31)

and in particular
\[
\mathbf{b}_{s,d}(x', x_d) = \frac{1}{2\sqrt{\pi}(s/2)} \int_0^\infty e^{-t (s-1)/2 - 1} e^{-x_d^2/(4t)} h_{d-1}(x', t) dt.
\]

Therefore, before going further, let us mention some properties of the heat-kernel or exponential factor \(e^{-c|x|^2/t}\) with \(c, t > 0\) and \(x \in \mathbb{R}^n\):

1. First, calculate the gradient in \(x\) to get \(\nabla e^{-c|x|^2/t} = 2c(x/t)e^{-c|x|^2/t}\), which yields exponential bounds of the form
\[
|\nabla e^{-c|x|^2/t}| \leq \frac{2c}{\sqrt{t}} \left( \frac{|x|}{\sqrt{t}} e^{-c|x|^2/t} \right),
\]
\[
\frac{|x|}{\sqrt{t}} e^{-c|x|^2/t} \leq (\text{sup}_{r > 0} \{re^{-r}\}) e^{-c(1-\varepsilon)|x|^2/t},
\]
for every \(c \geq \varepsilon > 0, t > 0\) and \(x \in \mathbb{R}^n, n = 1, 2, \ldots, d\). Similarly, this means that any \(k\) partial derivative in \(x\) is dominate by an exponential kernel with \(c - \varepsilon\) in lieu of \(c\) and a factor of \(t^{-k/2}\). Partial derivative in \(t\) are bounded with a \(t^{-k}\) factor.

2. Next, integrating in \(x\) over \(\mathbb{R}^n\) with the change of variable \(x = \sqrt{t}y\)
\[
\int_{\mathbb{R}^n} |x|^\alpha e^{-c|x|^2/t} dx = t^{(n+\alpha)/2} \int_{\mathbb{R}^n} |y|^\alpha e^{-c|y|^2/2} dy, \quad \forall \alpha > -n, t > 0,
\]
i.e., a \(t^{(n+\alpha)/2}\) factor is earned.

3. Use the equality
\[
e^{-c|x|^2/t} - e^{-c|y|^2/t} =
\]
\[
= 2c \int_0^1 \left( \frac{x - y}{\sqrt{t}} \right) \left( \frac{y + \theta(x - y)}{\sqrt{t}} \right) e^{-c|y + \theta(x - y)|^2/t} d\theta
\]

to obtain, either
\[
\left| e^{-c|x|^2/t} - e^{-c|y|^2/t} \right| \leq 2c \sup_{r > 0} \{ re^{-r^2} \} |x - y| t^{-1/2} \max \left\{ e^{-(c-\varepsilon)|x|^2/t}, e^{-(c-\varepsilon)|y|^2/t} \right\},
\]
or integrating in \( \mathbb{R}^n \),
\[
\int_{\mathbb{R}^n} \left| e^{-c|x+z|^2/t} - e^{-c|y+z|^2/t} \right| dz \leq 4c \sup_{r > 0} \{ re^{-r^2} \} |x - y| t^{(n-1)/2} \int_{\mathbb{R}^n} e^{-(c-\varepsilon)|z|^2} dz,
\]
for every \( c > \varepsilon > 0 \), \( t > 0 \) and \( x, y \) in \( \mathbb{R}^n \), \( n = 1, 2, \ldots, d \). Now, the interpolation inequality \( |a - b| \leq 2|a - b|^\alpha \max\{|a|, |b|\}^{1-\alpha} \) implies that
\[
\left| e^{-c|x|^2/t} - e^{-c|y|^2/t} \right| \leq C_\varepsilon |x - y|^\alpha t^{-\alpha/2} \max \left\{ e^{-(c-\varepsilon)|x|^2/t}, e^{-(c-\varepsilon)|y|^2/t} \right\},
\]
and
\[
\int_{\mathbb{R}^n} \left| e^{-c|x+z|^2/t} - e^{-c|y+z|^2/t} \right| dz \leq 2C_\varepsilon |x - y|^\alpha t^{-(n-\alpha)/2} \int_{\mathbb{R}^n} e^{-(c-\varepsilon)|z|^2} dz,
\]
with
\[
C_\varepsilon = 4c \sup_{r > 0} \{ re^{-r^2} \}.
\]
This is referred to as the H"older estimates for the hear-kernel in \( \mathbb{R}^n \).

Now, the \((d - 1)\)-dimensional heat-kernel can be estimated as follows
\[
\int_{\mathbb{R}^{d-1}} h_{d-1}(y' - x', t) dx' \leq C,
\]
\[
\int_{\mathbb{R}^{d-1}} |h_{d-1}(z' - x', t) - h_{d-1}(y' - x', t)| dx' \leq Ct^{-1/2} |z' - y'|,
\]
for any \( z', y' \) in \( \mathbb{R}^{d-1} \), \( x_d \) in \( \mathbb{R} \), \( t > 0 \) and a suitable constant \( C > 0 \) depending only on the dimension \( d \). Hence
\[
\int_{\mathbb{R}^{d-1}} b_{s,d}(y' - x', x_d) dx' \leq \frac{C}{2\sqrt{\pi} \Gamma(s/2)} \int_0^{\infty} e^{-t(t(s-1)/2-1-x_d^2/(4t))} dt
\]
and
\[
\int_{\mathbb{R}^{d-1}} |b_{s,d}(z' - x', x_d) - b_{s,d}(y' - x', x_d)| dx' \leq \frac{C}{2\sqrt{\pi} \Gamma(s/2)} \int_0^{\infty} e^{-t(t(s-1)/2-1-x_d^2/(4t))} \min \left\{ 1, t^{-1/2} |z' - y'| \right\} dt.
\]
Thus, the change of variables $x_d^2/(4t) = r$ yields
\[
2^{s-1} \int_0^\infty e^{-t\ell(s-1)/2-1} e^{-x_d^2/(4t)} dt =
= x_d^{s-1} \int_0^\infty e^{-x_d^2/(4r)(1-s)/2-1} e^{-r} dr \leq x_d^{s-1} \Gamma((1-s)/2)
\]
and
\[
2^{s-2} \int_0^\infty e^{-t\ell s/2-2} e^{-x_d^2/(4t)} dt =
= x_d^{s-2} \int_0^\infty e^{-x_d^2/(4r)(2-s)/2-1} e^{-r} dr \leq x_d^{s-2} \Gamma((2-s)/2),
\]
which implies the desired estimates. \hfill \Box

- **Remark 6.24.** It should be clear that with the technique used in Proposition 6.23, we can obtain the estimate
\[
\int_{\mathbb{R}^{d-1}} |b_{s,d}(x' + z', x_d) - 2b_{s,d}(x', x_d) + b_{s,d}(x' - z', x_d)| dx' \leq C_{d,s} |x_d|^{s-3}|z'|^2, \quad \forall x', z' \in \mathbb{R}^{d-1}, x_d \in \mathbb{R} \setminus \{0\},
\]
for some constant $C_{d,s}$ depending only on the dimension $d$. \hfill \Box

- **Remark 6.25.** Note that the singulary growth and Hölder continuity of a kernel $k(x)$ or $k(x, y)$ can be expressed by the conditions
\[
|k(x)| \leq C|x|^{-n}, \quad |k(x) - k(x')| \leq C|x - x'|^\alpha (|x|^{-n-\alpha} + |x'|^{-n-\alpha}),
\]
for every $x$ and $x'$ and some constant $C > 0$. It is clear that the Hölder condition is only useful when $|x - x'|$ is smaller than $|x|^{-n-\alpha}$ and $|x'|^{-n-\alpha}$, and it can be replaced by a more compact condition, namely
\[
|k(x) - k(x')| \leq C|x - x'|^\alpha (|x|^{-n-\alpha} \text{ if } 2|x - x'| < |x|,
\]
which, a priori seems a non-symmetric assumption. However, if $2|x - x'| < |x'|$ then $2|x - x'| \leq |x - x'| + |x'| \leq |x|$, i.e., the inequality also holds true if $2|x - x'| < |x'|$. Regarding the case where $2|x - x'| \geq |x'|$ or $2|x - x'| \geq |x|$, the singulary growth inequality take precedence, and the Hölder continuity estimate holds in this case. Certainly, this argument applies to a kernel $k(x, y)$ or a parabolic case like $k(x, t)$.

\hfill \Box

### 6.2.3 Fundamental Solutions

Recall that if $T$ is a distribution in $\mathbb{R}^d$ such that
\[
\langle T, \varphi \rangle = \lambda^k \langle T, \varphi \lambda \rangle, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d), \forall \lambda > 0, \varphi \lambda(x) = \lambda^{-d} \varphi(x/\lambda).
\]
then $T$ is called \textit{homogeneous of degree} $k$. Derivatives and Fourier transform preserve homogeneous distributions, i.e., if $T$ is homogeneous of degree $k$ then $\partial^\alpha T$ is homogeneous of degree $k - |\alpha|$ and $\hat{T}$ is homogeneous of degree $-d - k$. Usually, homogeneous functions (or distributions) are smooth on $\mathbb{R}^*_d = \mathbb{R}^d \setminus \{0\}$, i.e., the restriction $T|_{\mathbb{R}^*_d}$ or the multiplication $T\chi$ is a smooth distribution (for any smooth function $\chi$ satisfying $\chi = 0$ in a neighborhood of the origin). Note that the same argument used to find the Fourier transform of $e^{-\pi|x|^2}$ also shows that $\mathcal{F}[e^{i\lambda\pi|x|^2}] = (i\lambda)^{-d/2}e^{-i\pi|x|^2/\lambda}$, where in the calculation the main branch of $z \mapsto z^{-d/2}$ is taken.

The fundamental solution $F$ of the partial differential equation $(I - \Delta)u = f$ (i.e., the tempered distribution solution $u = F$ for $f = \delta$, the Dirac function) can be found by means of the Fourier transform, i.e., solving the equation $(1 + 4\pi^2|\xi|^2)\hat{F} = 1$, which yields $F = b_{2,d}$, the Bessel kernel of order 2. Similarly, for the partial differential equation $-\Delta u = f$, the fundamental solution $F$ is the Riesz potential of order 2, i.e., the formula $F[\delta(a)(\pi|x|)^{-a}] = \Gamma(d/2 - a)(\pi|x|)^{a - d/2}$ yields $(d - 2)F(x) = |x|^{2-d}/\omega_{d-1}$ for $d \geq 3$, with $\omega_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$ the area of the unit sphere in $\mathbb{R}^d$ and $F(x) = -\ln |x|/(2\pi)$ for $d = 2$. On the other hand, an application of the divergence Theorem yields

$$\int_\Omega (v\Delta u - u\Delta v) \, dx = \int_{\partial\Omega} (u\partial_\nu v - v\partial_\nu u) \, dx',$$

and the expression of the fundamental solution follows when taking $\Omega = \{y \in \mathbb{R}^d : |y - x| > \varepsilon\}$ and letting $\varepsilon$ vanishes.

The parabolic counterpart of the previous elliptic PDEs are the equations $(\partial_t - \Delta)u = f$, where now $u(x,t)$. In this case the fundamental solution $F$ is given by the heat-kernel, namely $F(x,t) = h_d(x,t) = (2\pi t)^{-d/2}e^{-|x|^2/(4t)}$. It is clear that for the equation $(\partial_t - \Delta + \lambda)u = f$, the expression take the form $F(x,t) = (2\pi t)^{-d/2}e^{-|x|^2/(4t) - \lambda t}$.

For the wave equation $(\partial^2_t - \Delta)u = f$, the fundamental solution is given by $F(x,t) = t^{1-d} F_1(x/t)$, with $F_1(x) = F(x,1)$, and $\hat{F_1}(\xi) = (\sin |\xi|)/|\xi|$. In general, if the fundamental solution is known then the solution $u$ is given by the convolution $F \ast f$.

Besides PDEs in the whole space, the interest could be on half-space, e.g., the equation $-\Delta u = 0$ in $\mathbb{R}^{d-1} \times (0,\infty)$ with a (Dirichlet) boundary condition $u = g$ in $\mathbb{R}^{d-1} \times \{0\}$, which yields the Poisson kernel

$$p(x) = \Gamma(d/2)\pi^{-d/2} x_d |x|^{-d}, \quad \forall t > 0, x = (x',x_d) \in \mathbb{R}^{d-1} \times (0,\infty),$$

$$\widehat{p(\cdot,x_d)}(\xi') = e^{-2\pi x_d|\xi'|}, \quad \forall t > 0, \xi' \in \mathbb{R}^{d-1}.$$
i.e., in terms of the heat-kernel. Similarly, for the elliptic PDE \((I - \Delta)u = 0\) in \(\mathbb{R}^{d-1} \times (0, \infty)\) and \(u = g\) in \(\mathbb{R}^{d-1} \times \{0\}\), the corresponding Poisson kernel is given by \(-\partial_x b_{2,d}(x)\).

Sobolev spaces and similar spaces appeared in the study of PDEs, where specific estimates could be obtained via fundamental solutions. The Sobolev spaces discussed are mainly designed for elliptic PDE, while the parabolic counterpart needs more elaboration. In any case, key estimates take the form

\[
\|\varphi\|_{W^{2,p}(\mathbb{R}^d)} \leq C_p \|(I - \Delta)\varphi\|_{L^p(\mathbb{R}^d)}, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d),
\]

or equivalently

\[
\|b_{2,d} \ast \varphi\|_{W^{2,p}(\mathbb{R}^d)} \leq C_p \|\varphi\|_{L^p(\mathbb{R}^d)}, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d),
\]

for some suitable constant \(C_p > 0\), for any \(1 < p < \infty\), and various variations necessary to include the boundary conditions. Estimates with \(p = \infty\) are also valid, but with the space \(W^{2+s',\infty}(\mathbb{R}^d)\), \(0 < s' < 1\), i.e., with the Hölder spaces denoted by \(C^{2,\alpha}_b(\mathbb{R}^d)\), e.g.,

\[
\|\varphi\|_{C^{2,\alpha}_b(\mathbb{R}^d)} \leq C_\alpha \|(I - \Delta)\varphi\|_{C^{\alpha}_b(\mathbb{R}^d)}, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d), \quad 0 < \alpha < 1,
\]

known as the Schauder’s estimates. For instance, the reader is referred to the research books Ladyzhenskaya and Ural’tseva [78] and Gilbarg and Trudinger [55] for elliptic PDEs, and to Ladyzhenskaya et al. [77] and Lieberman [81] for parabolic PDEs. Certainly, besides PDEs in the whole space \(\mathbb{R}^d\) or half-space \(\mathbb{R}^d_+\), also a smooth domain \(\Omega \subset \mathbb{R}^d\) is used.

Fundamental solutions is a name used within a given PDE in the whole or half space. When the PDE is set in a domain \(\Omega\) of \(\mathbb{R}^d\), the fundamental solution is known as the Green (kernel) function or as the Poisson (kernel) function (with various boundary conditions. Moreover, integro-partial differential equations can also be studied, e.g., see the books Garroni and Menaldi [49, 50].

### 6.3 Besov and Sobolev Relations

Depending on the parameters, Sobolev and Besov spaces coincide. Initially, Sobolev spaces impose partial derivatives to belong to some \(L^p\)-space (see Introduction to Sobolev Spaces Chapter 4) and the alternative way using Fourier transform to express the same requirement in the Hilbertian Sobolev Spaces Section 6.1. Besov spaces focus on properties of the modulus of continuity expressed with various parameters (e.g., see the textbooks Adams and Fournier [3, Chapter 7], Grafakos [58, Chapter 6], Leoni [79, Chapter 14]), but our interest is only on one parameter as defined below, besides \(1 \leq p \leq \infty\) used for the \(L^p\)-spaces. Note that interpolation theory is very useful for these spaces, e.g., Bergh and Löfström [18, Chapter 6, pp. 131–173] and Triebel [126, 128].
6.3.1 Besov spaces

Before given a definition, let us discuss some preliminaries. Recall the translation operator \( \tau_z f(x) = f(x + z) \) for any function \( f \) in \( L^p(\mathbb{R}^d) \), and define the first difference operator \( \Delta_z f(x) = (\tau_z - I) f(x) = [f(x + z) - f(x)] \), and by induction, the higher-order differences \( \Delta^k_z = \Delta_z (\Delta^k_z) \). Thus, note that the iterations of the translation operator produces \( \tau_z^k = \tau_{kz} \) to deduce the equality

\[
\Delta^k_z f(x) = (\tau_z - I)^k f(x) = \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} f(x + iz). \quad (6.32)
\]

Hence, the invariance of the translation in \( \mathbb{R}^d \) and the equality \( 2^k = \sum_{i=1}^{k} \binom{k}{i} \) yields the inequality \( \|\Delta^k_z f\|_p \leq 2^k \|f\|_p \) and \( \|\tau_z^k f\|_p \leq 2\|f\|_p \), with

\[
\|f\|_p = \left( \int_{\mathbb{R}^d} |f(x)|^p \, dx \right)^{1/p}
\]

being (of course) the \( L^p \)-norm in \( \mathbb{R}^d \). Let us now prove Marchaud’s inequality

**Proposition 6.26.** If \( k \) is positive integers then

\[
\|\Delta^{k+1}_z f\|_p \leq 2\|\Delta^k_z f\|_p \leq k \sum_{i=0}^{\infty} 2^{-ik}\|\Delta^{k+1}_z f\|_p,
\]

for any \( z \) in \( \mathbb{R}^d \) and \( f \) in \( L^p(\mathbb{R}^d) \).

**Proof.** The left-hand inequality follows immediately from inequalities \( \Delta^{k+1}_z f = \Delta_z (\Delta^k_z f) \) and \( \|\Delta_z f\|_p \leq 2\|f\|_p \).

Next, the equality \( \Delta^{2}_z = \tau_z^2 - I = (\tau_z + I)\Delta_z \) implies \( \|\Delta^{2}_z f\|_p \leq 2^{k}\|\Delta^k_z f\|_p \). Moreover,

\[
\Delta^k_z = (\tau_z^2 - I)^{k} = (\tau_z + I)^k \Delta^k_z,
\]

and

\[
(\tau_z + I)^k - 2^k I = \sum_{i=1}^{k} \binom{k}{i} (\tau_{iz} - I) = \sum_{i=1}^{k} \binom{k}{i} \sum_{j=0}^{i-1} \tau_{jz} \Delta_z
\]

yields

\[
\Delta^k_z = 2^k \Delta^k_z + \sum_{i=1}^{k} \binom{k}{i} \sum_{j=0}^{i-1} \tau_{jz} \Delta^{k+1}_z + \text{ or equivalently}
\]

\[
\Delta^k_z = 2^{-k} \left[ \Delta^{k}_z - \sum_{i=1}^{k} \binom{k}{i} \sum_{j=0}^{i-1} \Delta^{k+1}_z \tau_{jz} \right].
\]

Hence, use the translation invariance equality \( \|\Delta^{k+1}_z \tau_{jz} f\|_p = \|\Delta^{k+1}_z f\|_p \) to deduce

\[
\|\Delta^k_z f\|_p \leq 2^{-k} \left[ \|\Delta^{k}_z f\|_p + \sum_{i=1}^{k} \binom{k}{i} i \|\Delta^{k+1}_z f\|_p \right],
\]
which means,
\[ \| \Delta_z^k f \|_p \leq \frac{k}{2} \| \Delta_z^{k+1} f \|_p + 2^{-k} \| \Delta_z^k f \|_p. \]

This last inequality can be iterated with \( h \) replaced by \( 2h \) to deduce
\[ \| \Delta_z^k f \|_p \leq \frac{k}{2} \sum_{i=0}^{n} 2^{-ik} \| \Delta_z^{k+1} f \|_p + 2^{-k(n+1)} \| \Delta_z^{k+1} f \|_p. \]

Thus, because \( \| \Delta_z^{k+1} f \|_p \leq 2^k \| f \|_p \), the desired estimate follows by taking limit as \( n \to \infty \).

If \( k \) is a nonnegative integer and \( 1 \leq p, \theta \leq \infty \) then define the Besov semi-norm
\[ |f|_{s,p,\theta,k} = \left( \int_{\mathbb{R}^d} \| \Delta_z^k f \|_p |z|^{-d-s\theta} \, dz \right)^{1/\theta}, \quad 0 < s < k, \] (6.33)
and as usual, for \( \theta = \infty \) or \( p = \infty \) the integral is replaced by the essential supremum.

**Corollary 6.27.** If \( k < m \) are two positive integers and \( 0 < s < k \) then the seminorms \( | \cdot |_{s,p,\theta,k} \) and \( | \cdot |_{s,p,\theta,m} \) are equivalent.

**Proof.** By induction, it is clear that only the case \( m = k + 1 \) need to be considered. To this end, the left-hand inequality in Proposition 6.26 yields
\[ \int_{\mathbb{R}^d} \| \Delta_z^{k+1} f \|_p |z|^{-d-s\theta} \, dz \leq 2^\theta \int_{\mathbb{R}^d} \| \Delta_z^k f \|_p |z|^{-d-s\theta} \, dz, \]
\[ i.e., \ |f|_{s,p,\theta,k+1} \leq 2 |f|_{s,p,\theta,k}. \]

Similarly, use the right-hand inequality in Proposition 6.26 to obtain
\[ \int_{\mathbb{R}^d} \| \Delta_z^k f \|_p |z|^{-d-s\theta} \, dz \leq \left( k \sum_{i=0}^{\infty} 2^{-ik} \right) \theta \int_{\mathbb{R}^d} \| \Delta_z^{k+1} f \|_p |z|^{-d-s\theta} \, dz = \]
\[ = \left( k \sum_{i=0}^{\infty} 2^{-ik+is} \right) \theta \int_{\mathbb{R}^d} \| \Delta_z^{k+1} f \|_p |\zeta|^{-d-s\theta} \, d\zeta, \]
after the change of variables \( 2^i z = \zeta \). Since \( s < k \) the series is convergent and the inequality \( |f|_{s,p,\theta,k} \leq c |f|_{s,p,\theta,k+1} \) follows with \( c = k/(1 - 2^{-k-s}) \).

If \( \chi = \mathbb{1}_{(0,1)} \) and \( \chi^{*k} \) denotes the corresponding \( B \)-spline functions, i.e., the \( k \)-convolution defined by \( \chi^{*0} = \delta \), and
\[ \chi^{*(k+1)}(t) = \chi \ast \chi^{*k}(t) = \int_{-\infty}^{+\infty} \chi(t-s) \chi^{*k}(s) \, ds, \quad k = 0, 1, \ldots, \]
then for any smooth function $\varphi$ in $\mathbb{R}^d$ the following identity

$$
\triangle_z^k \varphi(x) = \int_{-\infty}^{+\infty} \chi^k(t) \sum_{|\alpha|=k} \frac{k!}{\alpha!} \partial^\alpha \varphi(x + tz)z^\alpha dt
$$

holds true. Indeed, this is first valid for the one-dimensional case, and then by means of the function $\phi(t) = \varphi(x + tz/|z|)$ and the equality $\triangle_z^k \varphi(t) = \triangle_z^k \varphi((x + tz)/|z|)$, the desired identity is proved. This shows that

$$
\|\triangle_z^k \varphi\|_p \leq C|z|^k \sum_{|\alpha|=k} \|\partial^\alpha \varphi\|_p,
$$

for a suitable constant $C > 0$ depending only on $k$ and the dimension $d$. Hence, the seminorm $|\varphi|_{s,p,\theta,k}$ is dominated by $\|\varphi\|_p$ plus $\sum_{|\alpha|=k} \|\partial^\alpha \varphi\|_p$. A little more elaborated arguments used of interpolation can be developed to obtain a converse type estimate (e.g., see Bennett and Sharpley [16, Section 5.4, pp. 331–347]), namely,

$$
\|\partial^\alpha \varphi\|_p \leq C t^{-|\alpha|}\left(\|\varphi\|_p + t^r \sum_{|\beta|=r} \|\partial^\beta \varphi\|_p\right), \quad \forall t > 0, \; |\alpha| \leq r,
$$

$$
\inf_{\psi} \left\{ \|\varphi - \psi\|_p + t^k \sum_{|\alpha|=k} \|\partial^\alpha \psi\|_p \right\} \leq C \sup_{|z| \leq t} \{ \|\triangle_z^k \varphi\|_p \}, \quad \forall t > 0,
$$

for any $\varphi$ in $\mathcal{S}(\mathbb{R}^d)$, and some constant $C > 0$ independent of $t$ and $\varphi$, and eventually deduce that

$$
\sum_{|\alpha|=k} \|\partial^\alpha \varphi\|_p^p \leq C \int_{|z| \leq 1} \|\triangle_z^{k+1} \varphi\|_p^p \frac{|z|^{-d-kp}}{|z|} dz, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d),
$$

for a suitable constant $C > 0$ depending only on the dimension $d$, $k = 1, 2, \ldots$, and $1 \leq p < \infty$. These arguments setup the stage for the interpolation theory to be applied to deduce that the intermediate spaces between $L^p(\mathbb{R}^d)$ and $W^{m,p}(\mathbb{R}^d)$ are the Besov spaces $B^{s,p,\theta}(\mathbb{R}^d)$.

The Besov space $B^{s,p,\theta}(\mathbb{R}^d)$ has seminorms given by (6.33), with $1 \leq \theta \leq \infty$, usually, $k$ is the integer part of $s$, i.e., $s = k + s'$, $k = \lfloor s \rfloor$, $0 < s' < 1$, and the seminorm is denoted by $| \cdot |_{s,p,\theta}$, where any possible choice of $k$ produces equivalent seminorms, see Corollary 6.27. Certainly, by adding the $L^p$-norm, the Besov norm $\| \cdot \|_{s,p,\theta}$ is defined. However, as mentioned above, our interest is on the two-parameter Besov spaces $B^{s,p}(\mathbb{R}^d) = B^{s,p,\theta}(\mathbb{R}^d)$ as defined below, without the explicit use of interpolation theory.

**Definition 6.28** (Besov spaces). A function $f$ in $L^p(\mathbb{R}^d)$ belongs to the Besov space $B^{1,p}(\mathbb{R}^d)$ if the seminorm

$$
|f|_{1,p} = \left( \int_{\mathbb{R}^d} dx \int_{|z| \leq 1} |f(x + z) - 2f(x) + f(x - z)|^p \frac{|z|^{-d-p}}{|z|} dz \right)^{1/p}
$$

[PDF page: 250] Chapter 6. Besov and Sobolev Spaces
is finite. Similarly, a function $f$ in $L^p(\mathbb{R}^d)$ belongs to the Besov space $B^{s,p}(\mathbb{R}^d)$, $0 < s < 1$ if the seminorm

$$|f|_{s,p} = \left( \int_{\mathbb{R}^d} dx \int_{|z| \leq 1} |f(x + z) - f(x)|^p |z|^{-d-sp} dz \right)^{1/p}$$

is finite. Certainly, $B^{0,p}(\mathbb{R}^d) = L^p(\mathbb{R}^d)$. Next, by means of induction and differentiation (referred to as reduction) the Besov spaces $B^{s,p}(\mathbb{R}^d)$ are constructed, i.e., if $s = k + s'$ with $0 < s \leq 1$ and $k$ a nonnegative integer then $B^{s,p}(\mathbb{R}^d)$ is the vector subspace of $B^{1,p}(\mathbb{R}^d)$ of all functions $f$ such that $\partial^\alpha f$ belongs to $B^{-|\alpha|,p}(\mathbb{R}^d)$, for every multi-index $\alpha$ of order $|\alpha| \leq k$, and the norm is given by

$$\|f\|_{s,p} = \left( \|f\|_p^p + \sum_{0 < |\alpha| < k} \|\partial^\alpha f\|_p^p + \sum_{|\alpha| = k} \|\partial^\alpha f\|_{s',p}^p \right)^{1/p}.$$ 

For $p = \infty$ the integral is replaced by the essential supremum and duality is used for $s < 0$, i.e., $B^{s,p}(\mathbb{R}^d)$ is the dual space of $B^{-s',p'}(\mathbb{R}^d)$ with $1/p' + 1/p' = 1$ and $1 \leq p < \infty$. 

First, note that the seminorm $|f|_{s,p}$ for $0 < s < 1$ coincides with the definition given in Chapter 4. Comparing with the definition of Sobolev spaces, this means that $W^{s,p}(\mathbb{R}^d) = B^{s,p}(\mathbb{R}^d)$ for any non-positive integer $s$, and $W^{m,p}(\mathbb{R}^d) \subset B^{m,p}(\mathbb{R}^d)$, for any positive integer $m$, but a priori, not necessarily equals. However, in view of specific estimates of the Fourier transform, for the Hiltertian case $p = 2$, both spaces are indeed equals, see the first subsection of the previous Section 6.1.

In view of the preliminaries, a function $f$ in $L^p(\mathbb{R}^d)$ belongs to the Besov space $B^{s,p}(\mathbb{R}^d)$ as in Definition 6.28 if and only if the seminorm $|f|_{s,p,p,k}$, given by (6.33), is finite, with any possible choice of $0 < s < k$. Note that our argument is not entirely self-contained, since the fact that if

$$|f|_{1+s',p} = \left( \int_{\mathbb{R}^d} dx \int_{|z| \leq 1} |f(x + z) - 2f(x) + f(x - z)|^p |z|^{-d-sp} dz \right)^{1/p}$$

is finite for some $0 < s' < 1$ then $\partial^\alpha f$ belongs to $B^{s',p}(\mathbb{R}^d)$ for any multi-index $\alpha$ with $|\alpha| = 1$, has not been proved. The use of Peetre’s scaling property to define the Besov spaces (e.g., see Bergh and Löfström [18, Section 6.2, pp. 139–146]) is very convenient to show in a direct way that $W^{s,p}(\mathbb{R}^d) \subset B^{s,p}(\mathbb{R}^d)$ if $p \geq 2$ and $W^{s,p}(\mathbb{R}^d) \supset B^{s,p}(\mathbb{R}^d)$ if $1 \leq p \leq 2$.

### 6.3.2 Bessel Potential Spaces

The (Hilbert) Sobolev spaces $H^s(\mathbb{R}^d)$ were defined in Section 6.1.1 via the Fourier transform in a way that Parseval-Plancherel’s equality makes them almost a weighted $L^2$-space. Trying to imitate this argument, and since the Fourier transform and mapping

$$f \mapsto (I - \Delta)^{s/2} f = \mathcal{F}^{-1} \left( (1 + |\cdot|^2)^{s/2} \mathcal{F}(f)(\cdot) \right)$$
are linear isomorphisms from either the smooth rapidly decreasing functions \( S(\mathbb{R}^d) \) or the tempered distributions \( S'(\mathbb{R}^d) \) onto itself, the Sobolev spaces \( H^{s,p} = H^{p,s}(\mathbb{R}^d) \) are defined as follows:

\[
    f \in H^{s,p}(\mathbb{R}^d) \quad \text{if and only if} \quad (I - \Delta)^{s/2} f \in L^p(\mathbb{R}^d),
\]

for any real number \( s \), and \( 1 \leq p \leq \infty \), with the norm \( f \mapsto \|f\|_{s,p} = \|(I - \Delta)^{s/2} f\|_p \). It is clear that with this definition, \( H^{s,p}(\mathbb{R}^d) \) is a Banach space which can be regarded as the completion of \( S(\mathbb{R}^d) \) with equivalent norm, so that \( (s,p) \) for any real number \( s \), \( 1 \leq p \leq \infty \), with the norm \( f \mapsto \|f\|_{s,p} = \|(I - \Delta)^{s/2} f\|_p \). It is clear that with this definition, \( H^{s,p}(\mathbb{R}^d) \) is a Banach space which can be regarded as the completion of \( S(\mathbb{R}^d) \) with equivalent norm, so that \( (s,p) \) for any real number \( s \), \( 1 \leq p \leq \infty \), with the norm \( f \mapsto \|f\|_{s,p} = \|(I - \Delta)^{s/2} f\|_p \). Moreover, by construction the mapping \( (I - \Delta)^{-s/2} \) is an isomorphism between \( H^{s+t,p}(\mathbb{R}^d) \) and \( H^{t,p}(\mathbb{R}^d) \), for any \( s, t \) real numbers. The paring \( \langle \cdot , \cdot \rangle \) of \( S \) and \( S' \) can be used to deduce that the dual space \((H^{s,p})'\) is indeed the space \( H^{-s,p'}\), with \( 1/p + 1/p' = 1 \). Therefore, our concern is only with the spaces \( H^{s,p} = H^{s,p}(\mathbb{R}^d) \), \( s > 0 \), since \( H^{0,p} = L^p \). Using the Bessel kernel\(^2\) \( b_s = b_{s,d} \), as in (6.19),

\[
    f \in H^{s,p}(\mathbb{R}^d), \quad s > 0 \quad \text{if and only if} \quad f = b_s * g, \quad g \in L^p(\mathbb{R}^d),
\]

and, by definition, \( \|f\|_{s,p} = \|g\|_p \). As in the hilbertian case, the key point of the project is to conclude that \( H^{s,p}(\mathbb{R}^d) = W^{s,p}(\mathbb{R}^d) \), as early defined in Section 4. Similarly, in the case of the Besov spaces, see Definition 6.28 above, the objective is to establish that

\[
    f \in B^{s,p}(\mathbb{R}^d), \quad s > 1 \quad \text{if and only if} \quad f = b_{s-1} * g, \quad g \in B^{1,p}(\mathbb{R}^d),
\]

\[
    f \in B^{s,p}(\mathbb{R}^d), \quad 0 < s < 1 \quad \text{if and only if} \quad g = b_{1-s} * f, \quad g \in B^{1,p}(\mathbb{R}^d),
\]

with equivalent norm, so that \( (I - \Delta)^{-s/2} \) is an isomorphism between \( B^{p,s+t}(\mathbb{R}^d) \) and \( B^{p,t}(\mathbb{R}^d) \), for any \( s, t \) real numbers.

As mentioned early, for any \( 0 < s' < 1 \leq p < \infty \), the norm \( \|\cdot\|_p \) in \( L^p(\mathbb{R}^d) \) and the seminorm

\[
    |f|_{s',p} = \left( \int_{|x - y| \leq 1} |f(x) - f(y)|^p |x - y|^{-d - ps'} \, dx \, dy \right)^{1/p},
\]

yield the norm \( \|\cdot\|_{s',p} = \|\cdot\|_p + |\cdot|_{s',p} \), which is used to obtain a Banach space \( W^{s',p}(\mathbb{R}^d) \) as the completion of the space \( S(\mathbb{R}^d) \), known as the \( (s',p) \)-Sobolev space. For \( p = \infty \), the integral is replaced by an essential supremum and the space \( W^{s',\infty}(\mathbb{R}^d) \) is indeed the space of bounded Hölder continuous functions \( C^\alpha(\mathbb{R}^d) \), with \( \alpha = s' \), where \( S(\mathbb{R}^d) \) is not a dense subspace. For \( s = 0 \), \( W^{0,p}(\mathbb{R}^d) \) is (by definition) the space of \( p \)-integrable functions \( L^p(\mathbb{R}^d) \). The Sobolev space \( W^{s,p}(\mathbb{R}^d) \), \( s = m + s' \), with \( k \) a positive integer and \( 0 \leq s' < 1 \) is defined by reduction, i.e., as the subspace of \( L^p(\mathbb{R}^d) \) of all functions \( f \) such that \( \partial^\alpha f \) belongs to \( W^{s-|\alpha|,p}(\mathbb{R}^d) \) for any multi-index \( \alpha \) of order \( |\alpha| \leq m \), with the norm

\[
    \|f\|_{s,p} = \left( \|f\|_p^p + \sum_{0 < |\alpha| < m} \|\partial^\alpha f\|_p^p + \sum_{|\alpha| = m} \|\partial^\alpha f\|_{s',p}^p \right)^{1/p}.
\]

\(^2\)sometimes the dimension \( d \) is omitted in the notation of the Bessel kernel.
For $p = \infty$ the integral is replaced by the essential supremum and duality is used for $s < 0$, i.e., $W^{s,p}(\mathbb{R}^d)$ is the dual space of $W^{-s,p'}(\mathbb{R}^d)$ with $1/p + 1/p' = 1$ and $1 \leq p < \infty$. The norm of the Besov spaces in Definition 6.28 has the same notation. The difference is only for integer values of $s$, where the seminorm $|·|_{s',p}$ in the Besov spaces is allowed to take the value $s' = 1$, i.e., $|·|_{1,p,p,2}$ as in (6.33).

Let us first summarize (and comment on) a number of properties already discussed, mainly on the Bessel potentials:

0.- First, for any $s > 0$,

$$\| (I - \Delta)^{-s/2} f \|_p \leq \| f \|_p, \quad \forall f \in L^p(\mathbb{R}^d), \ s > 0,$$

moreover, $(I - \Delta)^{-s/2}$ is an injective operator from $L^p(\mathbb{R}^d)$ into itself.

1.- The estimate on the partial derivative obtained in Corollary 6.20 yields

$$\| \partial_i (I - \Delta)^{-(s+1)/2} f \|_p \leq C_{s,d} \| (I - \Delta)^{-(s+1)/2} f \|_p + \| (I - \Delta)^{-s/2} f \|_p,$$

for any $f \in L^p(\mathbb{R}^d)$, $i = 1, \ldots, d$, $1 \leq p \leq \infty$, and a suitable constant $C_{s,d}$ depending only on the dimension $d$ and the order $s > 0$.

2.- As mentioned in Remark 6.22, via the $L^p$-multiplier Theorem 6.16 or the Riesz potential, and the so-called singular integrals, also the estimates

$$\| \partial_i (I - \Delta)^{-1/2} f \|_p \leq C \| f \|_p, \quad \forall f \in L^p(\mathbb{R}^d), \ i = 1, \ldots, d,$$

$$\| (I - \Delta)^{-s/2} f \|_{s,p} \leq C \| f \|_p, \quad \forall f \in L^p(\mathbb{R}^d), \ 0 < s < 1,$$

hold true, for any $1 < p < \infty$ and some suitable constant $C$ independent of $f$, where $|·|_{s,p}$ is the seminorm (6.36) with $s' = s$. The cases $p = 1$ or $p = \infty$ needs special treatment. In any way, details on the proofs of these estimates should be found in more advanced textbooks, as mentioned early.

3.- Besides the previous estimates just mentioned, we need

$$\| \partial_i (I - \Delta)^{-s} f \|_{s-1,p} \leq C \| f \|_{B^{1,p}}, \quad \forall f \in B^{1,p}(\mathbb{R}^d), \ 1 \leq i \leq d, \ s > 1,$$

$$\| (I - \Delta)^{1-s} f \|_{1,p} \leq C \| f \|_{W^{s,p}}, \quad \forall f \in W^{s,p}(\mathbb{R}^d), \ 0 < s < 1,$$

to handle the Besov spaces. Certainly, this involves singular integrals.

4.- Referring to (6.32), apply the difference operator $\Delta^k_z$ to the convolution $b_s * f$ to deduce $\| \Delta^k_z(b_s * f) \|_p \leq \| \Delta^k_z f \|_p$, which yields

$$\| (I - \Delta)^{-s} f \|_{t,p,k} \leq \| f \|_{t,p,k}, \quad \forall f \in L^p(\mathbb{R}^d), \ 0 < t < k, \ s > 0,$$

where $|·|_{t,p,k} = |·|_{t,p,p,k}$ is the Besov seminorm as defined by (6.33). However, singular integrals and/or $L^p$-multipliers are used to obtain the estimate

$$\| (I - \Delta)^{-s} f \|_{t+s,p,k} \leq C \| f \|_{t,p,k}, \quad \forall f \in L^p(\mathbb{R}^d), \ 0 < t + s < k, \ s > 0,$$

for a suitable constant $C$ depending only on the dimension $d$ and $s, t$. 

and many other properties and comments can be added.

**Theorem 6.29.** The Sobolev spaces \( H^{s,p}(\mathbb{R}^d) \) defined by (6.34) and \( W^{s,p}(\mathbb{R}^d) \) defined in Chapter 4 are the same with equivalent norms. For any \( s \) non-integer, the Besov space \( B^{s,p}(\mathbb{R}^d) \) is equal to the Sobolev space \( H^{s,p}(\mathbb{R}^d) \), while, for any \( s = m \geq 1 \) integer \( B^{m,p}(\mathbb{R}^d) \subset H^{m,p}(\mathbb{R}^d) \) if \( 1 \leq p \leq 2 \) and \( H^{m,p}(\mathbb{R}^d) \subset B^{m,p}(\mathbb{R}^d) \) if \( p \geq 2 \). Moreover, the linear mapping \((I - \Delta)^{s/2}\) is an isomorphism between \( H^{s+1,p}(\mathbb{R}^d) \) and \( H^{1,p}(\mathbb{R}^d) \), and also between \( B^{s+1,p}(\mathbb{R}^d) \) and \( B^{1,p}(\mathbb{R}^d) \). Furthermore, if \( 1 \leq p < \infty \) then the test functions \( \mathcal{D}(\mathbb{R}^d) \) are dense in either \( H^{s,p}(\mathbb{R}^d) \) or \( B^{s,p}(\mathbb{R}^d) \), and the dual space of \( H^{s,p}(\mathbb{R}^d) \) is the space \( H^{-s,p'}(\mathbb{R}^d) \) and the dual space of \( B^{s,p}(\mathbb{R}^d) \) is the space \( B^{-s,p'}(\mathbb{R}^d) \), with \( 1/p + 1/p' = 1 \).

**Proof Sketch.** Full details take a lot of work, for instance, the reader is referred to the books by Bergh and L"ofstr"om [18, Sections 6.2-6.4, pp. 139–153], Stein [113, Chapter V, pp. 116–165], Triebel [126, Chapter 2, pp. 151–244], and Peetre [99]. The papers Aronszajn and Smith [10] and Adams et al. [1] could be taken as key references. The reader may want to check also the textbook Leoni [79, Chapter 14, pp. 414-450] and Ziemer [140, Chapter 2, pp. 42–111].

Sobolev and Besov spaces were developed to study PDEs, and most of the statements in the Theorem have to do with a priori estimates. Thus, we are going to report only a couple of points. Actually, most of the techniques shown in the previous section on Riesz and Bessel potentials were developed to facilitate the understanding of this statements. Moreover, most estimates related to singular integrals (such as the \( L^p \)-multipliers) are only quoted.

For instance, if \( 1 < p < \infty \) and \( s \geq 1 \) then \( f \) belongs to \( H^{s,p}(\mathbb{R}^d) \) then \( f \) and \( \partial_j f \) belong to \( H^{s-1,p}(\mathbb{R}^d) \), for any \( j = 1, \ldots, d \). Indeed, \( f \) belongs to \( H^{s,p} \) if and only if \( f = b_s \ast g \) with \( g \) in \( L^p \). Use the Riesz transform \( R_j \) given by (6.22) and the \( L^p \)-multiplier \((-\Delta)^{1/2}(I - \Delta)^{-1/2}\), see Remark 6.17, to represent the partial derivative \( \partial_j f = -R_j(-\Delta)^{1/2}(I - \Delta)^{-1/2}g \) when \( s = 1 \), while for \( s > 1 \), Corollary 6.20 yields \(|\partial_j f| \leq C|b_{s-1} \ast [g + b_1 \ast g]| \). This shows that if \( f \) belongs to \( H^{s,p} \) then \( f \) and \( \partial_j f \) belong to \( H^{s-1,p} \).

Conversely, if \( \varphi \) belongs to \( \mathcal{S} \) then \( \varphi = b_1 \ast \psi \) for some \( \psi \) in \( \mathcal{S} \). By means of a density argument and the estimate \( \|\psi\|_p \leq C\|\varphi\|_{1,p} \), we deduce that if \( g \) and \( \partial_j g \) belong to \( H^{1,p} \) for any \( j = 1, \ldots, d \) then \( g = b_1 \ast h \) for some \( h \) in \( L^p \) satisfying \( \|h\|_p \leq C\|g\|_{1,p} \). The convolution property of the Bessel kernels yields \( f = b_{s-1} \ast g = b_s \ast h \) and \( \|f\|_{s,p} = \|h\|_p \leq C\|g\|_{1,p} \), which proves that \( f \) belongs to \( H^{1,p} \). Moreover, the norms \( \|f\|_{H^{s-1,p}} + \sum_j \|\partial_j f\|_{H^{s-1,p}} \) and \( \|f\|_{H^{s,p}} \) are equivalent.

These assertions reduce the proof to the case \( 0 < s < 1 \) for the Sobolev spaces and to \( 0 < s \leq 1 \) for the Besov spaces. In Remark 6.22 we gave some argument related to the estimate \( |b_s \ast \varphi|_{s,p} \leq C\|\varphi\|_p \). Moreover, a couple of similar estimates are also necessary to complete the assertion that \((I - \Delta)^{s/2}\) is a isomorphism between Sobolev or Besov spaces.

Except the inclusion relation between Sobolev and Besov spaces (which requires a special consideration), all other assertions follow from this isomorphism property.
6.3.3 Traces on Besov Spaces

Essentially most results obtained for the hilbertian case can be extended to the spaces $H^{s,p}$ and $B^{s,p}$. However, the arguments are much harder in general, and only some of them are briefly discussed below.

As in Definition 6.5, if $\Omega$ is an open subset of $\mathbb{R}^d$ then $H^{s,p}_0(\Omega)$ is the closure of the test functions $\mathcal{D}(\Omega)$ in $H^{s,p}(\mathbb{R}^d)$, while $H^{s}(\Omega)$ denotes the linear space of all restriction to $\Omega$ of tempered distributions in $H^{s,p}(\mathbb{R}^d)$, i.e., an element $f$ in $H^{s,p}(\Omega)$ is an element in $\mathcal{D}'(\Omega)$ which can be extended to become an element $f^e$ in $H^{s,p}(\mathbb{R}^d)$. In other words, if $|\cdot|_{\Omega}$ denotes the restriction to $\Omega$ and it is regarded as an operator from $H^{s,p}(\mathbb{R}^d)$ into $\mathcal{D}'(\Omega)$ then its kernel is the closed linear space $H^{s,p}_0(\mathbb{R}^d \setminus \Omega)$ and $H^{s,p}(\Omega)$ is the quotient space $H^{s,p}(\mathbb{R}^d)/H^{s,p}_0(\mathbb{R}^d \setminus \Omega)$. Clearly, with the quotient norm

$$\|f\|_{H^{s,p}(\Omega)} = \inf\{\|f^e\|_{H^{s,p}(\mathbb{R}^d)} : f^e \in H^{s,p}(\mathbb{R}^d), \ f^e|_{\Omega} = f\},$$

this is a Banach space. In particular, this definition yields the density of the restriction of test functions, i.e., $H^{s,p}(\Omega)$ is the closure of the linear subspace $\mathcal{D}(\mathbb{R}^d)|_{\Omega}$.

To reconcile this definition with Section 4, we need to study the extension operator, i.e., a linear continuous mapping $E$ from $W^{s,p}(\Omega)$ into $W^{s,p}(\mathbb{R}^d)$ satisfying $Ef = f$ in $\Omega$. In other words, if $R$ is the restriction (from a function defined on $\mathbb{R}^d$ to a function defined on $\Omega$) operator then how smooth $\Omega$ should be to ensure the equality $R(W^{s,p}(\mathbb{R}^d)) = W^{s,p}(\Omega)$, i.e., $H^{s,p}(\mathbb{R}^d) = W^{s,p}(\Omega)$.

As discussed in Stein [113, Chapter VI, pp. 166–195], this is true for Lipschitz domains, i.e., typically of the form $\Omega = \{(x',x_d) \in \mathbb{R}^d : x_d > \phi(x')\}$ for some Lipschitz function $\phi$. However, the arguments are easier when the domain is of class $C^{k,1}$, $k \geq 0$ an integer such that $k < s \leq 1 + k$, e.g., see Nečas [93, Section 2.3, pp. 60-77]. Certainly, this is also valid for the Besov spaces, i.e., an extension operator is necessary to compare $B^{1,p}(\mathbb{R}^d) = R(B^{1,p}(\mathbb{R}^d))$ with the space of all functions $f$ in $L^p(\Omega)$ such that the integral

$$\int_{\Omega} dx \int_{\{z: |z| \leq 1, x,x \pm z \in \Omega\}} |f(x+z) - 2f(x) + f(x-z)|^p |z|^{-d-2s} dz$$

is finite.

To define the Besov (or Sobolev) spaces on the boundary, we restate Definition 6.13

**Definition 6.30** (Besov/Boundary). Let $\Omega$ be a smooth domain in $\mathbb{R}^d$ of class $C^{m,1}$, $m \geq 0$, with a bounded boundary and hypographs $\phi_1, \ldots, \phi_n$ and associated smooth partition of the unity $\chi_1, \ldots, \chi_n$. If $0 \leq s \leq 1 + m$ then a function $f$ belongs to the Besov space $B^{s,p}(\partial \Omega)$ if and only if the function $f_k : y' \mapsto f(y', \phi_k(y')) \chi_k(y', \phi_k(y'))$ is in $B^{s,p}(\mathbb{R}^{d-1})$, for every $k$, see Definition 6.28. The norm is defined through the expression $\|f\|_{B^{s,p}(\partial \Omega)} = \sum_k \|f_k\|_{B^{s,p}(\mathbb{R}^{d-1})}$. □

Certainly, a definition as the above is used for the Sobolev spaces, replacing $B^{s,p}$ with $H^{s,p}$. It should be clear that $B^{s,p}(\partial \Omega)$ and $H^{s,p}(\partial \Omega)$ are Banach

spaces, that
\[
L^2(\partial \Omega) \simeq B^{0,p}(\partial \Omega) \simeq H^{0,p}(\partial \Omega),
\]
\[
B^{t,p}(\partial \Omega) \subset B^{s,p}(\partial \Omega) \text{ and } H^{t,p}(\partial \Omega) \subset H^{s,p}(\partial \Omega), \quad \forall s \leq t
\]
and that the definitions are independent of the particular hypographs and smooth partition of the unity used, see the beginning of Subsection 6.1.3.

Without repeating Subsection 6.1.3, we want to discuss a trace of a function in either \(B^{s,p}(\Omega)\) or \(H^{s,p}(\Omega)\), see Bergh and Lofström [18, Section 6.6, pp. 155–156], Triebel [127, Sections 2.5.7, 2.7.2 and 3.3.3, pp. 87–89, 131–139 and 199–202], as well as Leoni[79, Chapter 15, pp. 451-476], Nečas [93, Sections 2.4 and 2.5, pp. 77–102].

As mentioned early, the study of the trace operator for a sufficiently smooth domain \(\Omega\) is reduce to the case \(\Omega = \mathbb{R}_+^d\) with an argument of local coordinates. Thus, if \(n\) denotes the exterior unit normal direction on the boundary \(\partial \Omega\) then the trace operators
\[
T_0(\varphi) = \varphi|_{\partial \Omega}, \quad T_1(\varphi) = (\partial_n \varphi)|_{\partial \Omega}, \quad \ldots, \quad T_m(\varphi) = (\partial_n^m \varphi)|_{\partial \Omega}
\]
are defined for any smooth function \(\varphi\) defined on \(\Omega\) and any domain of class \(C^{m,1}\), for any nonnegative integer \(m\).

**Theorem 6.31.** If \(s > k + 1/p\) then the trace operator \(T_k\) given by (6.37) can be extended as a continuous linear operator from \(H^{s,p}(\Omega)\) or \(B^{s,p}(\Omega)\) onto \(B^{s-1/p,p}(\partial \Omega)\), \(1 < p < \infty\). Moreover, if \(m + 1/2 < s < m + 3/2\) with a positive integer \(m\) then the normal derivatives and trace operator
\[
f \mapsto T(f) = (f|_{\partial \Omega}, (\partial_n f)|_{\partial \Omega}, \ldots, (\partial_n^m f)|_{\partial \Omega})
\]
from \(H^{s,p}(\Omega)\) or \(B^{s,p}(\Omega)\) onto \(B^{s-1/2}(\partial \Omega) \times B^{s-1/2}(\partial \Omega) \times \cdots \times B^{s-1/2-m}(\partial \Omega)\), is also a continuous linear operator.

**Proof.** No mention to the extremes \(p = 1\) and \(p = \infty\), which require special treatment. As in Theorem 6.29, only certain aspects of the proof are given in detail. Every assertion is a matter of showing suitable estimates, so that the arguments in Theorem 6.11 (related to the hilbertian case \(p = 2\)) can be extended to this new situation. Except for the construction of the lift operators, only \(k = 0\), \(1/p < s \leq 1 < p < \infty\) need to be considered.

To prepare the arguments, recall Minkowski inequality for integrals, namely
\[
\left( \int_X \left( \int_Y |f(x,y)| \nu(dy) \right)^p \mu(dx) \right)^{1/p} \leq \int_Y \left( \int_X |f(x,y)|^p \mu(dx) \right)^{1/p} \nu(dy),
\]
and Hardy’s inequality (see Proposition 4.25), i.e.,
\[
\int_0^\infty \left[ t^\gamma \int_t^\infty \beta(\tau) d\tau \right]^{p/dt} \leq \gamma^{-p} \int_0^\infty \left[ t^{\gamma+1} \beta(t) \right]^{p/dt}, \quad \gamma > 0,
\]
\[
\int_0^\infty \left[ t^\gamma \int_0^t \beta(\tau) d\tau \right]^{p/dt} \leq (-\gamma)^{-p} \int_0^\infty \left[ t^{\gamma+1} \beta(t) \right]^{p/dt}, \quad \gamma < 0,
\]
for any $1 \leq p < \infty$, $\beta \geq 0$ Lebesgue integrable.

Firstly, note that if $f$ belongs to $H^{s,p}(\mathbb{R}^d)$ then there exists $\varphi$ in $L^p(\mathbb{R}^d)$ such that $f = b_{s,d} \ast \varphi$, $\|f\|_{H^{s,p}(\mathbb{R}^d)} = \|\varphi\|_{L^p(\mathbb{R}^d)}$, and $T_0(f) = (b_{s,d} \ast \varphi)(\cdot, 0)$, where $b_{s,d}$ is the Bessel kernel given by (6.19).

To check that $T_0$ maps $H^{s,p}(\mathbb{R}^d_+)$ or $B^{s,p}(\mathbb{R}^d_+)$ into $B^{s-1/p,p}(\mathbb{R}^{d-1})$ several estimates are necessary, the first estimate is relatively simple to obtain,

$$\|(b_{s,d} \ast \varphi)(\cdot, 0)\|_{L^p(\mathbb{R}^{d-1})} \leq C \|\varphi\|_{L^p(\mathbb{R}^d)}, \quad s > \frac{1}{p}, \quad (6.38)$$

for a suitable constant $C$ depending only on $s$, $p$ and the dimension $d$.

Indeed, use Hölder inequality with $1/p' + 1/p = 1$ to obtain

$$\int_{\mathbb{R}} |b_{s,d}(y', y_d)\varphi(x' - y', -y_d)| \, dy_d \leq \|b_{s,d}(y', \cdot)\|_{L^{p'}(\mathbb{R})} \|\varphi(x' - y', \cdot)\|_{L^p(\mathbb{R})},$$

which yields

$$| (b_{s,d} \ast \varphi)(x', 0) | \leq \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} b_{s,d}(y', y_d)\varphi(x' - y', y_d) \, dy_d \, dy' \leq \int_{\mathbb{R}^{d-1}} \|b_{s,d}(y', \cdot)\|_{L^{p'}(\mathbb{R})} \|\varphi(x' - y', \cdot)\|_{L^p(\mathbb{R})} \, dy'.$$

Hence, note that

$$\|\varphi\|_{L^p(\mathbb{R}^d)} = \left( \int_{\mathbb{R}^{d-1}} \|\varphi(x' - y', \cdot)\|^p_{L^{p'}(\mathbb{R})} \, dx' \right)^{1/p}$$

and apply Minkowski inequality for integrals to deduce

$$\|(b_{s,d} \ast \varphi)(\cdot, 0)\|_{L^p(\mathbb{R}^{d-1})} \leq \int_{\mathbb{R}^{d-1}} \|b_{s,d}(y', \cdot)\|_{L^{p'}(\mathbb{R})} \|\varphi\|_{L^p(\mathbb{R}^d)} \, dy',$$

i.e., estimate (6.38) holds with

$$C = \int_{\mathbb{R}^{d-1}} \left( \int_{\mathbb{R}} |b_{s,d}(y', y_d)|^{p'} \, dy_d \right)^{1/p'} \, dy'.$$

as long as $C$ is finite. To verify this, use again Minkowski inequality for integrals to obtain

$$\left( \int_{\mathbb{R}} |b_{s,d}(y', y_d)|^{p'} \, dy_d \right)^{1/p'} \leq \frac{1}{\Gamma(s/2)} \int_0^\infty e^{-t} t^{s/2 - 1} \|h_d(y', \cdot, t)\|_{L^{p'}(\mathbb{R})} \, dt,$$

where $h_d(y', y_d, t) = (4\pi t)^{-d/2} e^{y'^2/(4t)}$ is the $d$-dimensional heat-kernel. Use the equalities $h_d(y', y_d, t) = h_{d-1}(y', y_d, t)h_1(y_d, t)$, $-1 + 1/p' = -1/p$, and

$$\left( \int_{\mathbb{R}} |h_1(y_d, t)|^{p'} \, dy_d \right)^{1/p'} = (p')^{1/p'} (4\pi t)^{-1+1/p'},$$
to get $C = (p')^{1/p'}(4\pi)^{-1/p}\Gamma(s - 1/p)/\Gamma(s/2)$.

The next estimate is related to the seminorm $|\cdot|_{s,p}$ given by (6.36) in $\mathbb{R}^{d-1}$, namely,

$$
|\langle b_{s,d} \ast \varphi \rangle(\cdot,0)|_{W^{s-1/p,p}(\mathbb{R}^{d-1})} \leq C\|\varphi\|_{L^p(\mathbb{R}^d)}, \quad \frac{1}{p} < s < 1 + \frac{1}{p},
$$

(6.39)

for a suitable constant $C$ depending only on $s$, $p$ and the dimension $d$, where

$$
|g|^p_{W^{s-1/p,p}(\mathbb{R}^{d-1})} = \int_{\mathbb{R}^{d-1}} \int_{|z'| \leq 1} |g(x' + z') - g(x')|^p |z|^{-(d-1) - sp + 1} dz,
$$

with $0 < s < 1$.

To this end, consider the expressions

$$
g(x', y_d) = \int_{\mathbb{R}^{d-1}} b_{s,d}(y', y_d) \varphi(x' - y', y_d) dy',
$$

$$(b_{s,d} \ast \varphi)(x', 0) = \int_{\mathbb{R}} g(x', y_d) dy_d,$$

which yield

$$(b_{s,d} \ast \varphi)(x' + z', 0) - (b_{s,d} \ast \varphi)(x', 0) = \int_{\mathbb{R}} [g(x' + z', y_d) - g(x', y_d)] dy_d,$$

$$g(x' + z', y_d) - g(x', y_d) = \int_{\mathbb{R}^{d-1}} [b_{s,d}(y' - z', y_d) - b_{s,d}(y', y_d)] \varphi(x' - y', y_d) dy',$$

after a change of variables in $\mathbb{R}^{d-1}$. Now, note that $\|\varphi(\cdot - y', y_d)\|_{L^p(\mathbb{R}^{d-1})} = \|\varphi(\cdot; y_d)\|_{L^p(\mathbb{R}^{d-1})}$ and apply Minkowski inequality for integrals to deduce

$$
\|g(\cdot + z', y_d) - g(\cdot; y_d)\|_{L^p(\mathbb{R}^{d-1})} \leq \|\varphi(\cdot; y_d)\|_{L^p(\mathbb{R}^{d-1})} \int_{\mathbb{R}^{d-1}} |b_{s,d}(y' - z', y_d) - b_{s,d}(y', y_d)| dy'.
$$

Hence, use Proposition 6.23 to bound

$$
\int_{\mathbb{R}^{d-1}} |b_{s,d}(y' - z', y_d) - b_{s,d}(y', y_d)| dy' \leq C \min\{|y_d|^{s-1}, |y_d|^{s-2}|z'|\},
$$

and to obtain

$$
\|\langle b_{s,d} \ast \varphi \rangle(\cdot + z', 0) - \langle b_{s,d} \ast \varphi \rangle(\cdot, 0)\|_{L^p(\mathbb{R}^{d-1})} \leq \int_{\mathbb{R}} \|g(\cdot + z', y_d) - g(\cdot; y_d)\|_{L^p(\mathbb{R}^{d-1})} dy_d \leq \int_{|y_d| \geq |z'|} \|\varphi(\cdot; y_d)\|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-2}|z'| dy_d + \int_{|y_d| < |z'|} \|\varphi(\cdot; y_d)\|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-1} dy_d,
$$

\[258\] Chapter 6. Besov and Sobolev Spaces
which implies
\[
\left\| (b_{s,d} \ast \varphi)(\cdot + z', 0) - (b_{s,d} \ast \varphi)(\cdot, 0) \right\|_{L^p(\mathbb{R}^{d-1})}^p \leq \\
\leq C \left[ \int_{|y_d| \geq |z'|} \left| \varphi(\cdot, y_d) \right|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-2} |z'|^{1-s-(d-2)/p} dy_d + \\
+ \int_{|y_d| < |z'|} \left| \varphi(\cdot, y_d) \right|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-1} |z'|^{1-s-(d-2)/p} dy_d \right].
\]

Denote by $C[A + B]$ the right-hand term, note that $A$ and $B$ are radial functions of $z'$, and use spherical/polar coordinate to integrate in $dz'$ within the region $\{|z'| \leq 1\}$ and to check that
\[
\int_{\{|z'| \leq 1\}} (C[A + B])^p dz' \leq 2^{p-1} C^p \omega_{d-2} \int_0^1 (A^p + B^p) \rho^{d-2} d\rho,
\]
where $\omega_{d-2}$ is the area of the unit sphere $\{|z'| = 1\}$ in $\mathbb{R}^{d-1}$. Therefore
\[
\int_{\mathbb{R}^{d-1}} \frac{\left\| (b_{s,d} \ast \varphi)(\cdot + z', 0) - (b_{s,d} \ast \varphi)(\cdot, 0) \right\|_{L^p(\mathbb{R}^{d-1})}^p}{|z'|^{sp+(d-2)}} dz' \leq \\
\leq C' \int_0^1 \left[ \rho^{1-s+1/p} \int_{|y_d| \geq \rho} \left| \varphi(\cdot, y_d) \right|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-2} dy_d \right] \frac{d\rho}{\rho} + \\
+ C' \int_0^1 \left[ \rho^{-s+1/p} \int_{|y_d| < \rho} \left| \varphi(\cdot, y_d) \right|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-1} dy_d \right] \frac{d\rho}{\rho},
\]
with $C' = 2^{p-1} C^p \omega_{d-2}$. Thus, write the integrals in $|y_d| \geq \rho$ and in $|y_d| < \rho$ as
\[
\int_{\rho}^\infty \beta(y_d) y_d^{s-2} dy_d \quad \text{and} \quad \int_{0}^\rho \beta(y_d) y_d^{s-1} dy_d,
\]
with $\beta(y_d) = \left| \varphi(\cdot, y_d) \right|_{L^p(\mathbb{R}^{d-1})} + \left| \varphi(\cdot, -y_d) \right|_{L^p(\mathbb{R}^{d-1})}$,
and use Hardy’s inequality with $\gamma = 1 - s + 1/p > 0$ for the integral on $(\rho, \infty)$ and with $\gamma = -s + 1/p < 0$ for the integral on $(0, \rho)$ to deduce
\[
\int_0^1 \left[ \rho^{1-s+1/p} \int_{|y_d| \geq \rho} \left| \varphi(\cdot, y_d) \right|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-2} dy_d \right] \frac{d\rho}{\rho} \leq \\
\leq \gamma^{-p} \int_0^\infty \left[ \rho^{1-s+1/p} \rho^{-2+1} \beta(\rho) \right] \frac{d\rho}{\rho} = \gamma^{-p} \int_0^\infty (\beta(y_d))^p dy_d
\]
and
\[
\int_0^1 \left[ \rho^{-s+1/p} \int_{|y_d| \geq \rho} \left| \varphi(\cdot, y_d) \right|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-1} dy_d \right] \frac{d\rho}{\rho} \leq \\
\leq |\gamma|^{-p} \int_0^\infty \left[ \rho^{-s+1/p} \rho^{-1+1} \beta(\rho) \right] \frac{d\rho}{\rho} = |\gamma|^{-p} \int_0^\infty (\beta(y_d))^p dy_d.
\]
Collecting all pieces, the estimate (6.39) follows.

Now, the case $s - 1/p = 1$ should be considered, i.e., the estimate

$$
|\langle b_{s,d} \ast \varphi \rangle(\cdot, 0)\rangle_{B^{1,p}(\mathbb{R}^{d-1})} \leq C\|\varphi\|_{L^p(\mathbb{R}^d)}, \quad 1 + \frac{1}{p} = s,
$$

(6.40)

for a suitable constant $C$ depending only on $p$ and the dimension $d$, where

$$
|g|_{B^{1,p}(\mathbb{R}^{d-1})}^p = \int_{\mathbb{R}^{d-1}} \int_{|z'| \leq 1} |g(x' + z') - 2g(x') + g(x - z')|p|z'|^{-(d-1) - p} \, dz',
$$

see Definition 6.28, with $d$ replaced by $d - 1$.

The arguments are similar to those used for the previous estimate (6.39), but apply Remark 6.24 instead of Proposition 6.23, i.e., the bound

$$
\int_{\mathbb{R}^{d-1}} |b_{s,d}(y' + z', x_d) - 2b_{s,d}(y', x_d) + b_{s,d}(y' - z', y_d)| \, dy' \leq C \min \left\{ |y_d|^{s-1}, |y_d|^{s-3}|z'|^2 \right\}, \quad 0 < s < 2,
$$

for some constant $C$ depending only on the dimension $d$, to obtain

$$
\|\langle b_{s,d} \ast \varphi \rangle(\cdot + z', 0) - 2\langle b_{s,d} \ast \varphi \rangle(\cdot, 0) + \langle b_{s,d} \ast \varphi \rangle(\cdot - z', 0)\|_{L^p(\mathbb{R}^{d-1})} \leq C \left[ \int_{|y_d| \geq |z'|} \|\varphi(\cdot, y_d)\|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-3}|z'|^{2-s-(d-2)/p} \, dy_d + \int_{|y_d| < |z'|} \|\varphi(\cdot, y_d)\|_{L^p(\mathbb{R}^{d-1})} |y_d|^{s-1}|z'|^{-s-(d-2)/p} \, dy_d \right].
$$

As early, invoke Hardy’s inequality and write $g = \langle b_{s,d} \ast \varphi \rangle(\cdot, 0)$ to deduce

$$
\int_{\mathbb{R}^{d-1}} \int_{|z'| \leq 1} |g(x' + z') - 2g(x') + g(x - z')|p|z'|^{-(d-1) - sp} \, dz' = |\langle b_{s,d} \ast \varphi \rangle(\cdot, 0)\rangle_{B^{s,p}(\mathbb{R}^{d-1})}^p \leq C^p \|\varphi\|_{L^p(\mathbb{R}^d)}, \quad 0 < s < 2,
$$

some suitable constant $C$ depending only on $s$, $p$ and the dimension $d$, i.e., in particular, estimate (6.40) with $s = 1 + 1/p < 2$.

These three estimates cover all case for $H^{s,p}(\mathbb{R}^d)$ for $1 < p < \infty$ and $1/p < s \leq 1 + 1/p$. Now, if $s > 1 + 1/p$ and $f$ belongs to $H^{s,p}(\mathbb{R}^d)$ then $\partial_i f$ belongs to $H^{s-1,p}(\mathbb{R}^d)$, and an iteration of the argument complete the proof.

Since the norms $\|\cdot\|_{H^{p,s}}$ and $\|\cdot\|_{B^{p,s}}$ are equivalent for any non-integer $s$, for the Bessel spaces $B^{s,p}(\mathbb{R}^d)$ all cases are covered as above, but the case $s = 1$. Note that in proving this last estimate (6.40), we obtained

$$
|\langle b_{s,d} \ast \varphi \rangle(\cdot, 0)\rangle_{B^{s-1/p,p}(\mathbb{R}^{d-1})} \leq C\|\varphi\|_{L^p(\mathbb{R}^d)}, \quad 0 < s < 2,
$$
or equivalently

\[ |T(f)|_{B^{s-1/p,p}(\mathbb{R}^{d-1})} \leq C \|f\|_{H^{s,p}(\mathbb{R}^d)}, \quad 0 < s < 2, \]

where the constant \( C \) depends only on \( s, p \) and the dimension \( d \). However, the estimate

\[ |T(f)|_{B^{1-1/p,p}(\mathbb{R}^{d-1})} \leq C \|f\|_{B^{1,p}(\mathbb{R}^d)}, \]

is needed to cover the case \( s = 1 \). If \( 1 < p \leq 2 \) then \( B^{1,p}(\mathbb{R}^d) \subset H^{1,p}(\mathbb{R}^d) \) with continuous inclusion, so that this estimate follows from (6.40). For \( 2 < p < \infty \), some interpolation arguments could be used to conclude, e.g., see Bergh and L"ofstr"om [18, Theorem 6.6.1, pp. 165–166]).

To show that the trace operator is surjective, the reader may consult the books Nečas [93, Sections 2.4 and 2.5, pp. 77–102], Triebel [127, Sections 2.7.2, pp. 131–139], among others.

\[ \square \]
Appendix A

Exercises - Chapter (1) Abstract Integration

All exercises are re-listed here, but now, most of them have a (possible) solution. Certainly, this is not for the first reading. This part is meant to be read after having struggled (a little) with the exercises. Sometimes, there are many ways of solving problems, and depending of what was developed “in the theory”, solving the exercises could have alternative ways. In any way, some exercises are trivial while other are not simple. It is clear that what we may call “Exercises” in one textbook could be called “Propositions” in others.

(1.1) Daniell Integrals

Exercise 1.1. Fill in the details of the previous Remark 1.4. Moreover, consider in great detail the case of the Lebesgue-Stieltjes measures in \( \mathbb{R} \), i.e., \( I(1_{(a, b]}) = F(b) - F(a) \) for a given right-continuous increasing real-valued function. Furthermore, discuss the changes necessary to extend the arguments used for the length of a interval to the case of the hyper-volume of a \( d \)-dimensional interval (i.e., Lebesgue measure).

Proof. Consider the case of the Lebesgue-Stieltjes measure in \( \mathbb{R} \), i.e., \( E \) is the vector lattice space of step functions \( \varphi = \sum_{i=1}^{n} \alpha_i 1_{(a_i, b_i]} \) and the functional \( I \) is defined by linearity \( I(\varphi) = \sum_{i=1}^{n} \alpha_i I(1_{(a_i, b_i]}) \) with \( I(1_{(a_i, b_i]}) = F(b_i) - F(a_i) \) and \( F : \mathbb{R} \to \mathbb{R} \) is a given right-continuous increasing function.

It is simple to show that \( I \) is well defined, monotone and linear on \( E \), i.e.,

(a) if \( \varphi = \sum_{i=1}^{n} \alpha_i 1_{(a_i, b_i]} = \sum_{i=1}^{n} \alpha'_i 1_{(a'_i, b'_i]} \) then

\[
I(\varphi) = \sum_{i=1}^{n} \alpha_i [F(b_i) - F(a_i)] = \sum_{i=1}^{n} \alpha'_i [F(b'_i) - F(a'_i)],
\]

(b) if \( \varphi \leq \psi \) then \( I(\varphi) \leq I(\psi) \) and (c) if \( \alpha \) and \( \beta \) are constant then \( I(\alpha \varphi + \beta \psi) = \alpha I(\varphi) + \beta I(\psi) \).
To check the continuity condition (1.2-c), let \( \{ \varphi_n \} \) be a decreasing sequence of functions in \( E \) with \( \lim \varphi_n(x) = 0 \) for every \( x \). This implies that \( 0 \leq \varphi_n(x) \leq C \), for every \( x \in \mathbb{R} \), \( \varphi_n(x) = 0 \) for every \( x \) outside of a compact interval \( [a,b] \), and for every given \( \varepsilon > 0 \) and for every \( x \) in \( \mathbb{R} \) there exists and index \( \eta(x,\varepsilon) \) such that \( 0 \leq \varphi_n(x) \leq \varepsilon \) for any \( n \geq \eta(x,\varepsilon) \). Since each step function \( \varphi_n \) is discontinuous (actually has a jump) only at a finite number points (namely \( a_i \) and \( b_i \)), for any \( \varepsilon > 0 \) there exits a sequence containing all points of discontinuity for any \( \varphi_n \), relabeled \( \{ r_k \} \subset (a,b) \), and in view of the continuity form the right of \( F \) and the continuity from the left of \( x \mapsto F(x-) \), there exist two sequences sequences \( \{ p_k \} \) and \( \{ q_k \} \) such that \( a \leq p_k < r_k < q_k \leq b \) and \( 0 \leq [F(r_k) - F(p_k-)] + [F(q_k) - F(r_k)] < 2^{-k} \). Therefore, \( [F(q_k) - F(p_k-)] \leq 2^{-k} + [F(r_k) - F(r_k-)] \)

\[
\sum_k [F(q_k) - F(p_k-)] \leq 1 + \sum_k [F(r_k) - F(r_k-)] \leq 1 + [F(b) - F(a)],
\]

the series converges and so, there exist an index \( \kappa = \kappa(\varepsilon) \) such that the reminder \( \sum_{k>\kappa}[F(q_k) - F(p_k-)] < \varepsilon \).

On the other hand, for any point \( x \) of continuity \( x \) (i.e., in \( [a,b] \setminus \{ r_k \} \)) there exists an open interval \( I(x,n,\varepsilon) \) containing \( \{ x \} \) such that \( |\varphi_n(y) - \varphi_n(x)| \leq \varepsilon/2 \). Take \( J(x,\varepsilon) = I(x,\eta(x,\varepsilon),\varepsilon) \), to deduce that \( 0 \leq \varphi_n(y) \leq \varphi_n(x) + [\varphi_n(y) - \varphi_n(x)] \leq \varepsilon \), for every \( y \) in \( J(x,\varepsilon) \) and any \( n \geq \eta(x,\varepsilon) \). Since this family of intervals \( \{ J(x,\varepsilon) \} \) forms an open cover of the compact set \( K = [a,b] \setminus \bigcup_k (p_k,q_k) \), there exists a finite cover, i.e., \( x_1(\varepsilon), \ldots, x_j(\varepsilon) \) such that \( K \subset \bigcup_{i=1}^j J(x_i,\varepsilon) \).

Now, if \( N(\varepsilon) \) is the maximum index of \( \eta(x_i,\varepsilon) \), \( i = 1, \ldots, j \) and \( \eta(r_k), k = 1, \ldots, \kappa(\varepsilon) \) then \( 0 \leq \varphi_n(x) \leq \varepsilon \), for any \( x \) not in \( \{ r_k : k > \kappa(\varepsilon) \} \) and \( n \geq N(\varepsilon) \). Hence,

\[
\varphi_n(x) \leq \varepsilon \mathbb{1}_{[a,b]} + C\psi_{n,\varepsilon}(x), \quad \forall x,
\]

(A.1)

where \( \psi_{n,\varepsilon}(x) = 1 \) if \( x \) belongs to \( (p_k,q_k) \) for some \( k > \kappa(\varepsilon) \) with \( r_k \) being a point of discontinuity for \( \varphi_n \), and \( \psi_{n,\varepsilon}(x) = 0 \) otherwise. Because \( I(\psi_n) \) is bounded by a finite sum of terms of the form \( [F(q_k) - F(p_k-)] \) with \( k > \kappa(\varepsilon) \), this inequality yields

\[
I(\varphi_n) \leq \varepsilon I(\mathbb{1}_{[a,b]}) + CI(\psi_n) \leq \varepsilon [F(b) - F(a)] + C\varepsilon.
\]

Thus, taking limit as \( n \to \infty \) and then sending \( \varepsilon \to 0 \), we obtain \( I(\varphi_n) \to 0 \), which proves continuity condition (1.2-c).

Remark that the argument is shorter in the case of the Lebesgue measure where \( F(x) = x \). Indeed, the sequences \( \{ p_k \} \) and \( \{ q_k \} \) can be chosen \( p_k = r_k - \varepsilon 2^{-n-1} \) and \( q_k = r_k + \varepsilon 2^{-n-1} \) to ensure that \( \sum_k |q_k - p_k| \leq \varepsilon \). Therefore, by using Dini’s Theorem on the compact set \( K = [a,b] \setminus \bigcup_k (p_k,q_k) \), we deduce the estimate (A.1) with \( \psi_n(x) = 1 \) only if \( x \) belongs to some \( (p_k,q_k) \) with \( r_k \) being a point of discontinuity for \( \varphi_n \). Again, this implies \( I(\varphi_n) \leq \varepsilon (b-a) + C\varepsilon \) and the conclusion follows.

There is not too much changes for the Lebesgue measure in \( \mathbb{R}^d \), but the notation is more delicate.
Exercise 1.2. Let $E_i$ be a vector lattice of functions on a (Hausdorff) space $X_i$, for $i = 1, 2$. Denote by $E_1 \otimes E_2$ the vector lattice generated by functions of the form $\varphi_1(x_1)\varphi_2(x_2)$ with $\varphi_i$ in $E_i$, i.e., the smallest vector lattice containing the above class of functions. Verify that any element $\varphi(x_1, x_2)$ in $E_1 \otimes E_2$ satisfies: for every fixed $x_1$, the function $\varphi_{x_1} : x_2 \mapsto \varphi(x_1, x_2)$ belongs to $E_2$, and for every fixed $x_2$, the function $\varphi_{x_2} : x_1 \mapsto \varphi(x_1, x_2)$ belongs to $E_1$.

Proof. Indeed, let $F$ be the family of real-valued functions $\varphi$ from $X_1 \times X_2$ satisfying (a) for every fixed $x_1$, the function $\varphi_{x_1} : x_2 \mapsto \varphi(x_1, x_2)$ belongs to $E_2$, and (b) for every fixed $x_2$, the function $\varphi_{x_2} : x_1 \mapsto \varphi(x_1, x_2)$ belongs to $E_1$.

If $\varphi(x_1, x_2)$ has the form $\varphi_1(x_1)\varphi_2(x_2)$ with $\varphi_i$ in $E_i$, then the properties (a) and (b) hold true because $E_i$ is stable under the multiplication by constants. This proves that $F$ contains all product form functions.

Now, if $\varphi$ and $\psi$ belongs to $F$ and $a$, $b$ are constant, then $(a\varphi + b\psi)_{x_i} = (a\varphi_{x_i} + b\psi_{x_i})$ and $(\varphi \lor \psi)_{x_i} = (\varphi_{x_i}) \lor (\psi_{x_i})$ show that $a\varphi + b\psi$ and $\varphi \lor \psi$ are in $F$, because $E_i$ is a vector lattice, for $i = 1, 2$. Hence, $F$ is a vector lattice. This means that $F$ is equal to $E_1 \otimes E_2$.

In our definition of lattice (1.1) the Stone’s condition is included, but it is immediately to verify that $(1 \land \varphi)_{x_i} = 1 \land (\varphi)_{x_i}$, which means that if $E_i$ satisfies (1.1-c) for $i = 1, 2$ then $E_1 \otimes E_2$ has the same property. \qed

(1.1.1) Null or Negligible Sets

Exercise 1.3. Based on the technique of the previous Proposition 1.5, prove that if $N$ is a null set and $\psi$ and $\varphi$ are two functions in $E$ such that $\psi(x) = \varphi(x)$, for any $x$ in $X \setminus N$, then $I(\psi) = I(\varphi)$.

Proof. First, choose a increasing (nonnegative) sequence $\{\varphi_k\} \subset E$ such that $\varphi_k(x) \uparrow +\infty$, for every $x \in N$, and $I(\varphi_k) \leq 1$, for any $k \geq 1$.

Now, if $\psi = \varphi$ outside of $N$ and $\varepsilon > 0$ then the sequence $\{\phi_n = (\varphi - \psi - \varepsilon \varphi_n)^+\}$ satisfies $\varphi(x) - \psi(x) \leq \phi_n(x) + \varepsilon \varphi_n(x)$ and $\phi_n(x) \downarrow 0$, for every $x$. Hence $I(\phi_n) \downarrow 0$ and $I(\varphi) - I(\psi) \leq \varepsilon$, which implies that $I(\varphi) \leq I(\psi)$. Reversing the role of $\varphi$ and $\psi$, this proves that $I(\varphi) = I(\psi)$ as desired. Certainly, the above argument was partially used in Proposition 1.6. \qed

(1.1.2) Integrable Functions

Exercise 1.4. Prove that the following properties hold for the upper and the lower integrals: (a) $\bar{I}(f + g) \leq \bar{I}(f) + \bar{I}(g)$; (b) $\bar{I}(cf) = c\bar{I}(f)$, for any constant $c \geq 0$; (c) if $f \leq g$ then $\bar{I}(f) \leq \bar{I}(g)$ and $\bar{I}(g) \leq \bar{I}(f)$; (d) $\bar{I}(f) \leq \bar{I}(f)$ for any $f$, and $\bar{I}(g) = \bar{I}(g) = I(g)$ for $g$ in $\bar{E}$. Moreover, if $\{f_k\}$ is a sequence of nonnegative functions and $f = \sum_k f_k$ then $\bar{I}(f) = \sum_k \bar{I}(f_k)$.

Proof. Clearly, properties (a), (b), (c) and the first part of (d) are inherited from analogous properties of the infimum and supremum.

To check that $\bar{I}(g) = \bar{I}(g) = I(g)$ for $g$ in $\bar{E}$, just note that of any $g$ in $\bar{E}$, the infimum and supremum are really a minimum and a maximum, and both coincide.
Finally, if \{f_k\} is a sequence of nonnegative extended real-valued functions and \( f = \sum_k f_k \) and \( \sum_k I(f_k) < \infty \) then for any \( k \) and any \( \varepsilon > 0 \) there exists a sequence \{g_k\} \subset \bar{E} such that \( f_k \leq g_k \) and \( I(g_k) \leq I(f_k) + \varepsilon 2^{-k} \). Hence, \( g_k \geq 0, g = \sum_k g_k \) belongs to \( \bar{E} \) and \( I(g) = \sum_k I(g_k) \), in view of property (d) in Proposition 1.9). Thus

\[
I(f) \leq I(g) \leq \varepsilon + \sum_k I(f_k),
\]

and as \( \varepsilon \to 0 \) the desired inequality follows. Note that the case when \( \sum_k I(f_k) = \infty \) is certainly satisfied.

Exercise 1.5. With the previous notation, for any nonnegative simple function \( f = \sum_{i=1}^n a_i 1_{A_i} \geq 0 \) with \{\( A_i \)\} a finite sequence of disjoint subsets of \( X \), note that the expression

\[
I_*(f) = \sup \{ I(\varphi) : \varphi \leq f, \varphi \in \mathbb{K}^+ \},
\]

(A.2)
takes values in \([0, +\infty]\), and verify that (1) \( I_* \) is monotone, i.e., if \( 0 \leq f \leq g \) then \( I_*(f) \leq I_*(g) \), (2) super-additive, i.e., if \( f, g \geq 0 \) and \( f \wedge g = 0 \) then \( I_*(f + g) \geq I_*(f) + I_*(g) \), and (3) homogeneous, i.e., if \( c \geq 0 \) constant then \( I_*(c f) = c I_*(f) \), and that (4) \( I_*(\varphi) = I(\varphi) \), for every \( \varphi \) in \( \mathbb{K} \). Moreover, prove that the \( \mathbb{K} \)-tightness condition may be called \( \mathbb{K}^+ \)-tightness condition when written as (5) if \( \varphi \) and \( \psi \) are in \( \mathbb{K}^+ \) then \( I(\varphi) = I(\varphi \wedge \psi) + I_*(\varphi - \varphi \wedge \psi) \). Furthermore, show that (6) if \( f = \sum_{i=1}^n a_i 1_{A_i} \geq 0 \) with \{\( A_i \)\} a finite sequence disjoint measurable subsets of \( X \) (where measurability of a set \( A \) means that \( I(1_K) \leq I_*(1_K \cap A) + I_*(1_K \setminus A) \) for every \( K \) in \( \mathcal{K} \)) then \( I_*(f) = \sum_{i=1}^n a_i I_*(1_{A_i}) \). Finally, deduce that the unique linear extension of \( I \) could be given as the mapping \( \varphi \mapsto I_*(\varphi^+) - I_*(\varphi^-) \), with \( I_* \) given by (A.2).

Proof. (1)–(3) The monotonicity follows from the fact that if \( 0 \leq f \leq g \) then potentially more functions in \( \mathbb{K}^+ \) could satisfy the constraint \( \varphi \leq g \) than \( \varphi \leq f \), so that when taking the supremum the inequality \( I_*(f) \leq I_*(g) \) follows.

For the super-additivity, remark that if \( \varphi \leq f, \psi \leq g \) and \( f \wedge g = 0 \) with \( \varphi \) and \( \psi \) in \( \mathbb{K}^+ \) then \( \varphi \wedge \psi = 0 \), which implies that \( \varphi + \psi \) belongs to \( \mathbb{K}^+ \) and \( \varphi + \psi \leq f + g \). Thus \( I(\varphi) + I(\psi) \leq I_*(f + g) \), which yields the desired property. Also, it is clear that \( I_* \) inherits the homogeneity from \( I \).

(4) This follows from the monotonicity assumption on the initial definition of \( I \), i.e., if \( \varphi \leq f \) then \( I(\varphi) \leq I(f) \), provided that \( \varphi \) and \( f \) belong to \( \mathbb{K}^+ \).

(5) Because \( \mathbb{K}^+ \) is a lattice, only functions \( \varphi \geq \psi \) need consideration, i.e., we need to show that if \( \varphi \) and \( \psi \) are in \( \mathbb{K} \) satisfying \( \varphi \geq \psi \) then \( I(\varphi) = I(\psi) + I_*(\varphi - \psi) \). To this purpose, if \( \varphi \) and \( \psi \) belongs to \( \mathbb{K} \) then they assume only a finite number of values on sets belonging to the semi-lattice \( \mathcal{K} \), i.e., \( \varphi = \sum_{i=1}^n a_i 1_{A_i} \) and \( \psi = \sum_{j=1}^m b_i 1_{B_i} \), with \{\( A_i \)\} and \{\( B_i \)\} two finite sequences of disjoint sets in \( \mathcal{K} \) and \( a_1 < \cdots < a_n, b_1 < \cdots < b_m \). The condition \( \varphi \geq \psi \) yields \( A = \bigcup_{i=1}^n A_i \supset \bigcup_{j=1}^m B_j = B \) and if \( A_i \cap B_j \neq \emptyset \) then \( a_i \geq b_j \), and all these imply
that $\sum_i a_i \mathbb{1}_{A_i \cap B} \geq \psi$ and $\sum_i a_i \mathbb{1}_{A_i \setminus B} \geq \varphi - \psi$. Now, if $I$ is $\mathcal{K}$-tight then $I(\mathbb{1}_{A_i}) = I(\mathbb{1}_{A_i \cap B}) + I_*(\mathbb{1}_{A_i \setminus B})$, which yields

$$I(\varphi) = \sum_i a_i I(\mathbb{1}_{A_i}) = \sum_i [a_i I_*(\mathbb{1}_{A_i \cap B}) + a_i I_*(\mathbb{1}_{A_i \setminus B})] = \geq \sum_i [I_*(a_i \mathbb{1}_{A_i \cap B}) + I_*(a_i \mathbb{1}_{A_i \setminus B})] \geq I_*(\psi) + I_*(\varphi - \psi).$$

Because $I_*(\psi) = I(\psi)$, this shows that if $I$ satisfies the $\mathcal{K}$-tightness condition then the $\mathcal{K}^+$-tightness property is also satisfied. For the converse, note the following property: if $I_*$ is given by (A.2) then $I_*(\mathbb{1}) = \sup\{I(\mathbb{1}_K) : \mathbb{1}_K \leq \mathbb{1}_A, K \in \mathcal{K}\}$. Moreover, this fact was implicitly used in the previous inequality.

(6) One side of the inequality follows from the super-additivity, i.e., $I_*(f) \geq \sum_{i=1}^n a_i I_*(\mathbb{1}_{A_i})$. To show the converse inequality, the definition of supremum ensures that for every $r < I_*(f)$ there exists a function $\varphi \leq f$ in $\mathcal{K}^+$ such that $r < I(\varphi)$. Now, write $\varphi = \sum_{j=1}^m \alpha_j \mathbb{1}_{K_j}$ with $\{K_j\}$ a finite sequence of disjoint sets in $\mathcal{K}$ and use the facts that $I_*$ (as proved in Proposition 1.22) is linear on characteristic functions of measurable sets, and that $\{A_i\}$ is a finite sequence of disjoint measurable sets satisfying $\bigcup_{i=1}^n A_i \supset K_j$ to deduce the equality $I(\mathbb{1}_{K_j}) = \sum_{i=1}^n I_*(\mathbb{1}_{K_j \cap A_i})$. Hence, from this last equality, the expression of $\varphi$ and the inequality $\sum_{j=1}^m \alpha_j \mathbb{1}_{K_j \cap A_i} \leq a_i \mathbb{1}_{A_i}$, as well as the homogeneity, super-additivity and monotonicity of $I_*$, follows

$$I(\varphi) = \sum_{j=1}^m \alpha_j I(\mathbb{1}_{K_j}) = \sum_{j=1}^m \alpha_j \sum_{i=1}^n I_*(\mathbb{1}_{K_j \cap A_i}) \leq \sum_{i=1}^n I_*(\sum_{j=1}^m \alpha_j \mathbb{1}_{K_j \cap A_i}) \leq \sum_{i=1}^n a_i I_*(\mathbb{1}_{A_i})$$

Thus, $r \leq \sum_{i=1}^n a_i I_*(\mathbb{1}_{A_i})$, which yields $I_*(f) \leq \sum_{i=1}^n a_i I_*(\mathbb{1}_{A_i})$.

Actually, the relevant point in this context is that if $\varphi = \sum_{i=1}^n a_i \mathbb{1}_{E_i}$, with $\{E_i\}$ a finite sequence of disjoint sets belonging to the ring generated by the semi-lattice $\mathcal{K}$, then $I_*(\varphi) = \sum_{i=1}^n a_i I_*(\mathbb{1}_{E_i})$, i.e., the equality holds true for any function $\varphi \geq 0$ in the vector lattice $E$.

(1.1.3) Measurable Functions

Exercise 1.6. Let $E_i$ be a vector lattice of functions on a (Hausdorff) space $X_i$, for $i = 1, 2$, and set $X = X_1 \times X_2$ and $E = E_1 \otimes E_2$, see Exercise 1.2. Assume that a pre-integral $I_1$ is given on $E_i$, $i = 1, 2$, and such that for every $\varphi$ in $E$, the function $x_2 \mapsto I_1(\varphi(\cdot, x_2))$ belongs to $E_2$. Prove that the iterate expression $I(\varphi) = I_2(I_1(\varphi))$ defines a pre-integral on $E$. Based on results of this section, try to show that for any $I$-integrable function $f$ there exists a $I_2$-null set $N_2$ such that the function $x_1 \mapsto f(x_1, x_2)$ is $I_1$-integrable for every $x_2 \in X_2 \setminus N_2$, e.g., see Taylor [122, Section 7.2, pp. 329–334].

Proof. Perhaps we should check that the vector lattice $E$ satisfies Stone’s condition (1.1-c), see Exercise 1.2. Moreover, to show that $I$ is pre-integral, only part (c) of condition (1.2) needs some discussion. Indeed, if $\{\varphi_n\}$ is a decreasing sequence in $E$ pointwise convergent to zero, then Exercise 1.2 has established that the $x_2$-section functions $x_1 \mapsto \varphi_{n,x_2}(x_1) = \varphi_n(x_1,x_2)$ belong to $E_1$, and certainly, for each $x_2$, the sequence $\{\varphi_{n,x_2}\}$ is pointwise decreasing to zero. Because $I_1$ is a pre-integral then $I_1(\varphi_{n,x_2}) \downarrow 0$, for every $x_2$ in $X_2$. By assumption, the function $x_2 \mapsto I_1(\varphi_{n,x_2}) = I_1(\varphi_n(\cdot,x_2))$ belongs to $E_2$, and now, because $I_2$ is a pre-integral then $I(\varphi_n) = I_2(I_1(\varphi_n)) \downarrow 0$, proving that $I$ is indeed a pre-integral on $E$.

It is clear that $I(f) = I_2(I_1(f)) = I_1(I_2(f))$ for any function $f$ belonging to the vector space spanned by the product functions $\varphi_1(x_1)\varphi_2(x_2)$ with $\varphi_i$ in $E_i$, but not obviously true for any $f$ in the vector lattice $E$. This is certainly holds when each $E_i$ is a vector lattice of simple functions (since any vector subspace is also vector lattice). Any way, this is not part of our discussion.

The point is to check the form of $I$-null sets, with respect to $I_1$-null and $I_2$-null sets. If $f$ is an $I$-integrable function then, in view of Remark 1.14 and Definitions 1.10 and 1.19, there exits a sequence $\{\varphi_n\}$ in $E$ satisfying $\sum_n I(|\varphi_n|) < \infty$, and $f(x) = \sum_n \varphi_n(x)$ whenever $\sum_n |\varphi_n(x)| < \infty$, i.e., $N = \{x \in X : \sum_n |\varphi_n(x)| = \infty\}$ is a $I$-null set and $f$ agree with the pointwise limit of the series outside of $N$, and $I(f) = \sum_n I(\varphi_n)$.

Since $I(\varphi_n) = I_2(I(\varphi_n))$, for each $\varphi_n(x), x = (x_1,x_2)$, in $E$, we deduce that the series $I_1(\varphi_{x_2}) + \sum_n I_1(\varphi_{n,x_2})$ converges to $I(f_{x_2})$, for every $x_2$ outside of a $I_2$-null set, i.e., the function $x_1 \mapsto f(x_1,x_2)$ is $I_1$-integrable for every $x_2$ outside of a $I_2$-null set.

Exercise 1.7. Let $S$ be a (Stone) vector lattice, see (1.1), of bounded (real-valued) functions defined on $X$. First (1) show that if $f, g, h \geq 0$ and $f + g \geq h$ with $f, g, h$ in $S$ then we can write $h = h_1 + h_2$ with $h_i$ in $S$, $0 \leq h_1 \leq f$ and $0 \leq h_2 \leq g$. Next, let $I$ be a linear functional on $S$ such that there exists a constant $C$ satisfying $|I(f)| \leq C\|f\|$ for every function $f$ in $S$, where $\|f\| = \sup\{|f(x)| : x \in X \}$. Define

$$I^+(f) = \sup_{0 \leq h \leq f} \{I(h)\}, \quad \text{and} \quad I^-(f) = \inf_{0 \leq h \leq f} \{I(h)\},$$

for every $f \geq 0$ in $S$, and later $I^\pm(f) = I^\pm(f^+) - I^\pm(f^-)$. Prove (2) that $I^+$ and $I^-$ are two linear (nonnegative) functionals such that $I = I^+ - I^-$. Moreover, (3) if $I$ is a signed pre-integral (i.e., besides being linear it has the monotone convergence property $I(f_n) \to 0$ whenever $f_n \downarrow 0$ pointwise decreasing to 0) then so are $I^+$ and $I^-$.

Proof. This has been adapted from Bogachev [19, Section 7.8, pp. 99–107].

(1) If $f, g, h \geq 0$ and $f + g \geq h$ with $f, g, h$ in $S$ then define $h_1 = f \wedge h$ and $h_2 = h - h_1$ to have $h_i$ in $S$, $0 \leq h_1 \leq f$ and $h_2 \geq 0$. To check that $h_2 \leq g$, pick an $x$ to see that either (a) $h_1(x) = h(x)$ and then $h_2(x) = 0$ or (b) $h_1(x) = f(x)$ and then $h_2(x) = h(x) - h_1(x) \leq h(x) - f(x) \leq g(x)$, i.e., in both cases, $h_2(x) \leq g(x)$. 

(2) Since $|I(h)| \leq C\|h\|$ implies $I^+(f) \leq C\|f\|$, the functionals $I^+$ and $I^-$ take values in $[0, +\infty)$. It is clear that if $c$ is a nonnegative constant and $f \geq 0$ a function in $S$ then $I^+(cf) = cI^+(f)$. Moreover, in view of (1),

$$I^+(f + g) = \sup\{I(h) : 0 \leq h \leq f + g, \ h \in S\} = \sup\{I(h_1) + I(h_2) : 0 \leq h_1 \leq f, \ 0 \leq h_2 \leq g, \ h_i \in S\} = I^+(f) + I^+(g),$$

which prove that $I^+$ is linear (and analogously for $I^-$) as defined for nonnegative functions. To check this property for any function on $S$, note that if $f = f_1 - f_2$ with $f_i \geq 0$ then $f^+ + f^- = f_1 + f^-$ yields $I^\pm(f^+) + I^\pm(f^-) = I^\pm(f_1) + I^\pm(f^-)$, i.e., $I^\pm(f) = I^\pm(f_1) - I^\pm(f^-)$. The homogeneity $I^\pm(cf) = cI^\pm(f)$, for any constant $c$ and function $f$ in $S$, also follows immediately. Hence, the linearity on the whole vector lattice $S$ is established.

By definition the functional $I^+$ is nonnegative (or positive) and $I^+(f) \geq I(f)$ for any $f \geq 0$, which implies that the functional $I^+ - I$ is also nonnegative. Moreover, for any $f \geq 0$,

$$I^+(f) - I(f) = \sup_{0 \leq h \leq f} \{I(h) - I(f)\} = -\inf_{0 \leq h \leq f} \{I(f - h)\} = I^-(f),$$

i.e., $I = I^+ - I^-$. Moreover, if $|I| = I^+ + I^-$ then $|I|(f) = \sup_{0 \leq h \leq f} \{|I(h)|\}$ for every $f \geq 0$ in $S$, and $I(f) \leq (I^+(1) + I^-(1))\|f\|$, for every $f$ in $S$.

(3) If $I$ is a pre-integral on $S$ then let us check that so is $I^\pm$. Indeed, if $\{f_n\}$ is a decreasing sequence in $S$ which is pointwise converging to 0, and $\varepsilon > 0$ then there exits another sequence $\{\varphi_n\}$ in $S$ such that $I^\pm(f_n) - 2^{-n}\varepsilon < I^\pm(\varphi_n)$. Next, define by induction $g_1 = \varphi_1$ and $g_n = \min\{g_{n-1}, \varphi_n\}$ for $n \geq 2$, to see that $I^\pm(f_n) \leq I(g_n) + \varepsilon \sum_{k=1}^{n} 2^{-k}$. Indeed, the equality $g_n + \max\{g_{n-1}, \varphi_n\} = g_{n-1} + \varphi_n$ yields

$$I(g_n) + I(\max\{g_{n-1}, \varphi_n\}) \geq I(g_{n-1}) + I^\pm(f_n) - 2^{-n}\varepsilon.$$

On the other hand, the inequalities $g_{n-1} \leq \varphi_{n-1} \leq f_{n-1}$ and $\varphi_n \leq f_n \leq f_{n-1}$, and the induction assumption for $n - 1$ imply

$$I(\max\{g_{n-1}, \varphi_n\}) \leq I^\pm(f_{n-1}) \leq I(g_{n-1}) + \varepsilon \sum_{k=1}^{n-1} 2^{-k}.$$

Combining both inequalities, the desired inequality

$$I(g_{n-1}) + I^\pm(f_n) - 2^{-n}\varepsilon \leq I(g_n) + I(g_{n-1}) + \varepsilon \sum_{k=1}^{n-1} 2^{-k}$$

follows. This means that $g_n = \min_{k \leq n}\{\varphi_k\} \leq f_n$ satisfies $I^\pm(f_n) \leq I((g_n) + \varepsilon$. Since $I$ is a pre-integral, we have $I(g_n) \downarrow 0$, which implies that $I^\pm(f_n) \downarrow 0$, i.e., $I^\pm$ is also a pre-integral.
(1.2) Uniform Integrability

(1.2.1) Main Properties

(1.2.2) Mean Convergence

(1.2.3) Convergence in Norm

Exercise 1.8. Consider the Lebesgue measure on the interval $(0, \infty)$ and define the functions $f_i = (1/i) 1_{(i, 2i)}$ and $g_i = 2i 1_{(2^{-i-1}, 2^{-i})}$ for $i \geq 1$. Prove that (a) the sequence $\{f_i : i \geq 1\}$ is uniformly integrable of any order $p > 1$, but not of order $0 < p \leq 1$. On the contrary, show that (b) the sequence $\{g_i : i \geq 1\}$ is uniformly integrable of any order $0 < p < 1$, but the sequence is not equi-integrable of any order $p \geq 1$.

Proof. Since $0 \leq f_i(x) \leq f(x)$ with $f(x) = \min\{1, 2/x\}$ and $f$ belongs to $L^p([1, \infty[)$ for every $p > 1$, the sequence $\{f_i : i \geq 1\}$ is uniformly integrable of any order $p > 1$. Clearly, for $0 < p \leq 1$, the difficulty is the tightness condition. If $A$ is a subset of $(1, \infty)$ with finite Lebesgue measure then the Lebesgue dominate convergence implies

$$\lim_{i} \int_A |f_i(x)|^p dx = 0, \quad \forall p > 0,$$

and because

$$\int_{(0, \infty)} |f_i(x)|^p dx = i^{1-p} + 1^{1-p} > 0, \quad \forall i \geq 1, \ 0 < p \leq 1,$$

we deduce that for every $\varepsilon > 0$ and any set $A \subset (1, \infty)$ with finite Lebesgue measure there exists an index $i \geq 1$ such that

$$\int_{A^c} |f_i(x)|^p dx > \varepsilon,$$

i.e., the sequence $\{f_i : i \geq 1\}$ is neither uniformly integrable nor equi-integrable of order $0 < p \leq 1$. Note that $f_i(x) \to 0$ as $i \to \infty$ for every $x$ in $(0, \infty)$.

If $0 < p < 1$ then

$$\int_{(0, \infty)} |g_i|^p dx = 2^i(p-1) \to 0 \quad \text{as} \quad i \to \infty,$$

which show that the sequence $\{g_i : i \geq 1\}$ is uniformly integrable of any order $0 < p < 1$. However, if $p \geq 1$ then open interval $I_i = (2^{-i-1}, 2^{-i})$ satisfies

$$\int_{I_i} |f_i(x)|^p dx = 2^i(p-1) \geq 1 \quad \forall i \geq 1$$

but the Lebesgue measure of $I_i$ vanish as $i \to \infty$, which proves that the sequence $\{g_i : i \geq 1\}$ is not equi-integrable integrable of order $p \geq 1$. Note that $g_i(x) = 0$ for every $x \geq 1$ and $g_i(x) \to 0$ as $i \to \infty$ for every $x$ in $(0, \infty)$. \qed
(1.3) Vector-valued Integrals

(1.3.1) Metric Space of Measurable Functions

(1.3.2) With Values in a Banach Space
(2.1) Previous Background

(2.1.1) Simple Spectral Analysis

(2.1.2) Three Basic Results

(2.1.3) Introduction with Examples

**Exercise 2.1.** Given a domain $E$ in the Euclidean space $\mathbb{R}^d$ and $0 < \alpha < 1$ we say that a function $f : E \rightarrow \mathbb{R}$ is Hölder continuous in $E$ with exponent $\alpha$ if there exists a constant $C$ such that $|f(x) - f(y)| \leq C|x - y|^\alpha$, for every $x, y$ in $E$ (and the limiting case $\alpha = 1$ is called Lipschitz continuous), and the smallest of all those constant $C$ is denoted by $[f]_{\alpha,E}$, i.e.,

$$[f]_{\alpha,E} = \sup_{x,y \in E, x \neq y} \{|f(x) - f(y)| |x - y|^{-\alpha}\}.$$  

For the limiting case $\alpha = 0$, we use $C^0(E) = C(E)$. Now, denote by $C^{0,\alpha}(E)$ the space of all Hölder (Lipschitz) continuous functions on $E$. Sometime, the notation $C^{0,\alpha}(E) = C^{\alpha}(E)$, with $0 < \alpha < 1$, could be used. Assume $E$ a bounded set and prove that $C^{0,\alpha}(E)$ are Banach spaces with the norm

$$\|f\|_{\alpha,E} = [f]_{\alpha,E} + \sup_{x \in E} |f(x)|, \quad 0 < \alpha \leq 1$$

Consider also the case when $E$ is unbounded and discuss the spaces $C^{m,\alpha}_b(E)$ defined as a combination of $C^m_b(E)$ and $C^{0,\alpha}(E)$.

**Proof.** The only point to discuss is the completeness of the space. For this purpose, let $\{u_n\}$ be a Cauchy sequence in $C^{0,\alpha}(E)$, i.e., $\|u_n - u_m\|_{\alpha,E} \rightarrow 0$ as $n, m \rightarrow \infty$. Because this sequence is also a Cauchy sequence in $C^0(E)$, there exists a function $u$ such that $u_n(x) \rightarrow u(x)$ in the uniform norm. Thus, given $\varepsilon > 0$ find $N = N(\varepsilon)$ such that $\|u_n - u_m\|_{\alpha,E} < \varepsilon$, for every $n, m \geq N$, and write

$$|[u_n(x) - u_m(x)] - [u_n(y) - u_m(y)]| \leq [u_n - u_m]_{\alpha,E}|x - y|^\alpha \leq \varepsilon|x - y|^\alpha,$$
send \( m \to \infty \) to deduce \(|u_n(x) - u(x)| - |u_n(y) - u(y)|| \leq \varepsilon |x - y|^{\alpha} \), which implies that the convergence is also in the \( C^{0,\alpha}(E) \) norm.

The interested reader may remark that (e.g., see Kufner et al. [76]) the space \( C^{0,\alpha}(E) \) is not separable, even when \( E \) is compact. Also note that the interpolation inequality

\[
[f]_{\alpha', E} = \sup_{x,y \in E} \{ |x - y|^{-\alpha'} |f(x) - f(y)| \} \leq \left( \sup_{x,y \in E} \{ |x - y|^{-\alpha} |f(x) - f(y)| \} \right)^{\alpha'} \left( \sup_{x,y \in E} |f(x) - f(y)| \right)^{1 - \frac{\alpha'}{\alpha}} \leq 2 \left( [f]_{\alpha, E} \right)^{\frac{\alpha'}{\alpha}} \left( \sup_{x \in E} |f(x)| \right)^{1 - \frac{\alpha'}{\alpha}}, \quad \forall f \in C^{0,\alpha}(E),
\]

shows that if \( 0 \leq \alpha' < \alpha \leq 1 \) then \( C^{0,\alpha}(E) \subset C^{0,\alpha'}(E) \subset C^{0}(E) \). Usually, the \( E \) is a compact domain, so that any continuously differentiable functions belongs to \( C^{0,\alpha}(E) \), any \( 0 < \alpha \leq 1 \), For instance, (1) an absolutely continuous function \( f \) on a bounded interval \( I \subset \mathbb{R} \) with a derivative almost everywhere equal to an element \( f' \) in \( L^p(I) \), with \( p > 1 \), belongs to \( C^{0,\alpha}(I) \) for any \( 0 < \alpha \leq 1 - 1/p \); (2) the inequality \( (\sqrt{x} - \sqrt{y})^2 \leq x + y - 2 \min\{x,y\} = |x - y| \) shows that function \( f(x) = x^\frac{1}{p} \) belongs to \( C^{0,\alpha}(I) \), for any \( 0 < \alpha \leq 1/2 \) with 0 in \( I \), but it does not belongs to \( C^{0,\beta}(I) \) for any \( 1/2 < \beta \leq 1 \); (3) the function \( f(x) = 1/\ln|x| \) and \( f(x) = 0 \), for \( x \) within the interval \( \alpha = [-1/2, 1/2] \) is a continuous function that does not belong to any \( C^{0,\alpha}(I), 0 < \alpha \leq 1 \).

If the domain \( E \) is unbounded then a continuous function on the closure \( \bar{E} \) is not necessarily bounded, so that the notation \( C_b^{0,\alpha}(E) \) is necessary, the above argument shows that \( C_b^{0,\alpha}(E) \) is a Banach space too. Moreover, instead of only bounded, we may impose other conditions, e.g., “vanishing at infinite”, i.e., functions \( f \) such that for every \( \varepsilon > 0 \) there exists a compact set \( K = K_{\varepsilon} \subset E \) such that \( \|f\|_{\alpha, E-K} \leq \varepsilon \).

The space \( C_b^{n,\alpha}(E) \) are defined in the same way, the space of all function \( f \) defined on the domain \( E \) with real (or complex) values which are continuously differentiable and bounded on \( E \) up to the order \( n \), and all derivative of order \( n \) belongs to \( C_b^{0,\alpha}(E) \). As expected, this space is a complete endowed with the norm

\[
\|f\|_{C_b^{n,\alpha}(E)} = \sum_{k=1}^{n} \|\partial^k f\|_{C_b^0(E)} + \|\partial^n f\|_{C_b^{0,\alpha}(E)},
\]

where \( \partial^k \) means all derivatives of order \( k \). Sometimes, for \( 0 < \alpha < 1 \) the notation \( C_b^{n,1+\alpha}(E) \) is used, but we remark the differences between the spaces \( C_b^{n,1}(E) \) and \( C_b^{n+1}(E) \).
(2.2) Compactness and Separability

(2.2.1) Linear Functionals

(2.2.2) Nonlinear Functional

Exercise 2.2. With the notation of Exercise 2.1, let \( \{f_n\} \) be a bounded sequence in the Hölder space \( C^{0,\alpha}(K) \) with \( K \subset \mathbb{R}^d \) and \( 0 < \alpha \leq 1 \). Prove that if \( 0 < \alpha' < \alpha \) and \( K \) is compact then there exists a subsequence \( \{f_{n_k}\} \) and a function \( f \) in \( C^{0,\alpha}(K) \) such that \( f_{n_k} \to f \) in \( C^{0,\alpha'}(K) \).

Proof. Because the topology in \( C^{0,\alpha}(K) \) is sequential, this statement is equivalent to the following: any bounded set in \( C^{0,\alpha}(K) \) is relative compact in \( C^{0,\alpha'}(E) \).

From Arzela-Ascoli Theorem 2.9 follows that there exists a a subsequence \( \{f_{n_k}\} \) and a continuous function \( f \) such that \( f_{n_k} \to f \) uniformly on the compact set \( K \subset \mathbb{R}^d \). Since \( \|f_n\|_{\alpha,K} \leq C \) for some constant, \( |f_{n_k}(x) - f_{n_k}(y)| \leq C|x - y|^{\alpha} \), \( \forall x, y \in K, \forall k \)

which shows that \( f \) belongs to \( C^{0,\alpha}(K) \). Moreover, the interpolation estimate

\[ [g]_{\alpha',K} \leq ([g]_{\alpha,K})^{{\alpha'}/\alpha} \left( \sup_{x,y \in K} |g(x) - g(y)| \right)^{1-\alpha'/\alpha}, \forall g \in C^{0,\alpha}(K), \]

applied to the function \( g = f_{n_k} - f \) yields \( \|f_{n_k} - f\|_{\alpha',K} \to 0 \) as desired.

(2.2.3) Baire Category Arguments

(2.3) Three Essential Principles

(2.3.1) Uniformly Boundedness Principle

(2.3.2) Open Mappings Theorem

(2.3.3) Closed Graph Theorem

Exercise 2.3. Let \( V \) be a finite-dimensional linear subspace of a topological vector space \( X \) and \( p \) be a continuous seminorm on \( X \) such that \( p(v) = 0 \) and \( v \) in \( V \) imply \( v = 0 \). Take a basis \( \{v_1, \ldots, v_n\} \) in \( V \) and consider the continuous linear mapping \( c = (c_1, \ldots, c_n) \) from \( \mathbb{R}^n \) into \( X \) defined by \( T c = c_1 v_1 + \cdots + c_n v_n \). First (1) minimize the real-valued function \( c \mapsto p(Tc) \) over the region \( \{c : |c_1| + \cdots + |c_n| = 1\} \), and then (2) prove the estimate

\[ |c_1| + \cdots + |c_n| \leq K p(Tc), \forall c \in \mathbb{R}^n \]

for some constant \( K > 0 \). Finally, (3) deduce that \( T^{-1} : V \to \mathbb{R}^n \) is also continuous and therefore \( V \) is closed in \( X \).
Proof. (1) Note that the region \( R = \{ c \in \mathbb{R}^d : |c_1| + \cdots + |c_n| = 1 \} \) is compact and the function \( f : c \mapsto p(Tc) \) is continuous and strictly positive, so the minimum value \( 1/K \) is positive, i.e., \( f(c) \geq 1/K \) for every \( c \) in \( R \).

(2) Since \( f \) is positive homogeneous, \( f(rc) = |r|f(c) \) for every scalar \( r \). Given any \( c \neq 0 \) in \( \mathbb{R}^n \) define \( 1/r = |c_1| + \cdots + |c_n| \) to see that \( rc \) belongs to \( R \), which yields

\[
|c_1| + \cdots + |c_n| = \frac{1}{r} \leq Kf(c) = Kp(Tc), \quad \forall c \in \mathbb{R}^n,
\]
as desired.

(3) In view of the above estimate, the inverse \( T^{-1}(v) = c \) if \( v = c_1v_1 + \cdots + c_nv_n \) is a continuous linear function from \( V \) into \( \mathbb{R} \) with the norm \( c \mapsto |c_1| + \cdots + |c_n| \), and the argument is complete, see also Remark 2.15. \( \square \)

Exercise 2.4. On a given topological vector space \( X \), (1) recall the definition of sequentially compact and bounded sets, and (2) prove that any sequentially compact set \( A \subset X \) is also a bounded set. Next, (3) show that every topological vector space \( X \) having a compact neighborhood of zero is finite dimensional.

Proof. (1) Recall that in a topological vector space \( X \), a set \( K \) is called sequentially compact if every sequence in \( K \) has a convergence subsequence, and a set \( B \) is called bounded if for every neighborhood \( U \) of zero there exists a scalar \( s > 0 \) such that \( B \subset sU \), for every \( t > s \). It is clear that if \( X \) is a lctvs then the definition of bounded set becomes: for every neighborhood \( U \) of zero there exists a scalar \( t > 0 \) such that \( B \subset tU \), and if the space \( X \) is a normed space then this is equivalent to \( \sup_{x \in B} \|x\| < \infty \). Moreover, in a topological space, a set \( F \) having the property

\[
\forall \{x_k \} \subset F, \quad x_k \to x \quad \text{implies} \quad x \in F
\]
is called sequentially closed. It is clear that any closed set \( F \) is sequentially closed, and if the converse holds true then the topology is called a sequential topology. Certainly, any compact set is sequentially compact, and the converse holds true in any sequential topology. Similarly, a set is called (sequentially) relative compact (or pre-compact) if its (sequential) closure is compact. A typical example of a sequential topology is the one given by a metric.

Note that in a metric space \((X,d)\) a set \( B \) is called \( d \)-bounded if its diameter \( d(B) = \sup \{d(x,y) : x,y \in B \} \) is finite, and in general, this notion does not agree with the concept of a bounded set in a topological vector space (e.g., if \( d \) is an invariant metric on \( X \) then \( d'(x,y) = d(x,y)/(d(x,y) + 1) \) is another invariant metric yielding the same topology where the diameter of the whole space is finite). However, both definition agree on a normed space.

(2) Let \( K \) be a sequentially compact subset of a given topological vector space \( X \). To prove that \( K \) is bounded, choose a neighborhood \( U \) of zero and suppose that for every \( n > 0 \) there exists a point \( x_n \) in \( K \setminus t_nU \), with \( t_n > n \). Because \( K \) is sequentially compact, there is a subsequence \( x_{n_k} \to x \). However,
(1/t_n)x_n does not belongs to \( U \) and the continuity of the scalar multiplication implies that \((1/t_{n_k})x_{n_k} \to 0\), which is a contradiction.

(3) Take an open set \( O \) containing the origin with a compact closure \( \overline{O} \). The family \( \{x + \frac{1}{2}O : x \in X\} \) is an open cover of \( \overline{O} \), which must have a finite subcover, i.e., there exit \( x_1, \ldots, x_n \) in \( X \) such that \( \overline{O} \subset (x_1 + \frac{1}{2}O) \cup \cdots \cup (x_n + \frac{1}{2}O) \). If \( Y \) is the vector space spanned by \( \{x_1, \ldots, x_n\} \) then \( \overline{O} \subset Y + \frac{1}{2}O \), and because \( Y \) is a subspace, \( \frac{1}{2}O \subset \frac{1}{2}Y + \frac{1}{4}O = Y + \frac{1}{4}O \) and \( O \subset Y + \frac{1}{2}O = Y + Y + \frac{1}{4}O = Y + \frac{1}{2}O \), and, by induction, \( O \subset Y + 2^{-k}O \), for every \( k \geq 1 \). If \( y \) belongs to \( Y + 2^{-k}O \) for every \( k \geq 1 \) then \( y = y_k + 2^{-k}z_k \) with \( y_k \) in \( Y \) and \( z_k \) in \( O \). In view of (2), \( \overline{O} \) is bounded, i.e., for every open set \( V \) there exists \( s > 0 \) such that \( \overline{O} \subset tV \), for every \( t \geq s \), and so \( 2^{-k}z_k \) belongs to \( 2^{-k}tV = V \), if \( t = 2^k \), which means that \( 2^{-k}z_k \to 0 \). Hence, \( y \) belongs to the sequential closure of \( Y \), and so, \( y \) belongs to \( \overline{Y} = Y \), after invoking Remark 2.15. This shows that \( O \subset Y \). Now, for any \( x \) in \( X \), the continuity of the multiplication implies that the sequence \( \{x_k = (1/k)x, k \geq 1\} \) converges to zero, and so, \((1/k)x \) belongs to \( O \) for any \( k \) sufficiently large, i.e., \( x \) belongs to \( kO \subset kY = Y \), which proves that \( X = Y \). \( \square \)

(2.3.4) Hahn-Banach Theorem

(2.4) More on Lebesgue Spaces

(2.4.1) Weak Convergence

(2.4.2) Totally Bounded Sets

**Exercise 2.5.** If \( A \) is a totally bounded set of a normed space \( (X, \| \cdot \|) \) then prove that the convex hull (or convex envelope) \( \text{co}(A) \) of \( A \) (i.e., the smallest convex set containing \( A \)) is also totally bounded. In particular, the closed convex hull of a compact set of a Banach space is also compact. *Hint:* Use the following argument (1) if \( F \subset X \) is a finite set then the convex hull \( \text{co}(F) \) of \( F \) is a totally bounded set. Next, let \( A \) be a totally bounded subset of \( X \) and let \( B_1 \) be an open balls containing the origin. By using the previous result, (2) find a finite set \( F \) such that \( A \subset F + B_1 \) and deduce that \( \text{co}(A) \) lies inside \( K + B_1 \) for some totally bounded set \( K \). Now, take any two open balls \( B_1 \) and \( B \) containing the origin and satisfying \( B_1 + B_1 \subset B \). Finally, because \( K \) is totally bounded, (3) find another finite \( E \) such that \( \text{co}(A) \subset (E + B_1) + B_1 \subset E + B \), and deduce that \( \text{co}(A) \) is indeed totally bounded.

*Proof. (1)* First, note that \( x \) belongs to \( \text{co}(F) \) if and only if \( x \) is a convex combination of points in \( F \), i.e., if and only if there exist \( n \geq 2 \), \( a_i \) in \( [0,1] \), and points \( x_i \), for \( i = 1, \ldots, n \) such that \( \sum_{i=1}^{n} a_i = 1 \) and \( x = \sum_{i=1}^{n} a_i x_i \). Thus, if \( F \) is a finite set, say \( F = \{x_1, \ldots, x_n\} \), then consider the dyadic approximation in \( [0,1] \), i.e., \( D_k = \{j2^{-k} \in [0,1]: j = 0, 1, \ldots, 2^k\} \), \( k = 1, 2, \ldots \), with \( D_k(a) = \max\{d \in D_k: d \leq a\} \), and the finite set \( F_k = \{y = \sum_{i=1}^{n} d_i x_i : d_i \in D_k, 1 = \sum_{i=1}^{n} d_i \in \text{co}(F)\} \). Hence, for any point \( x \) in \( \text{co}(F) \) there exist \( a_i \) in \( [0,1] \) such that \( \sum_{i=1}^{n} a_i = 1 \) and \( x = \sum_{i=1}^{n} a_i x_i \). For each \( a_i \) define \( d_i = D_k(a_i) \) for

\[ i = 1, \ldots, n-1 \text{ and } d_n = 1 - d_1 - \cdots - d_{n-1} \] to deduce that \( y = \sum_{i=1}^{n} d_i x_i \) belongs to \( F_k \) and
\[
\|x - y\| \leq \sum_{i=1}^{n} |a_i - d_i| \|x_i\| \leq 2(n - 1)2^{-k} (\max_i \|x_i\|).
\]
Therefore, given any \( \varepsilon > 0 \) there exists \( k \) such that \( 2(n - 1)2^{-k} < \varepsilon \), which implies that any point in \( \text{co}(F) \) is within a distance less than \( \varepsilon \) from the finite set \( F_k \), i.e., \( \text{co}(F) \) is totally bounded.

(2) Since the open \( B_1 \) contains the origin, there exists \( \varepsilon > 0 \) such that \( \|x\| \leq \varepsilon \) implies \( x \) is in \( B_1 \), and because \( A \) is totally bounded there exits a finite set \( F \) such that every point in \( A \) lies within a distance less than \( \varepsilon \) from \( F \). This yields \( A \subset F + B_1 \). The ball \( B_1 \) is convex, and in view of (1), the convex hull \( K = \text{co}(F) \) is totally bounded, therefore \( \text{co}(A) \subset K + B_1 \).

(3) Because \( K \) is totally bounded and \( B_1 \) is a ball containing the origin, invoke the property (1) to find another finite set \( E \) such that \( K \subset E + B_1 \). Hence, the inclusion (2) implies \( \text{co}(A) \subset (E + B_1) + B_1 \subset E + B \). Since the ball \( B \) is also arbitrary, this shows that \( \text{co}(A) \) is totally bounded.

Finally, remark that in a Banach space (i.e., a complete normed space) a set is totally bounded if and only if it is pre-compact. Recall that closure and the interior of a convex set is convex, and that the convex hull of an open set is open. However, the convex hull of a closed set is not necessarily closed. In a finite-dimensional space, the convex hull of a compact set is compact.

Exercise 2.6. Banach-Saks Theorem states that if \( \{f_n\} \) is a weakly convergence sequence to \( f \) in \( L^p(\Omega, \mathcal{F}, \mu), 1 \leq p < \infty \) then there exists a subsequence \( \{f_{n_k}\} \) such that the arithmetic means \( g_k = (f_{n_1} + \cdots + f_{n_k})/k \) strongly converges to \( f \), i.e., \( \|g_k - f\|_p \to 0 \). Prove this result for a Hilbert space \( H \) with scalar product (\( \cdot, \cdot \)) and norm \( \| \cdot \| \), in particular for \( p = 2 \). Hint: First reduce the problem to the case where \( f = 0 \), and \( \|f_n\| \leq 1 \) for every \( n \geq 1 \). Next, construct a subsequence satisfying \( |(f_{n_i}, f_{n_{i+1}})| \leq 1/k \), for every \( i = 1, \ldots, k \), and deduce that \( \|g_k\|^2 \leq 3/k \). see Riesz and Nagy [107, Section 38, pp. 80–81].

Proof. First, if \( f_n \to f \) weakly then \( \|f_n\|_p \leq C \), for every \( n \), and then \( f_n - f \to 0 \) weakly. Hence, the sequence of functions \( f'_n = (f_n - f)/(2C) \) converges weakly to 0 and \( \|f'_n\|_p \to 1 \).

Now, let \( \{f_n\} \) be a sequence weakly convergence sequence to 0 in a Hilbert space \( H \) satisfying \( \|f_n\| \leq 1 \). Beginning with \( f_{n_1} = f_1 \), note that \( (f, f_n) \to 0 \), as \( n \to \infty \), for every \( f \) in \( H \), to choose \( f_{n_2} \) such that \( |(f_{n_1}, f_{n_2})| \leq 1 \), and next, by induction, to choose \( f_{n_k} \) such that \( |(f_{n_i}, f_{n_{i+1}})| \leq 1/k \), for every \( i = 1, \ldots, k \).

Define \( g_k = (f_{n_1} + \cdots + f_{n_k})/k \) to check that
\[
\|g_k\|^2 \leq \frac{2}{k^2} \sum_{1 \leq i < j}^{k} |(f_{n_i}, f_{n_j})| + \frac{1}{k^2} \sum_{i=1}^{k} |(f_{n_i}, f_{n_i})|,
\]
and because the first sum has \( k(k-1) \) terms, all bounded by \( 1/(k-1) \) and the second sum has \( k \) terms all bounded by 1, deduce that \( \|g_k\|^2 \leq 3/k \). This shows that the sequence \( g_k \) strongly converges to 0.
(2.5) Basic Interpolation Ideas

(2.5.1) Preliminary Interpolation

(2.5.2) Marcinkiewicz Interpolation Theorem

(2.5.3) Riesz-Thorin Interpolation Theorem

(2.5.4) Intermediate Spaces
Exercise 3.1. Use the argument in Exercise 2.5 to show that the closed convex hull of a totally bounded subset $A$ in a Fréchet space is a compact set.

**Proof.** Revise the arguments in Exercise 2.5 as follows. Let $A$ be a totally bounded subset $A$ of a Fréchet space $(X,d)$.

If $F$ is a finite set then its closed convex hull $(co(F)$ or the smallest closed convex hull containing $F$, is indeed, its convex hull $co(F)$. Indeed, if $F = \{x_1, \ldots, x_n\}$ and $\{z_k\}$ is a sequence in $co(F)$ converging to $z$ then $z_k = \sum_{i=1}^{n} a_i^k x_i$ with $a_i^k$ in $[0,1]$. Thus, there exists a convergent subsequence $\{a_i^k\}$, $a_i^k \to a_i$ as $j \to \infty$, which implies that $z = \sum_{i=1}^{n} a_i x_i$, i.e., the convex hull $co(F)$ is closed. Hence, $F$ is a finite set then its closed convex hull $co(F)$ is compact.

If $B_1$ is an open ball set and $A$ is totally bounded then there exists finite set $F \subset A$ such that $A \subset F + B_1/2$. Hence, if $z = \sum_{i=1}^{n} a_i z_i$ with $a_i$ in $[0,1]$ and $z_i$ in $A$ then there exists $x_i$ in $F$ such that $z_i - x_i$ belongs to $B_1/2$, which yields $co(A) \subset co(F) + B_1/2$. Because $co(F)$ is compact we have $co(F) + B_1/2 = co(F) + B_1/2$, i.e., $co(A) \subset co(F) + B_1$, and $co(F) = K$ is a compact convex set in the Fréchet space $X$.

Next, if $B_1$ and $B$ are two balls containing the origin and satisfying $B_1 + B_1 \subset B$ then repeating the argument with the totally bounded set $K$, there exists a finite set $E$ such $co(A) \subset (E + B_1) + B_1 \subset E + B$. This proves that $co(A)$ is totally bounded, and because it is also closed, the closed convex hull $co(A)$ it is compact.

The concept of totally bounded is initially defined for a metric space, i.e., a set $A$ is totally bounded if for every $\varepsilon > 0$ there exists a finite subset $F = F_\varepsilon$ of $A$ such that all points in $A$ are within a distance less than $\varepsilon$ from the finite set $F$. However, a locally convex topological vector space $X$ which is not a Fréchet space does not have a metric, a set $A$ is totally bounded if and only if for every
\[ \varepsilon > 0 \] and for any continuous seminorm \( p \) there exists a finite set \( F = F_{\varepsilon,p} \) such that for any point \( x \) in \( A \) there exist a point \( y \) in \( F \) satisfying \( p(x - y) < \varepsilon. \) The above argument extends to this case, i.e., the closed convex hull \( \overline{\mathcal{C}}(A) \) of a totally bounded subset \( A \) in a complete locally convex topological vector space \( X \) is a compact set.

As expected, in the case of an inductive topology \( X = \bigcup_k X_k \), a set \( A \) in \( X \) is totally bounded if and only if \( A \) is totally bounded in some \( X_k. \)

**Exercise 3.2.** Following Remark 2.16, let \( N \) be a closed (vector) subspace of a locally convex topological vector space \( X \) with a separating family of seminorms \( \{ p_i : i \in I \} \). The quotient space \( X/N \) is the space of cosets \( \bar{x} = x + N. \) Verify that \( X/N \) is a vector space and that \( \{ \bar{p}_i : i \in I \} \) with

\[
\bar{p}_i(\bar{x}) = \inf_{x \in \bar{x}} p_i(x), \quad \forall \bar{x} \in X/S
\]

is a separating family of seminorms for \( X/N \), i.e., \( X/N \) becomes a lctvs. Next show that if \( X \) is complete, metrizable or separable then so is \( X/N. \)

**Proof.** To show that \( X/N \) is a lctvs, the only question to discuss is the separating property of the family of seminorms \( \{ \bar{p}_i \}. \) To this purpose, first recall that a point \( x \) belongs to the closure of \( N \) (in this case to \( N \) because \( N \) is closed) if and only if every open set containing \( x \) intercept \( N \), i.e., \( N \cap \{ y : p_i(y-x) < \varepsilon \} \neq \emptyset \), for every \( i \) in \( I \) and every \( \varepsilon > 0. \) Now, we proceed by contradiction, if the family is not separating, i.e., there exists \( \bar{x} \neq 0 \) such that \( \bar{p}_i(\bar{x}) = 0 \) for any \( i \) in \( I \), then any \( x \) in \( \bar{x} \) does not belongs to \( N \) and for every \( \varepsilon > 0 \) there exists \( n \) in \( N \) such that \( p_i(x + n) < \varepsilon, \) which means that \( x \) belongs to the closure of \( N \), which is a contraction.

Since \( \bar{p}_i(\bar{x}) \leq p_i(x) \) for every \( x \) in \( \bar{x} \), it is clear that if a sequence \( \{ x_k \} \) is dense in \( X \) then the sequence \( \{ \bar{x}_k \} \) is dense in \( X/N \), i.e., if \( X \) is separable then so is \( X/N. \) Moreover, because the family of seminorms have the same cardinal, if \( X \) is metrizable then so is \( X/N. \)

To check the completeness, take a Cauchy sequence \( \{ \bar{x}_k \} \) in \( X/N \) and, by contradiction, suppose that there is not accumulation point, i.e., for every \( \bar{x}, \) there exists \( \varepsilon > 0, \) and \( i \) in \( I \) such that \( \bar{p}_i(\bar{x}_k - \bar{x}) > \varepsilon, \) for every \( k, \) i.e., \( p_i(x_k - x) > \varepsilon, \) for every \( x_k \) in \( \bar{x}_k \) and \( x \) in \( \bar{x}. \) Hence, choosing \( x_k \) and \( x \) so \( \bar{p}_i(\bar{x}_k - \bar{x}_\ell) + 1/(k+\ell) > p_i(x_k - x_\ell), \) we obtain a Cauchy sequence \( \{ x_k \} \) without accumulation point, which contract the fact that \( X \) is complete.

**Exercise 3.3.** Let \( A \) and \( B \) be two closed subsets of a topological vector space \( X. \) Give an example where \( A + B = \{ a + b : a \in A, b \in B \} \) is not necessarily closed. Next show (1) if \( A \) or \( B \) is (sequentially) compact then \( A + B \) is closed and (2) if \( A \) and \( B \) are independent closed vector subspaces, i.e., \( A \cap B = \{ 0 \}, \) and \( X \) is \( F \)-space (complete and metrizable) then \( A + B \) is closed. Finally, (3) deduce that if \( A \) and \( B \) are closed vector subspaces and \( \overline{A} + \overline{B} \) is finite dimensional and \( X \) is \( F \)-space then \( A + B \) is also closed. What about the general case? Hint: for (2) note that the mapping \( (a, b) \mapsto a + b \) is a one-to-one application from \( A \times B \) onto \( A + B, \) and use the open mapping theorem as
in Remark 2.22 to deduce that any Cauchy sequence of the form \( \{a_n + b_n\} \) is pre-mapped from Cauchy sequences \( \{a_n\} \) and \( \{b_n\} \); for (3) use Remark 2.15 to know that any finite dimensional subspace of a topological vector space is necessarily closed.

**Proof.** The sequences \( A = \{n + 1/n : n \geq 2\} \) and \( B = \{-n + 1/n : n \geq 2\} \) are closed and unbounded subsets of \( \mathbb{R} \), but \( A + B \) does not contain the origin, yet, \((n+1/n) + (-n+1/n) = 2/n \to 0\), which means that \( A + B \) is not a closed set.

(1) If \( A \) is closed and \( B \) is compact then take any sequence \( \{a_n + b_n = c_n\} \) in \( A + B \) convergent to some limit \( c \). Because \( \{b_n\} \subset B \) and \( B \) is compact, there exists a subsequence \( b_{n_k} \) convergent to some \( b \) in \( B \), and the continuity of the addition shows that \( a_{n_k} = c_{n_k} - b_{n_k} \) must converge to some \( a \). Since \( A \) is closed, the limit \( a \) belongs to \( A \), and then \( c_{n_k} = a_{n_k} + b_{n_k} \to a + b \), proving that the limit \( c \) belongs to \( A + B \), i.e., \( A + B \) is a closed set.

(2) If \( A \) and \( B \) are independent (non-null) closed vector subspaces, i.e., \( a + b = 0 \) implies \( a = b = 0 \), then any element \( c \) in \( A + B \) can be uniquely written as \( c = a + b \), with \( a \) in \( A \) and \( b \) in \( B \). Because \( A \) and \( B \) are closed vector spaces of a \( F \)-space, they are in themselves \( F \)-spaces, and \( A + B \) is metrizable space, a priori, not necessarily complete. The open mapping Theorem 2.20 (in the form of Remark 2.22) can be applied to the continuous and onto mapping \( T : (x, y) \mapsto x + y \), from the product space \( A \times B \) into \( A + B \), to deduce that \( T \) is an open operator. This implies that if \( \{c_n = a_n + b_n\} \) is a Cauchy sequence, so are \( \{a_n\} \) and \( \{b_n\} \). Hence, if \( a_n + b_n = c_n \to c \) then \( \{c_n\}, \{a_n\} \) and \( \{b_n\} \) are Cauchy sequences. Since the space \( X \) is complete, all sequences are convergent, i.e., \( a_n \to a \) and \( b_n \to b \), and because \( A \) and \( B \) are closed, the limit points \( (a, b) \) belong to \( A \times B \) and \( c = a + b \). This shows that \( A + B \) is closed.

(3) In view of Remark 2.15, any finite dimensional subspace of a topological vector space is closed. Now, if \( A \) and \( B \) are (non-null) closed vector subspaces of a \( F \)-space \( X \) and \( A \) is finite dimensional then choose a base \( \{x_1, \ldots, x_r\} \) for \( A \) and if necessary, re-ordered the vectors in such a way that \( x_1, \ldots, x_k \) belongs to \( A \cap B \), and \( x_{k+1}, \ldots, x_r \) belongs to \( A \) but not to \( B \). Denote by \( A' \) the (closed) vector space generated by the vector \( \{x_{k+1}, \ldots, x_r\} \). Because \( B \) is a vector space, all linear combination of the vectors \( \{x_1, \ldots, x_r\} \) belong to \( B \), which proves that \( A + B = A' + B \). Since \( A' \) and \( B \) are linearly independent, apply the previous result (2) to deduce that \( A' + B \) is closed, i.e., \( A + B \) is closed.

Alternatively, because the image \( T(A \times B) = A + B \) is necessarily a \( F \)-space, the space \( A + B \) is complete, which implies that \( A + B \) is a complete subset of the \( F \)-space \( X \), and so \( A + B \) is closed. In this argument, the fact that \( A \) and \( B \) are independent is not used, i.e., if \( A \) and \( B \) are two closed vector subspaces then so is \( A + B \). \( \square \)

**Exercise 3.4.** Prove that a locally convex (Hausdorff space) is normable (i.e., there exists a norm yielding the same topology) if and only if its zero vector has a bounded neighborhood. For instance, the reader may consult the book Al-Gwaiz [4, Theorem 1.6, p.15], among others.
Proof. Recall that, by definition, a subset $B$ of a topological vector space $X$ is bounded if it can be absorbed by any neighborhood of zero, i.e., for every open subset $O$ containing 0 there exists $t > 0$ such that $B \subset tO = \{tv : v \in O\}$.

It is clear that the unit open ball $\{x \in X : \|x\| < 1\}$ is a bounded neighborhood of the zero. Conversely, if $X$ is a lctvs with a bounded neighborhood $U$ of zero then there exists a continuous seminorm $p$ such that $\{x \in X : p(x) < 1\} = B \subset U$. To check that $p$ is indeed a norm, take $x$ such that $p(x) = 0$. Since the ball $B$ is a bounded neighborhood of zero, there exists $r > 0$ such that $rx$ belongs to $B$, and any neighborhood $V$ of zero must absorb $B$, i.e., there exists $t > 0$ such that $B \subset tV$. Hence the vector $y = (r/t)x$ belongs to every neighborhood $V$ of zero, and since the lctvs is a Hausdorff space (separate points), the point $y$ must be zero, i.e., $x = 0$, which means that $p$ is a norm. 

### (3.1.1) Dual Spaces

### (3.1.2) Inductive Limits

**Exercise 3.5.** On a barrel lctvs $X$ and its dual space $X'$, (1) show that a weakly* bounded sequence in the dual space $X'$ is also strongly bounded. Finally, assume that $X$ satisfies the Heine-Borel property, i.e., every closed and bounded set is compact, and (2) prove that any sequence is strongly convergence in the dual space $X'$ if and only if it is weakly* convergence.

**Proof.** (1) By definition, if $\{f_n\}$ is a weakly* bounded sequence then for every $x$ there exists a constant $C = C(x)$ such that $|\langle f_n, x \rangle| \leq C$, for every $n$, i.e., it is pointwise bounded. Because $X$ is a barrel lctvs, the uniform bounded principle Theorem 2.17 can be applied to deduce that the sequence $\{f_n\}$ is equi-continuous, i.e., there exists a continuous seminorm $p$ on $X$ such that $|\langle f_n, x \rangle| \leq p(x)$, for every $x$ in $X$ and any $n$, i.e., the sequence $\{f_n\}$ is strongly bounded.

(2) We have to show that if a sequence $\{f_n\}$ is weakly* convergent then it is also strongly convergent, i.e., if $\langle f_n, x \rangle \to 0$ for every $x$ then $\langle f_n, x \rangle \to 0$ uniformly for $x$ within any bounded set $B$ of $X$. However, a weakly* convergence sequence is weakly* bounded, and so, invoking (1), the sequence is also strongly bounded, i.e., there exists a continuous seminorm $p$ such that $|\langle f_n, x \rangle| \leq p(x)$, for every $x$ in $X$ and $n \geq 1$. Because $B$ is relatively compact in $X$, for any $\varepsilon > 0$ there exists a finite number of points $x_1, \ldots, x_r$ in $B$ such that $p(x - x_i) < \varepsilon$ implies that $x$ belongs to $B$. Therefore, the estimate

$$\sup_{x \in B} |\langle f_n, x \rangle| \leq \min_i p(x - x_i) + \max_i |\langle f_n, x_i \rangle| \leq \varepsilon + \sum_{i=1}^r |\langle f_n, x_i \rangle|,$$

implies the desired conclusion. 

### (3.1.3) Test Function Spaces

**Exercise 3.6.** Similar to Exercise 2.1, discuss the spaces $C^0_0(\Omega)$. 

Proof. The argument is similar to the one used on the space $C_0(\Omega)$. First, define the subspace spaces $C_\theta^\prime(K)$ of functions in $C_\theta^0(\Omega)$ with support in $K$ (or vanishing on the boundary $\partial K$ and extended by zero). The seminorms

$$
[f]_{\theta,\Omega} = \sup_{x,y \in \Omega, \ x \neq y} \{|f(x) - f(y)| \ |x - y|^{-\theta}\},
$$

for any compact domain $K$ of $\Omega$, and the sup-norm

$$
\|f\|_{\infty,\Omega} = \sup_{x \in \Omega} |f(x)|,
$$

make $C_\theta^0(\Omega)$ a Banach space (in particular, a complete lctvs). Then $C_\theta^0(\Omega)$ with the above seminorms yield the inductive limit topology on $C_\theta^0(\Omega)$, i.e., $f_n \to f$ if and only if (a) there exists a compact domain $K$ of $\Omega$ such that all $f_n$ belong to the same $C_\theta^\prime(K)$, and (b) $f_n \to f$ in $C_\theta^\prime(K)$. It is clear that this means that (a) there exists a compact $K$ of $\Omega$ such that $f_n(x) = 0$ for every $x$ outside $K$ and any $n$, and (b) $\|f_n - f\|_{\infty,K} + [f_n - f]_{\theta,\Omega} \to 0$ as $n \to \infty$.

Referring to Exercise 2.1, we deduce that the inclusion of $C_\theta^0(\Omega)$ into $C_\theta^\prime(\Omega)$, with $0 \leq \theta' < \theta \leq 1$ is compact and that $C_\theta^0(\Omega)$ is not separable.

Exercise 3.7. Let $Z_a$ be the space of (complex) entire functions $f : \mathbb{C} \to \mathbb{C}$ of exponential type $a > 0$, namely, for each $k \geq 0$ there exists a constant $C_p$ such that

$$(1 + |z|)^k |f(z)| \leq C_k e^{a|y|}, \quad \forall z = x + iy \in \mathbb{C}.$$ 

Consider the family of seminorms given by

$$p_k(f) = \sup \{e^{-a|y|}(1 + |z|)^k |f(z)| : z = x + iy \in \mathbb{C}\},$$

and discuss the “inductive limit generated”, see Friedman [45, Section 2.3, pp 33–34].

Proof. To define the inductive limits we consider the spaces $Z_{a,k}$ of (complex) entire functions $f : \mathbb{C} \to \mathbb{C}$ of exponential type $a > 0$ and such that $e^{-a|y|}(1 + |z|)^k |f(z)| \to 0$ as $|z| \to \infty$. It is clear that $p_k(\cdot)$ is a norm on $Z_{a,k}$, and because the uniform limit of complex entire functions is a complex entire function, the space $Z_{a,k}$ is a Banach space. However in this case, $Z_{a,k} \supset Z_a$ and $Z_a$ is a complete lctvs with the countable family of seminorms $\{p_k(\cdot) : k = 0,1,\ldots\}$, i.e., a Fréchet space. Thus, there is not need to introduce a inductive limit generated. By the way, note that the only entire function with compact support is the identically zero function.

(3.2) Calculus with Distributions

Exercise 3.8. Verify that for $x$ in $\mathbb{R}$, the expression

$$\langle p.v.(1/x), \varphi \rangle = \lim_{\varepsilon \to 0} \int_{|x| > \varepsilon} \frac{\varphi(x)}{x} \, dx = \int_0^\infty \frac{\varphi(x) - \varphi(-x)}{x} \, dx,$$
defines a distribution in $\mathbb{R}$. Moreover, let $f$ be a continuous function $f$ in $\mathbb{R}^d \setminus \{0\}$ which is positively homogeneous of degree $-d$ and has mean zero on the unit sphere $\{x : |x| = 1\}$, i.e.,

$$f(\lambda x) = \lambda^{-d} f(x), \quad \forall x \in \mathbb{R}^d, \lambda > 0 \quad \text{and} \quad \int_{|x|=1} f(x')dx' = 0,$$

where $dx'$ denotes the surface area measure on the unit sphere. Show that the expression

$$\langle p.v. (f) , \varphi \rangle = \lim_{\varepsilon \rightarrow 0} \int_{|x| > \varepsilon} f(x) \varphi(x) dx,$$

defines a distribution in $\mathbb{R}^d$.

**Proof.** Because the region $\{|x| > \varepsilon\}$ is symmetric, we can write

$$\int_{|x| > \varepsilon} \frac{\varphi(x)}{x} dx = \int_{\varepsilon}^{\infty} \frac{\varphi(x) - \varphi(-x)}{x} dx,$$

which prove that the expression $p.v. (1/x)$ defines a linear functional on $D(\mathbb{R})$. Moreover,

$$\varphi(x) - \varphi(-x) = \int_{-1}^{1} x \varphi'(xt) dt,$$

i.e.,

$$\langle p.v. (1/x) , \varphi \rangle = \int_{0}^{\infty} dx \int_{-1}^{1} \varphi'(xt) dt.$$

Thus, if a sequence $\{\varphi_n\}$ in $D(\mathbb{R})$ (actually, $C^1_0(\mathbb{R})$ suffices) satisfies (1) there exists $r > 0$ such that $\varphi_n(x) = 0$ if $|x| > r$ for every $n$, and (2) $\varphi_n \rightarrow \varphi$ and $\varphi'_n \rightarrow \varphi'$ uniformly, then $\langle p.v. (1/x) , \varphi_n \rangle \rightarrow \langle p.v. (1/x) , \varphi \rangle$, i.e., expression $p.v. (1/x)$ defines a distribution on $\mathbb{R}$. This is usually refers to as the principal value (valuer principal) of $1/x$.

For higher powers, e.g., $1/x^2$, the improper (symmetric) integral defined as the limit

$$\lim_{\varepsilon \rightarrow 0} \int_{|x| > \varepsilon} \frac{\varphi(x)}{x} dx$$

does not have a proper meaning, and the principal value is replaced by the Hadamard finite part, i.e.

$$\langle f.p. (1/x^2) , \varphi \rangle = \int_{0}^{\infty} \frac{\varphi(x) + \varphi(-x) - 2\varphi(0)}{x^2} dx,$$
and
\[ \langle f.p.(1/x^3), \varphi \rangle = \int_0^\infty \frac{\varphi(x) - \varphi(-x) + 2x\varphi'(0)}{x^3} dx, \]
and so on for other powers. Note that in the distribution sense we have
\[ [p.v.(1/x)]' = f.p.(-1/x^2), \]
\[ [f.p.(1/x^2)]' = f.p.(-2/x^3), \]
and so on.

Similarly, because \( f \) is homogeneous of degree \( n \), by means of spherical coordinates we obtain
\[ \int_{|x| > \varepsilon} f(x)\varphi(x) dx = \int_{\varepsilon}^\infty \rho^{d-1}d\rho \int_{|x| = 1} f(\rho x')\varphi(\rho x') dx' = \]
\[ = \int_{\varepsilon}^\infty \rho^{-1}d\rho \int_{|x| = 1} f(x')[\varphi(\rho x') - \varphi(0)] dx'. \]
Essentially the same one-dimensional argument above yields
\[ \langle p.v.(f), \varphi \rangle = \int_0^\infty d\rho \int_{|x'| = 1} f(x')dx' \int_0^1 x' \cdot \nabla \varphi(t \rho x') dt, \]
which proves that the expression \( p.v.(f) \) is indeed a distribution in \( \mathbb{R}^d \).

**Exercise 3.9.** Consider the function \( x \mapsto \ln |x| \) as a distribution in \( \mathbb{R}^d \) and calculate its first order derivatives.

**Proof.** The function \( \ln |x| \) is in \( L^1(\mathbb{R}^d) \) and its derivative in the distribution sense with respect to \( x_1 \) is given by
\[ \langle \partial_1 \ln |x|, \varphi \rangle = -\int_{\mathbb{R}^d} \ln |x| \partial_1 \varphi(x) dx \]
\[ = -\lim_{\varepsilon \to 0} \int_{\mathbb{R}^{d-1}} dx' \int_{|x_1| > \varepsilon} \ln |x| \partial_1 \varphi(x') dx_1, \]
where \( x = (x_1, x') \) and \( \mathbb{R}^{d-1} \) refers to the variable \( x' = (x_2, \ldots, x_d) \). Integrating by parts,
\[ -\int_{|x_1| > \varepsilon} \ln |x| \partial_1 \varphi(x) dx_1 = \ln |(\varepsilon, x')| (\varphi(-\varepsilon, x') - \varphi(\varepsilon, x')) + \]
\[ + \int_{|x_1| > \varepsilon} \frac{x_1}{|x|^2} \varphi(x) dx_1. \]
Since
\[ \varphi(-\varepsilon, x') - \varphi(\varepsilon, x') = \int_{-1}^1 \varepsilon \partial_1 \varphi(\varepsilon t, x') dt, \]
the limit of the integral with \( \ln |(\varepsilon, x')| \) vanishes and we deduce that
\[ \langle \partial_1 \ln |x|, \varphi \rangle = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^{d-1}} dx' \int_{|x_1| > \varepsilon} \frac{x_1}{|x|^2} \varphi(x) dx_1 = \]
\[ = \int_{\mathbb{R}^{d-1}} dx' \int_0^\infty \frac{x_1}{|x|^2} (\varphi(x_1, x') - \varphi(-x_1, x')) dx. \]
Note that
\[ \varphi(x_1, x') - \varphi(-x_1, x') = x_1 \int_{-1}^{1} \partial_1 \varphi(tx_1, x') \, dt, \]
and that \( x \mapsto x_1^2/|x|^2 \) in a locally integrable functions. Briefly, the chain rule yields \( \partial_1 \ln |x| = \frac{x}{|x|^2} \) and the derivative with respect to \( x_i \) in the sense of distribution of \( \ln |x| \) is the principal value of \( x \mapsto x_i/|x|^2 \), e.g.,
\[ \langle \partial_1 \ln |x|, \varphi \rangle = \int_{\mathbb{R}^{d-1}} dx' \int_{0}^{\infty} \frac{x_1^2}{|x|^2} \, dx_1 \int_{-1}^{1} \partial_1 \varphi(tx_1, x') \, dt. \]
Remark the particular one-dimensional case, \( (\ln |x|)' = p.v.(1/x) \) as in the previous Exercise 3.8. Moreover, considering only the one-dimensional half-space, in particular, if \( \Omega = \mathbb{R}^d \), then
\[ \langle \ln(x^+)', \varphi \rangle = -\langle \ln(x^+), \varphi' \rangle = \lim_{\varepsilon \to 0} \left[ \int_{\varepsilon}^{\infty} x \varphi(x) \, dx + \varphi(0) \ln \varepsilon \right] = \int_{0}^{\infty} \frac{1}{x} (\varphi(x) - \varphi(0)) \, dx, \]
i.e., \( (\ln x^+)' = f.p.(1/x^+) \) in \( \mathcal{D}'(\mathbb{R}) \), which means the Hadamard finite part. It is clear that no singularity exists when \( (\ln x)' = 1/x \) is regarded as a distribution in \( \mathbb{R}^+ \). In any case, note that \( p.v.(1/|x|) = f.p.(1/x^+) - f.p.(1/x^-) \).

**Exercise 3.10.** Discuss (a) the translation operator \( \tau_h \) defined as \( \tau_h \varphi(x) = \varphi(x+h) \) with \( \Omega_h = h + \Omega \) and (b) the reflection operator \( \tilde{\varphi}(x) = \varphi(-x) \).

**Proof.** It is clear that the translation operator \( \tau_h \) is a linear continuous mapping from \( \mathcal{D}(\Omega) \) into \( \mathcal{D}(\Omega_h) \), with \( \Omega_1 = h + \Omega \). If \( T_f \) is the distribution associated with a locally integrable function \( f \) then
\[ \langle \tau_h T_f, \varphi \rangle = \int_{\Omega} (x+h) \varphi(x) \, dx = \int_{\Omega_h} f(y) \varphi(y-h) \, dy = \langle T_f, \tau_{-h} \varphi \rangle. \]
Thus, for any distribution in \( \Omega \) or equivalently, for any element in \( \mathcal{D}'(\Omega) \), the translation operator is defined by \( \langle \tau_h T, \varphi \rangle = \langle T, \tau_{-h} \varphi \rangle \), for any \( \varphi \) in \( \mathcal{D}(\Omega_h) \). Hence, \( \tau_h \) is also a linear continuous mapping from \( \mathcal{D}'(\Omega) \) into \( \mathcal{D}'(\Omega_h) \). In particular, if \( \Omega = \mathbb{R}^d \) then \( \tau_h \) maps \( \mathcal{D}'(\mathbb{R}^d) \) into itself.

Considering test functions with support in a fixed compact \( K \subset \Omega \), i.e., on the subspace \( \mathcal{D}_K(\Omega) \), the translation \( \tau_h \) can be regarded as a linear continuous operator from \( \mathcal{D}_K(\Omega) \) into itself provided \( h \) is sufficiently small, i.e., \( |h| \) smaller than the distance from the compact \( K \) to the boundary \( \partial \Omega \).

The reflection operator \( \tilde{\varphi}(x) = \varphi(-x) \) makes sense as a linear continuous mapping from \( \mathcal{D}(\Omega) \) into \( \mathcal{D}(\Omega) \), with \( \Omega = \{-x : x \in \Omega\} \). As in the case of the translation operator, the reflection operator is defined for distribution by transposition, i.e., \( \langle \tilde{T}, \varphi \rangle = -\langle T, \tilde{\varphi} \rangle \), for any \( \varphi \) in \( \mathcal{D}(\Omega) \).

If \( \Omega \) is symmetric, i.e., \( \Omega = \Omega \), then the reflection is a linear continuous mapping from \( \mathcal{D}(\Omega) \) into itself, in particular, this applies to \( \Omega = \mathbb{R}^d \). \( \square \)
Exercise 3.11. For a unit vector $e$ in $\mathbb{R}^d$, consider the expression $\Lambda_{e,t}\varphi(x) = [\varphi(x + te) - \varphi(x)]/t$, for $t > 0$. Discuss (a) the directional rate operator $\Lambda_{e,t}$ as defined on either $\mathcal{D}(\mathbb{R}^d)$ or $\mathcal{D}(\Omega)$, and (b) extend the definition of $\Lambda_{e,t}$ as a linear continuous operator on the spaces of distributions, i.e., on $\mathcal{D}$, $\mathcal{E}$ and $\mathcal{S}$. Moreover, also discuss (c) the iteration $\Lambda_{e,t}\Lambda_{e,-t}$ written as $\Lambda_{e,t}^2\varphi(x) = \varphi(x + et) - \varphi(x - et) - 2\varphi(x)/t^2$, and then (d) consider the continuity of the directional derivative $\lim_{t \to 0} \Lambda_{e,t}$ and the Hessian $\lim_{t \to 0} \Lambda_{e,t}^2$ as operator acting on distributions.

Proof. (a) Considering the directional rate operator $\Lambda_{e,t}$ as a linear continuous on $\mathcal{D}(\mathbb{R}^d)$ is immediately, while on the space $\mathcal{D}(\Omega)$, two steps are necessary. For instance, we may consider the inclusion $\mathcal{D}(\Omega) \subset \mathcal{D}(\mathbb{R}^d)$ and define $\Lambda_{e,t}$ from $\mathcal{D}(\Omega)$ into $\mathcal{D}(\mathbb{R}^d)$, for every $t > 0$ and any unit vector $e$ in $\mathbb{R}^d$. Alternatively, we may first consider the directional rate operator $\Lambda_{e,t}$ as acting on the subspace $\mathcal{D}_K(\Omega)$, with $t > 0$ smaller than the distance from the compact $K$ to the boundary $\partial \Omega$.

(b) By transposition, the directional rate operator $\Lambda_{e,t}$ is defined as linear continuous operator from $\mathcal{D}$ (or $\mathcal{E}$, or $\mathcal{S}$) into itself. If $T_f$ is the distribution associated with a locally integrable function $f$ then

$$\langle \Lambda_{e,t}T_f, \varphi \rangle = \int_{\mathbb{R}^d} \Lambda_{e,t} f(x) \varphi(x) dx = \int_{\mathbb{R}^d} f(y) \Lambda_{e,-t} \varphi(y) dy = \langle T_f, \Lambda_{e,-t} \varphi \rangle.$$ 

for any $\varphi$ in $\mathcal{D} = \mathcal{D}(\mathbb{R}^d)$. Thus, for any distribution in $\mathbb{R}^d$ or equivalently, for any element in $\mathcal{D}' \supset S' \supset \mathcal{E}'$, the translation operator is defined by $\langle \Lambda_{e,t}T, \varphi \rangle = \langle T, \Lambda_{e,-t} \varphi \rangle$, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$. Hence, $\Lambda_{e,t}$ becomes a linear continuous mapping from $\mathcal{D}'$ (or $\mathcal{E}'$, or $\mathcal{S}'$) into itself.

(c) Initially, $\Lambda_{e,t}$ is defined for any unit vector $e$ and any $t > 0$, but clearly, this also make sense for any $t < 0$. The iteration

$$\Lambda_{e,t}\Lambda_{e,-t}\varphi(x) = \Lambda_{e,-t}\Lambda_{e,t}\varphi(x) = \frac{\varphi(x + et) + \varphi(x - et) - 2\varphi(x)}{t^2},$$

which is denoted by $\Lambda_{e,t}^2$ is a linear continuous operator from $\mathcal{D}$ (or $\mathcal{E}$, or $\mathcal{S}$) into itself. Moreover, again by transposition, the expression $\langle \Lambda_{e,t}^2T, \varphi \rangle = \langle T, \Lambda_{e,t}^2 \varphi \rangle$, defines $\Lambda_{e,t}^2$ as a linear continuous mapping from $\mathcal{D}'$ (or $\mathcal{E}'$, or $\mathcal{S}'$) into itself. Note the symmetry of $\Lambda_{e,t}^2$ and the fact that $\Lambda_{e,t}\Lambda_{e,-t} = \Lambda_{e,-t}\Lambda_{e,t} = 0$.

(d) Because the derivative operator is continuous, the limits as $t \to 0$ are well defined as linear continuous operators from $\mathcal{D}$ (or $\mathcal{E}$, or $\mathcal{S}$) into itself, and also, from $\mathcal{D}'$ (or $\mathcal{E}'$, or $\mathcal{S}'$) into itself. It is also clear that considering the unit vector as a parameter (or even taken any vector, not necessarily of unit length) the directional derivative $\lim_{t \to 0} \Lambda_{e,t} = e \cdot \nabla$ is linear in $e$ and the Hessian $\lim_{t \to 0} \Lambda_{e,t}^2 = (e \cdot \nabla)^2$ is bilinear in $e$. Actually, the directional derivative can be regarded as the gradient operator, i.e., mapping real-valued functions (or distributions) into vector-valued functions (or distributions), meaning $(e_i \cdot \nabla)$ with $\{e_i\}$ an orthonormal basis in $\mathbb{R}^d$. The Hessian can be regarded as mapping real-valued functions (or distributions) into symmetric matrix-valued functions.
(or distributions), i.e., the Hessian matrix operator $H = (H_{ij})$ can be defined as $H_{ij} = (a_{ij} \cdot \nabla)^2 - b_{ij} \cdot \nabla)^2)/2$, where the unit vectors $a_{ij}$ and $b_{ij}$ are given by $a_{ij} = (e_i + e_j)/\sqrt{2}$ and $b_{ij} = (e_i - e_j)/\sqrt{2}$.

### 3.2.1 Positivity, Differentiability and Integrability

**Exercise 3.12.** Let $f$ be a real-valued function defined on a convex open set $\Omega$ of $\mathbb{R}^d$. Recall that $f$ is called convex whenever $f(sx + ty) \leq sf(x) + tf(y)$, for every $x, y$ in $\Omega$ and any $s, t \geq 0$, $s + t = 1$. Also, $f$ is called concave if $-f$ is convex. Assuming that $f$ is twice continuously differentiable, (a) prove that $f$ is convex if and only if the Hessian matrix $D^2 f$ is nonnegative definite, i.e., $(v, D^2 f(x)v) \geq 0$ for every $v$ in $\mathbb{R}^d$ and any $x$ in $\Omega$. Now, a function $f$ is called semi-convex (or semi-concave) if there exists a twice continuously differentiable function $g$ such that $f + g$ is convex (or concave). Prove that (b) if locally integrable function $f$ is semi-convex and also semi-concave then the Hessian matrix $D^2 f$, regarded as a matrix-valued distribution, is actually a locally bounded matrix-valued function.

**Proof.** (a) This part is rather standard. For given $x$ and $y$ in $\Omega$ consider the function $F(r) = f((1 - r)x + ry) = f(x + r(y - x))$ for any $r$ in $[0, 1]$. Because $F(0) = f(x)$ and $F(1) = f(y)$, it is clear that $f$ is convex if and only if $r \mapsto F(r)$ is convex in $[0, 1]$ for every $x, y$. Now, the function of one variable $F$ is twice continuously differentiable and therefore, $F$ is convex if and only if $F'' \geq 0$. Hence, by means of the chain rule, $F'(r) = (y - x)Df((x + r(y - x))$ and $F''(r) = ((y - x), D^2f((x + r(y - x))(y - x)) \geq 0$, for every $r$ in $[0, 1]$, i.e., $f$ is convex if and only if the Hessian matrix is nonnegative definite.

(b) We make use of the Hessian approximation $\Lambda^2_{e,t}\varphi(x) = \varphi(x + et) + \varphi(x - et) - 2\varphi(x)]/t^2$ and its extension to distributions $\langle \Lambda^2_{e,t}T, \varphi \rangle = \langle T, \Lambda^2_{e,t}\varphi \rangle$, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$. Note that any convex function satisfies $\Lambda^2_{e,t}f(x) \geq 0$ for every $x$ and $t > 0$. The Hessian of a distribution $T$ is denoted by the matrix-valued distribution $D^2T$ of by the real-valued distribution $D^2T(u, v)$, for any vector $u, v$ in $\mathbb{R}^d$.

If $f$ is a semi-convex function then there exists a twice continuously differentiable function $g$ such that $f + g$ is convex. Therefore $\Lambda^2_{e,t}[f(x) + g(x)] \geq 0$ for every $x$ and $t > 0$. Considering the distribution $T_f$ induced by $f$ we deduce

$$\langle \lim_{t \to 0} \Lambda^2_{e,t}T_f, \varphi \rangle \geq -\langle \lim_{t \to 0} \Lambda^2_{e,t}g, \varphi \rangle = -\int_\Omega \langle e, D^2g(x)e \rangle \varphi(x)dx,$$

for every test function $\varphi \geq 0$. This proves that $D^2T_f(e, e) + (e, D^2g(\cdot)e) \geq 0$. Moreover, if $f$ is also semi-concave then there exists a twice continuously differentiable function $h$ such that $f + h$ is concave. This yields $T_f(e, e) + (e, D^2h(\cdot)e) \leq 0$. Invoking Proposition 3.20, the distribution $T_f$ can be identified with a locally integrable function, denoted by $D^2f$, which coincides with the pointwise Hessian of $f$. Note that pointwise Hessian of a convex (or concave) function $f$ is defined as the monotone limit of $\Lambda^2_{e,t}f(x)$, for every $x$. It is clear that the limit $\Lambda^2_{e,t}f(x)$ exists also when $f$ is semi-convex or semi-concave.
In general, we say that a distribution $T$ is convex (or concave) if
\[ \langle \Lambda^2_{e,t} T, \varphi \rangle = \langle T, \Lambda^2_{e,t} \varphi \rangle \geq 0, \quad \forall t > 0, e, \]
for any $\varphi \geq 0$ in $D(\mathbb{R}^d)$. This implies that $\Lambda^2_{e,t} T$ is a nonnegative measure for every $t > 0$ and so is the limit distribution $\lim_{t \to 0} \Lambda^2_{e,t} T = D^2 T(e,e)$. Because
\[ \Lambda^2_{e,t} \varphi(x) = \int_0^1 dr \int_{-1}^1 (e, \varphi(x + trse)e) ds, \]
we deduce that $T$ is a convex (or concave) distribution if and only if $D^2 T(e,e)$ is a nonnegative (nonpositive) Radon measure, for every unit vector $e$ in $\mathbb{R}^d$, i.e., $D^2 T(e,e)$ is a nonnegative (nonpositive) element of the dual space $C_0^0(\Omega)'$.

Next, recall that a (signed) Radon measure $\mu$ can be identified with an element in $C_0^0(\Omega)'$, and that the restriction of an element $T$ of $C_0^0(\Omega)'$ to any compact $K \subset \Omega$ is actually identified with a (signed) Radon measure on $K$. Thus, a distribution $T$ is semi-convex (or semi-concave) if there exists an element $S$ in the dual space $C_0^0(\Omega)'$ such that $T + S$ is a convex (or concave) distribution. This means that the only semi-convex (or semi-concave) distributions are the elements of the dual space $C_0^0(\Omega)'$, i.e., distributions of order zero. \hfill \Box

**Exercise 3.13.** Give more detail on assertion (e) above, namely, use Proposition 3.20 to show that for any open interval $I$ in $\mathbb{R}$ and any element $T$ in $\mathcal{D}'(I)$ we have (1) if $T' \geq 0$ then $T = T_f$ is the distribution associated to some increasing function $f$; (2) if $T'' \geq 0$ then $T = T_f$ is the distribution associated to some convex function $f$. Moreover, (3) if $T'$ is a signed Radon measure on any compact sub-interval of $I$ then $T = T_f$ is the distribution associated to some function $f$ with bounded variation on every compact sub-interval of $I$ (i.e., $f$ has locally bounded variation on the open interval $I$); and finally (4) if $T''$ is a signed Radon measure on any compact sub-interval of $I$ then $T = T_f$ is the distribution associated to some function $f$ which is a difference of two convex functions.

**Proof.** Proposition 3.20 applied to $T' \geq 0$ implies that the distribution $T' = \mu$ is actually a (non-negative) Radon measure on $I$. Thus, if $a$ belongs to $I$ then the cad-lag non-decreasing function $f$ defined by $x \mapsto \mu([a,x])$ for $x \geq a$ and $x \mapsto \mu([x,a])$ for $x < a$ satisfies
\[ \langle T', \varphi \rangle = \int_{-\infty}^{+\infty} \varphi(x) \, df(x), \quad \forall \varphi \in \mathcal{D}(\mathbb{R}), \]
where the integral can be also considered in the Riemann-Stieltjes sense, where integration by parts shows that
\[ \langle T', \varphi \rangle = -\int_{-\infty}^{+\infty} \varphi'(x) \, f(x) \, dx \quad \forall \varphi \in \mathcal{D}(\mathbb{R}). \]
On the other hand, if $\varphi_0$ is a test function satisfying
\[
\int_{-\infty}^{+\infty} \varphi_0(x) \, dx = 1
\]
then the distribution $T$ can be represented as
\[
\langle T, \varphi \rangle = \langle T, \varphi_0 \rangle \langle 1, \varphi \rangle + \langle T', \int_{-\infty}^{+\infty} [\langle 1, \varphi \rangle \varphi_0(x) - \varphi(x)] \, dx \rangle,
\]
which yields
\[
\langle T, \varphi \rangle = \langle T, \varphi_0 \rangle \langle 1, \varphi \rangle + \int_{-\infty}^{+\infty} [\varphi(x) - \langle 1, \varphi \rangle \varphi_0(x)] f(x) \, dx,
\]
i.e., except for a constant $T$ is identified with $f$, namely, with
\[
g = f + \langle T, \varphi_0 \rangle - \int_{-\infty}^{+\infty} \varphi_0(x) f(x) \, dx
\]
we have $T = T_g$.

Similarly, if the second derivative $T'' \geq 0$ then $T'$ is identified with a non-increasing function $g$ in view of the previous assertion (1). Therefore, an antiderivative $f$ of $g$ can be identified with $T$, and $f$ is convex because $f' = g$ is a non-increasing function.

Now again, if $T$ is a distribution (element in $D(I)$) such that its first derivative $T' = \mu$ is a signed Radon measure on any compact sub-interval of $I$ then choose $a$ in $I$ to see that the cad-lag function $f$ defined by $x \mapsto \mu([a,x])$, for $x \geq a$, and $x \mapsto \mu([x,a])$, for $x < a$, has bounded variation on each compact sub-interval of $I$ and satisfies
\[
\langle T', \varphi \rangle = \int_{-\infty}^{+\infty} \varphi(x) \, df(x), \quad \forall \varphi \in D(\mathbb{R}),
\]
where the integral can be also considered in the Riemann-Stieltjes sense, where integration by parts shows that
\[
\langle T', \varphi \rangle = -\int_{-\infty}^{+\infty} \varphi'(x) f(x) \, dx \quad \forall \varphi \in D(\mathbb{R}).
\]
Then except for a constant, the distribution $T$ is identified with $f$, which has locally bounded variation on the interval $I$. Thus $f = f_+ - f_-$, where each $f_+$ and $f_-$ ia a non-increasing function, and $T = c + T_{f_+} - T_{f_-}$ for some constant $c$.

A similar argument as above shows that if $T$ is a distribution (element in $D(I)$) such that its second derivative $T'' = \mu$ is a signed Radon measure on any compact sub-interval of $I$ then $T$ is identified with the difference of two convex functions. □
Exercise 3.14. For the powers distributions $|x|^z = \exp(z \ln|x|)$, $(x^+)^z = \exp(z \ln(\max\{x, 0\}))$ and $(x^-)^z = \exp(z \ln(-\min\{x, 0\}))$ in $\mathcal{D}'(\mathbb{R})$, remove the singularity at 0 to show that they are well defined for any $z$ in $\mathbb{C}$, which is not a negative integer, e.g., see Al-Gwaiz[4, Section 2.8, pp. 63–72].

Proof. Only one case need consideration, namely, the one-dimensional distribution induced by the power function $(x^+)^z$. The other cases are deduced from this one.

For a test function $\varphi$, use Taylor polynomials of order $n$ to write

$$\varphi(x) = \sum_{k=0}^{n-1} x^k \frac{\varphi^{(k)}(0)}{k!} + x^n \int_0^1 (1-t)^{(n-1)} \frac{\varphi^{(n)}(tx)}{(n-1)!} dt,$$

where $\varphi^{(k)}(x)$ is the $k$-derivative. If $z = a + ib$ with $-2 < a + n < -1$ for some $n = 0, 1, \ldots$, then the function $x \mapsto (x^+)^{z+n}$ is a locally integrable function and the expression

$$\left\langle (x^+)^z, \varphi(x) - \sum_{k=0}^{n-1} x^k \frac{\varphi^{(k)}(0)}{k!} \mathbb{1}_{x \leq 1} \right\rangle = \int_0^1 x^{z+n} dx \int_0^1 (1-t)^{(n-1)} \frac{\varphi^{(n)}(tx)}{(n-1)!} dt$$

defines a distribution in $\mathbb{R}$. Moreover the expression

$$\left\langle (x^+)^z, \sum_{k=0}^{n-1} x^k \frac{\varphi^{(k)}(0)}{k!} \mathbb{1}_{x \leq 1} \right\rangle = \sum_{k=0}^{n-1} \int_0^1 x^{z+k} \frac{\varphi^{(k)}(0)}{k!} dx = \sum_{k=0}^{n-1} \frac{\varphi^{(k)}(0)}{k!(z+k)}$$

also defines a distribution in $\mathbb{R}$. Hence, adding both expressions, the power function $(x^+)^z$ is a distribution if $z$ is not a negative integer.

If $z = -n$ for some integer $n \geq 1$ then the finite part of the power distribution $(x^+)^{-n}$ can be defined as follows. First take a test function $\chi$ such that $\chi(x) = 1$ if $-1/2 < x < 1/2$ and $\chi(x) = 0$ if $|x| > 1$. Because $(x^+)^{-n} = (x^+)^{-n} \chi(x) + (x^+)^{-n}(1 - \chi(x))$ and the functions $x \mapsto (x^+)^{-n}(1 - \chi(x))$ is locally integrable, to define power distribution $(x^+)^{-n}$, we need only to give the meaning of $(x^+)^{-n} \chi(x)$ as its finite part. Therefore

$$\left\langle f.p.(x^+)^{-n} \chi(x), \varphi(x) \right\rangle = \left\langle (x^+)^{-n} \chi(x), \varphi(x) - \sum_{k=0}^{n-1} x^k \frac{\varphi^{(k)}(0)}{k!} \right\rangle = \int_0^1 \chi(x) dx \int_0^1 (1-t)^{(n-1)} \frac{\varphi^{(n)}(tx)}{(n-1)!} dt,$$

which is indeed an element in $\mathcal{D}'(\mathbb{R})$. The fact that a test function $\chi$ was chosen instead of $\mathbb{1}_{0<x<1}$ is needed to be able not to discuss the situation near $x = 1$. Indeed, the pointwise product $(x^+)^{-n} \chi(x)$ is defined for any (non necessarily
smooth) pointwise function $\chi$, but as a distribution $T = f.p.(x^+)^{-n}\chi$ make (a priori) sense only when $\chi$ is a test function, i.e., $\langle T, \chi \varphi \rangle = \langle T \chi, \varphi \rangle$.

Note that also we have
\[
\langle f.p. (x^+)^{-1}, \varphi(x) \rangle = \lim_{\varepsilon \to 0} \left[ \int\limits_\varepsilon^{\infty} (x^+)^{-1} \varphi(x) dx - \ln(\varepsilon) \varphi(\varepsilon) \right] = -\langle \ln(x^+), \varphi'(x) \rangle,
\]
and recalling the expression of the pointwise $n$-derivative $(\ln x)^{(n)} = (-1)^{n-1}(n-1)!x^{-n}$, and the (iterated) integration by parts
\[
\int u v^{(n)} dx = \sum_{k=0}^{n-1} (-1)^k u^{(k)} v^{(n-k-1)} + (-1)^n \int u^{(n)} v dx,
\]
we deduce
\[
\langle f.p. (x^+)^{-n}, \varphi(x) \rangle = \lim_{\varepsilon \to 0} \left[ \int\limits_\varepsilon^{\infty} (x^+)^{-n} \varphi(x) dx + (-1)^n \ln \varepsilon \varphi^{(n-1)} + \sum_{k=1}^{n-1} (-1)^{k-1} \varepsilon^{-k} \frac{\varphi^{(n-k-1)}(\varepsilon)}{(k-1)!} \right] = (-1)^n(n-1)! \langle \ln(x^+), \varphi^{(n)}(x) \rangle,
\]
i.e., the $n$-derivative of the logarithm is given by $(\ln(x^+))^{(n)} = (-1)^{n-1}(n-1)! f.p.(x^+)^{-n}$ in the distribution sense $\mathcal{D}'(\mathbb{R})$.

**Exercise 3.15.** With the previous notation on fractional integrals, verify that $\Phi_{\nu} * \Phi_{\mu} = \Phi_{\nu+\mu}$ and deduce that $I_t^\nu I_t^\mu = I_t^{\nu+\mu}$. Moreover, if $p, q$ belong to $[0, \infty)$ and $0 < \nu < 1$ then $I_t^\nu$ is a bounded operator from $L^p$ into $L^q$ if $1 < p < 1/\nu$ and $q = p/(1 - p\nu)$, i.e., such that
\[
\int_0^\infty dt \left\| \int_0^t (t-s)^{\nu-1} f(s) ds \right\|^q \leq C \|f\|_p^q,
\]
for some a constant $C = C_{p,q,\nu}$.

**Proof.** Recall that $\Phi_{\nu}(t) = t^{\nu-1}/\Gamma(\nu)$ and the Gamma function is given by (3.7). Thus, to show that $\Phi_{\nu} * \Phi_{\mu} = \Phi_{\nu+\mu}$ calculate
\[
(\Phi_{\nu} * \Phi_{\mu})(t) = \frac{1}{\Gamma(\nu) \Gamma(\mu)} \int_0^t (t-s)^{\nu-1} s^{\mu-1} ds
\]
with the change of variable $s = tx$ to get
\[
(\Phi_{\nu} * \Phi_{\mu})(t) = \frac{1}{\Gamma(\nu) \Gamma(\mu)} t^{\nu+\mu-1} \int_0^1 (1-x)^{\nu-1} x^{\mu-1} dx.
\]
The conclusion follows from the classic equality involving the Gamma and the Beta functions, namely
\[
\int_0^1 (1-x)^{\nu-1} x^{\mu-1} dx = B(\nu, \mu) = \frac{\Gamma(\nu) \Gamma(\mu)}{\Gamma(\nu + \mu)}.
\]
To prove this equality, begin exchanging the order of integration in the convolution
\[
\int_0^\infty dx \int_y^\infty e^{-(x-y)}(x-y)^{\nu-1}e^{-y\mu-1}dy = 
\int_0^\infty dx \int_y^\infty e^{-(x-y)}(x-y)^{\nu-1}e^{-y\mu-1}dx = \Gamma(\nu)\Gamma(\mu).
\]

Next, observe that the change of variable \(y = xt\) yields
\[
\int_x^\infty e^{-(x-y)}(x-y)^{\nu-1}e^{-y\mu-1}dy = e^{-x}x^{\nu+\mu-1} \int_0^1 (1-t)^{\nu-1}t^{\mu-1}dt,
\]
and the desired equality follows.

Since
\[
I^{\nu}\phi = \frac{1}{\Gamma(\nu)} \int_0^t (t-s)^{\nu-1}f(s)ds,
\]
the previous relation implies that \(I^{\nu}I^{\mu} = I^{\nu+\mu}\). Given \(1 < p < 1/\nu\), to apply Hölder inequality note that \(q = p/(1-\nu p)\) satisfies \(1/p + 1/q = 1\),
\[
\left| \int_0^t (t-s)^{\nu-1}f(s)ds \right|^q \leq \|f\|_p^q \int_0^t (t-s)^{(\nu-1)q}ds = \frac{\|f\|_p^q}{(\nu-1)q + 1}
\]
with \((\nu-1)q + 1 = (1-p)/(1-\nu p) > 0\) and \(C = (1-\nu p)/(1-p)\).

\[\square\]

**(3.2.2) Support and Finite Order**

**Exercise 3.16.** Show that the expressions \(\langle T, \varphi \rangle = \sum_k 2^k \varphi(1/k)\) and \(\langle S, \varphi \rangle = \sum_k 2^k \varphi^{(k)}(1/k)\) define two distributions \((0, 1) \subset \mathbb{R}\), where \(T\) is of order 0 while \(S\) is not of finite order. Check that the support of each of them is not compact. Can you modify the above expressions to produce a distribution which is not of finite order and has a compact support?

**Proof.** It is clear that the support of the distributions \(T\) or \(S\) is the set \(K = \{1/k : k = 1, 2, \ldots\}\), which is not compact. Indeed, if \(\varphi\) is supported outside of \(K\) (i.e., the support of \(\varphi\) is a subset of \((0, 1) \setminus K\)) then \(\langle T, \varphi \rangle = 0\) and \(\langle S, \varphi \rangle = 0\), because \(\varphi(n)(1/k) = 0\) for every \(k = 1, 2, \ldots\) and order \(n\) of derivative.

Since
\[
\sum_k |\varphi(1/k)| \leq 2^n \sup_{1/n < x < 1} |\varphi(x)|, \quad \forall \varphi \in \mathcal{D}(\mathbb{R})
\]
\(T\) is a distribution of order 0, while the estimate
\[
\sum_k |\varphi^{(k)}(1/k)| \leq 2^n \sup_{1/n < x < 1, m \leq n} |\varphi^{(m)}(x)|, \quad \forall \varphi \in \mathcal{D}(\mathbb{R})
\]
shows that \(S\) is also a distribution. It is also clear that \(S\) can not be of finite order.

Finally, Proposition 3.28 proves that there is no distribution with compact support which is not of finite order.

\[\square\]
Exercise 3.17. Let \( \{x_k\} \) be a sequence of points in \( \Omega \) such that the distance from \( x_k \) to the boundary \( \partial \Omega \) goes to zero, or such that \( |x_k| \to \infty \) if \( \Omega = \mathbb{R}^d \). Define \( \langle T, \varphi \rangle = \sum_k \varphi(x_k) \) and \( \langle S, \varphi \rangle = \sum_k \partial_1^k \varphi(x_k) \). Discuss if \( T \) and \( S \) are distributions, and if so, find their order and support.

Proof. Given a compact set \( K \) of \( \Omega \), there is only a finite number of points \( \{x_k\} \) in \( K \), which is denoted by \( n(K) \). Thus, the estimates

\[
\sum_k |\varphi(x_k)| \leq n(K) \sup_{x \in K} |\varphi(x)|, \quad \forall \varphi \in \mathcal{D}_K(\Omega)
\]

and

\[
\sum_k |\partial_1^k \varphi(x_k)| \leq n(K) \sup_{x \in K, k \leq n(K)} \{|\partial_1^k \varphi(x)| \}, \quad \forall \varphi \in \mathcal{D}_K(\Omega)
\]

show that \( T \) and \( S \) are elements in \( \mathcal{D}'(\Omega) \), and that \( T \) belongs to the dual space \( C_0^\infty(\Omega)' \), i.e., \( T \) is a distribution of order 0.

It is also clear that the support of both distributions is the non compact set \( \{x_k\} \). Certainly, \( S \) is not a distribution of finite order. \(\square\)

Exercise 3.18. Consider the distribution \( \langle T, \varphi \rangle = \sum_{|\alpha| \leq n} c_\alpha \partial^\alpha \varphi(x_0) \), where \( c_\alpha \) are constants, and \( x_0 \) is a point in \( \Omega \). (a) Verify that the support of \( T \) is the point \( x_0 \) and that the order is \( n \) if for some \( \alpha \) with \( |\alpha| = n \) we have \( c_\alpha \neq 0 \).

(b) Prove that the only distributions on \( \Omega \) with support equal to a simple point \( \{x_0\} \) are finite linear combinations of the derivative of the Dirac delta at \( x_0 \), i.e., as \( T \) above.

Proof. (a) If \( \varphi \) is a test function with support inside an open subset \( \Omega' \) of \( \Omega \), and \( x_0 \) does not belong to \( \Omega' \) then the function all its derivatives vanish at \( x_0 \), i.e., \( \varphi(x_0) = \partial^\alpha \varphi(x_0) = 0 \). Hence, the support of \( T \) is the singleton \( \{x_0\} \). It is clear that the order of \( T \) is \( \max\{|\alpha| : c_\alpha \neq 0\} \leq n \).

(b) If a distribution \( T \) has a singleton \( \{x_0\} \) as its support then Proposition 3.28 implies that \( T \) must be of finite order since its support is a compact set, i.e., there exists \( n \) and a constant \( C > 0 \) such that

\[
|\langle T, \varphi \rangle| \leq C \sum_{|\alpha| \leq n} \sup_{x \in \Omega} |\partial^\alpha \varphi(x)|, \quad \forall \varphi \in C^\infty(\Omega).
\]

Now, take a sequence \( \{\chi_k\} \) of test functions such that \( \chi_k = 1 \) if \( |x - x_0| \leq 1/k \) and write \( \varphi = (1 - \chi_k) \varphi + \chi_k \varphi \) to obtain \( \langle T, \varphi \rangle = \langle T, \chi_k \varphi \rangle \). If \( \partial^\alpha \varphi(0) = 0 \) for every multi-index \( \alpha \) of order \( |\alpha| \leq n \) then

\[
\sup_{x \in \Omega} |\partial^\alpha (\chi_k \varphi(x))| \to 0.
\]

Actually, this is the argument of Proposition 3.31.
Therefore, for the function \( \psi(x) = \varphi(x) - \sum_{|\alpha| \leq n} x^\alpha \partial^\alpha \varphi(0)/\alpha! \) we deduce that \( \langle T, \psi \rangle = 0 \), i.e.,

\[
|\langle T, \varphi \rangle| = \sum_{|\alpha| \leq n} c_\alpha \partial^\alpha \varphi(0), \quad \text{with} \quad c_\alpha = \langle T, x^\alpha/\alpha! \rangle,
\]
as desired. \( \square \)

**Exercise 3.19.** Verify that if \( u \) is a function in \( C^n(\mathbb{R}^d) \) then, for any \( |\alpha| \leq n \), the function defined by \( U_\alpha(x, x) = 0 \) and

\[
U_\alpha(x, y) = \left| \partial^\alpha u(x) - \sum_{|\beta| \leq n - |\alpha|} \frac{\partial^{\alpha+\beta} u(y)(x-y)^\beta}{\beta!} \right| |x-y|^{\alpha-n}, \quad \forall x \neq y,
\]
is continuous on \( \mathbb{R}^d \times \mathbb{R}^d \). Actually, the converse of is called Whitney’s Extension Theorem, i.e., given continuous functions \( u_\alpha, |\alpha| \leq n \), on a compact set \( K \) of \( \mathbb{R}^d \), define the functions \( U_\alpha(x, y) \) on \( K \times K \) by means of the above expression replacing \( \partial^\alpha u(x) \) with \( u_\alpha(x) \) and \( \partial^{\alpha+\beta} u(y) \) with \( u_{\alpha+\beta}(y) \). If \( U_\alpha \) are continuous on \( K \times K \) then there exits a function \( u \) in \( C^n(\mathbb{R}^d) \) such that \( \partial^\alpha u = u_\alpha \) and

\[
\sum_{|\alpha| \leq n} \sup_K |\partial^\alpha u| \leq C \left[ \sum_{|\alpha| \leq n} \sup_{K \times K} |U_\alpha| + \sum_{|\alpha| \leq n} \sup_K |u_\alpha| \right],
\]
for some constant \( C \) depending only on \( K \), e.g., see Hörmander [68, Section 2.3, pp. 44–52].

**Proof.** A first key fact is the following: if \( K \) is a compact set then there exists a smooth partition of the unity \( \sum_i \chi_i = 1 \) on \( \mathbb{R}^d \setminus K \) such that no point is in the support of infinity many functions \( \chi_i \), the diameter of the support of \( \chi_i \) is at most twice the distance to \( K \), and \( |\partial^\alpha \chi_i(x)| \leq C_\alpha (d_K(x))^{-|\alpha|} \), for every \( x \) in \( \mathbb{R}^d \setminus K \), where \( d_K(x) \) is the distance form the point \( x \) to \( K \). This is based on the co-called cutoff functions, e.g., see Hörmander [68, Section 1.4, pp. 25–32].

The fact that the functions \( U_\alpha \) are continuous is only a re-statement of Taylor formula for a continuously differentiable function of order \( n - |\alpha| \). The converse is the interesting part, i.e., Whitney’s Extension Theorem. \( \square \)

**Exercise 3.20.** By means of the inequality (3.9), prove the inclusions \( S \subset \mathcal{D}_{L^p} \subset \mathcal{D}_{L^q} \subset \hat{B} \), for any \( 1 \leq p \leq q < \infty \) as well as the density of \( \mathcal{D} \) in any of those spaces.

**Proof.** Recall that, for \( 1 \leq p < \infty \)

\[
\mathcal{D}_{L^p} = \{ \varphi \in C^\infty(\mathbb{R}^d) : \partial^\alpha \varphi \in L^p(\mathbb{R}^d), \; \forall \alpha \},
\]
and \( \hat{B} \) is the space of functions \( \varphi \) in \( C^\infty(\mathbb{R}^d) \) such that for any multi-index \( \alpha \) we have

\[
\partial^\alpha \varphi \in L^\infty(\mathbb{R}^d) \quad \text{and} \quad \sup_{|x| \geq r} |\partial^\alpha \varphi(x)| \to 0 \quad \text{as} \quad r \to \infty.
\]

Also, a function \( \varphi \) belongs to \( S \) if and only if \( \varphi \) is in \( C^\infty(\mathbb{R}^d) \) and for any multi-index \( \alpha \) and any \( n \geq 0 \) we have

\[
\sup_{x \in \mathbb{R}^d} \{|\partial^\alpha \varphi(x)|(1 + |x|^2)^n\} < \infty.
\]

The estimate (3.9) write as: for any \( a > 0 \) there exists a constant \( C = C_{a,d} \) such that

\[
|\varphi(x)| \leq C \sum_{|\alpha| \leq d} \int_{|x-y| \leq a} |\partial^\alpha \varphi(y)| \, dy, \quad \forall \varphi \in C^d(\mathbb{R}^d).
\]

If \( \varphi \) belongs to \( S \) then

\[
|\partial^\alpha \varphi(x)| \leq \{ |\partial^\alpha \varphi(x)|(1 + |x|^2)^n \} |(1 + |x|^2)^n|
\]

and choosing \( n \) sufficiently large, this implies that \( \varphi \) belongs to \( D_{L^p} \). Because the elements of \( D_{L^p} \) are smooth functions, \( \partial^\alpha \varphi \) is locally bounded for any multi-index \( \alpha \), which is also expressed by the estimate (3.9). Actually, by means of Hölder inequality

\[
\int_{|x-y| \leq a} |\partial^\alpha \varphi(y)| \, dy \leq C \left( \int_{|x-y| \leq a} |\partial^\alpha \varphi(y)|^p \, dy \right)^{1/p},
\]

for some constant \( C \) (depending on \( a, p \) and \( d \)), which implies that \( D_{L^p} \subset D_{L^\infty} \).

Therefore because \( L^p \cap L^\infty \subset L^q \), we obtain \( D_{L^p} \subset D_{L^q} \) for every \( 1 \leq p < q \leq \infty \). Moreover, the same estimate (3.9) implies that \( \varphi(x) \to 0 \) as \( |x| \to \infty \), i.e., \( D_{L^p} \subset \hat{B} \).

It is clear the countable family of seminorm \( \|\partial^\alpha \varphi\|_p \) makes \( D_{L^p} \) (and \( \hat{B} \) with \( p = \infty \)) a Fréchet space, for \( 1 \leq p \leq \infty \).

If an increasing sequence \( \{\varphi_k\} \) in \( D(\mathbb{R}^d) \) satisfies \( \varphi_k \to 1 \) in \( E(\Omega) \) then \( \varphi_k \varphi \to \varphi \) in \( S \) or \( D_{L^p} \) or \( \hat{B} \), depending on where the function \( \varphi \) belong to. Thus, the density is proved. However, note that \( D(\mathbb{R}^d) \) is not dense \( D_{L^\infty} \), which is usually denote by \( B \), without the ‘dot’. For instance, the interested reader may check the book by Schwartz [112, Section VI.8, pp. 199–205], among others.  

\[ \square \]

### (3.3) More Operations and Localization

#### (3.3.1) Product of Distributions

**Exercise 3.21.** Complete the previous statements: show that (1) \( D(\Omega_1) \otimes D(\Omega_2) \) is dense in \( D(\Omega_1 \times \Omega_2) \); (2) \( T_1 \otimes T_2 \) can be uniquely extended to a distribution in \( \Omega_1 \times \Omega_2 \); and (3) the support of the tensor product of two distributions \( T_1 \otimes T_2 \) is the Cartesian product of their support \( \text{supp}(T_1) \times \text{supp}(T_2) \).
Proof. Recall that $D(\Omega_1) \otimes D(\Omega_2)$ is the vector space generated by the tensor product functions $\varphi_1 \otimes \varphi_2$, with $\varphi_i$ in $D(\Omega_i)$, where $\varphi_1 \otimes \varphi_2(x_1, x_2) = \varphi_1(x_1)\varphi_2(x_2)$, and $x_i$ belongs to $\Omega_i \subset \mathbb{R}^{d_i}$, $i = 1, 2$.

Given a function $\varphi$ in $D(\Omega_1 \times \Omega_2)$, Weierstrass’ Approximation Theorem ensures the existence of a sequence $\{p_n(x_i, x_2) : n \geq 1\}$ of polynomials such that $p_n$ (as well as any derivative) converges to $\varphi$, uniformly over any compact subset of $\Omega_1 \times \Omega_2$. Now, choose functions $\alpha_i$ in $D(\Omega_i)$ such that $\alpha_1(x_1)\alpha_2(x_2) = 1$ for any $x = (x_1, x_2)$ belonging to the support of $\varphi$. Under these conditions, the function $\varphi_n(x) = \alpha(x_1)\beta(x_2)p_n(x_1, x_2)$ belongs to the vector space $D(\Omega_1) \otimes D(\Omega_2)$ and the sequence $\{\varphi_n\}$ converges $\varphi$ in the topology of $D(\Omega_1 \times \Omega_2)$, i.e., (1) is true.

The tensor product $T_1 \otimes T_2$ of two distributions $T_i$ in $D(\Omega_i)$, $i = 1, 2$, is initially defined on the vector (tensor) space $D(\Omega_1) \otimes D(\Omega_2)$ as

$$
\langle T_1 \otimes T_2, \varphi_1 \otimes \varphi_2 \rangle = \langle T_1, \varphi_1 \rangle \langle T_2, \varphi_2 \rangle, \quad \forall \varphi_i \in D(\Omega_i), \ i = 1, 2,
$$

and the density shown in (1) proves that the extension must be unique. Moreover, since $x_1 \mapsto \langle T_2, \varphi(x_1, \cdot) \rangle$ belongs to $D(\Omega_1)$, the expression

$$
\langle T_1 \otimes T_2, \varphi \rangle = \langle T_1(T_2, \varphi(x_1, x_2)) \rangle, \quad \forall \varphi \in D(\Omega_1 \times \Omega_2),
$$

is also valid. Furthermore, if $\varphi_n \to \varphi$ in $D(\Omega_1 \times \Omega_2)$ then the projections $\varphi_n(\cdot, x_2)$ and $\varphi_n(x_1, \cdot)$ also converge to $\varphi(\cdot, x_2)$ and $\varphi(x_1, \cdot)$, i.e., the converge $\langle T_2, \varphi_n(x_1, \cdot) \rangle \to \langle T_2, \varphi(x_1, \cdot) \rangle$, is not only pointwise, but also in the topology of $D(\Omega_1)$. Hence,

$$
\langle T_1 \otimes T_2, \varphi_n \rangle \to \langle T_1 \otimes T_2, \varphi \rangle, \quad \text{as} \quad n \to \infty,
$$

proving the continuity of $T_1 \otimes T_2$.

Regarding the support of the tensor product $T_1 \otimes T_2$, it is clear that if $\varphi$ has its support in $(\mathbb{R}^{d_1} \setminus \Omega_1) \times \Omega_2$ or in $\Omega_1 \times (\mathbb{R}^{d_1} \setminus \Omega_2)$ then the product expression implies that $\langle T_1 \otimes T_2, \varphi \rangle = 0$, i.e., $\text{supp}(T_1) \times \text{supp}(T_2)$ contains the support of $T_1 \otimes T_2$. Conversely, if a point $x^0 = (x_1^0, x_2^0)$ belongs to $\text{supp}(T_1) \times \text{supp}(T_2)$ then there exist $\varphi_i$ satisfying $\varphi_i(x_i^0) \neq 0$, and $\langle T_i, \varphi_i \rangle \neq 0$, $i = 1, 2$, which yields

$$
\langle T_1 \otimes T_2, \varphi_1 \otimes \varphi_2 \rangle \neq 0.
$$

Hence, the equality (3) is obtained.

\[\square\]

**Exercise 3.22.** Reconsider the previous question as follows: If $T$ is a distribution in $\Omega_1 \times \Omega_2$ then we can define a continuous linear operator $\mathcal{T} : D(\Omega_2) \to D'(\Omega_1)$ by the formula

$$
\langle \mathcal{T} \psi, \varphi \rangle = \langle T, \varphi \otimes \psi \rangle, \quad \forall \varphi \in D(\Omega_1), \ \psi \in D(\Omega_2).
$$

Prove that the application $T \mapsto \mathcal{T}$ is injective and surjective.
Proof. The main argument is the density of the vector (tensor) space $\mathcal{D}(\Omega_1) \otimes \mathcal{D}(\Omega_2)$, considered as a subspace of space of test functions $\mathcal{D}(\Omega_1 \times \Omega_2)$.

Indeed, if $\mathcal{T} = \tilde{T}$ then

$$\langle T, \varphi \otimes \psi \rangle = \langle \tilde{T}, \varphi \otimes \psi \rangle, \quad \forall \varphi \in \mathcal{D}(\Omega_1), \ \psi \in \mathcal{D}(\Omega_2),$$

i.e., $T = \tilde{T}$.

Similarly, if $\mathcal{T}$ is continuous linear operator $\mathcal{T}: \mathcal{D}(\Omega_2) \to \mathcal{D}'(\Omega_1)$ then define

$$\langle T, \varphi \otimes \psi \rangle = \langle \mathcal{T}\psi, \varphi \rangle, \quad \forall \varphi \in \mathcal{D}(\Omega_1), \ \psi \in \mathcal{D}(\Omega_2).$$

Thus $T$ is a linear continuous functional initially define on the vector (tensor) space $\mathcal{D}(\Omega_1) \otimes \mathcal{D}(\Omega_2)$, which is considered as a subspace of $\mathcal{D}(\Omega_1 \times \Omega_2)$. Hence, by density, it can be extended to a distribution on $\Omega_1 \times \Omega_2$.

Exercise 3.23. Let $x = (x', x_d)$ a point in $\mathbb{R}^d$, with $x'$ in $\mathbb{R}^{d-1}$, and $\mathbb{R}_+^d = \mathbb{R}^{d-1} \times [0, \infty)$. If $f$ belongs to $C^\infty(\mathbb{R}_+^d)$ we denote its zero-extension to the whole $\mathbb{R}^d$ by $\overline{f}$, i.e., $\overline{f}(x', x_d) = f(x', x_d)$ if $x_d \geq 0$, and $\overline{f}(x', x_d) = 0$ if $x_d < 0$. Consider $\overline{f}$ as a distribution on $\mathbb{R}^d$ and prove that its first derivative in the normal direction $x_d$, is given by the formula

$$\partial_d \overline{f} = \partial_d f + J, \quad \langle J, \varphi \rangle = \int_{\mathbb{R}_+^{d-1}} f(x', 0) \varphi(x', 0) \, dx', \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d).$$

Moreover, by means of the Dirac function, give a formula for the $n$-derivative in the normal direction $x_d$, $\partial_d^n \overline{\varphi}$, in term of a tensor product of distributions. Furthermore, obtain a similar formula in general, for any derivative $\partial^\alpha \overline{\varphi}$ for any multi-index $\alpha$.

Proof. Indeed, by definition,

$$\langle \partial_d \overline{f}, \varphi \rangle = -\langle f, \partial_d \varphi \rangle = -\int_{\mathbb{R}_+^d} f(x', x_d) \partial_d \varphi(x', x_d) \, dx' \, dx_d,$$

and by means of the integration by parts, this is equal to

$$\int_{\mathbb{R}_+^{d-1}} \varphi(x') \left[ f(x', 0) \varphi(x', 0) + \int_0^\infty \partial_d f(x', x_d) \varphi(x', x_d) \, dx_d \right],$$

which prove the first part.

Using the Dirac function $\delta$ in the variable $x_d$, i.e., as a distribution

$$\langle \delta, \varphi \rangle = \varphi(0), \quad \forall \varphi \in \mathcal{D}(\mathbb{R}),$$

the relation just proved $\partial_d \overline{f} = \partial_d f + J$ becomes $\partial_d \overline{f} = \partial_d f + 1 \otimes \delta$, where number 1 is understood as (or identified with) the distribution 1 on $\mathbb{R}^{d-1}$, i.e.,

$$\langle 1, \varphi \rangle = \int_{\mathbb{R}_+^{d-1}} \varphi(x') \, dx', \quad \forall \varphi \in \mathcal{D}(\mathbb{R}_+^{d-1}),$$

[November 11, 2016]
and $\otimes$ is the tensor product of distributions. Certainly, an iteration yields

$$
\partial^n_d f = \partial^n_0 f + \sum_{k=1}^n 1 \otimes \delta^{(k-1)}, \quad \forall n = 0, 1, \ldots,
$$

where $\delta^{(k)}$ is the derivative of order $k$, with $k = 0$ meaning the initial Dirac distribution $\delta$.

For a multi-index $\alpha = (\alpha', \alpha_d)$ the formula becomes

$$
\partial^\alpha f = \partial^\alpha_0 f + \sum_{k=1}^{\alpha_d} 1 \otimes \delta^{(k-1)},
$$

since $\partial^\beta f = \partial^\beta_0 f$, for any multi-index $\beta = (\beta_1, \ldots, \beta_{d-1}, 0)$.

\(\square\)

### 3.3.2 Convolution of Distributions

**Exercise 3.24.** Recall the differential operator $\Delta = \sum_{i=1}^d \partial_i^2$. A fundamental distributional solution associated with the iterated Laplacian $\Delta^k$ is a distribution $E = E_{kd}$ on $\mathbb{R}^d$ such that $\Delta^k(E \ast \delta) = \delta$, where $\delta$ is the Dirac delta measure, $\langle \delta, \varphi \rangle = \varphi(0)$. Verify that $E = |x|^{2k-d}(a_{kd} \ln |x| + b_{kd})$ is a fundamental distributional solution associated $\delta^k$ in $\mathbb{R}^d$, where one of the constants $a_{kd}$ or $b_{kd}$ vanishes, namely, if $2k - d < 0$ or $d$ is odd then $a_{kd} = 0$, and otherwise $b_{kd} = 0$. Note that if $2k - d > 0$ then $E$ belongs to $C^{2k-d-1}$ and complete the following argument. First, consider a distribution $T$ with compact support and verify that $T = E \ast (\Delta^k T)$. Next, if $\Delta^k T$ is a distribution of order $n$ (i.e., it belongs to the dual space of $C^n$) with a compact support and $2k - d - 1 \geq n$ then $T$ is the distribution associated to the function $x \mapsto \langle \Delta^k T, E(x - \cdot) \rangle$, and therefore $T$ belongs to $C^n$.

**Proof.** First note that for dimension $d = 1$, a fundamental solution for the $k$-order (ordinary) differential equation $F^{(k)} = \delta$ is the Heaviside function $F_1 = H$, $H(x) = 1$ if $x > 0$ and $H(x) = 0$ otherwise (to which a constant may be added) for $k = 1$. Now for $k = 2$, any indefinite integral of $H$ plus a constant solves $F'' = \delta$, for instant, $F_2(x) = xH(x)$ or $F_2(x) = xH(x) - x/2 = |x|/2$. Thus, $F_k(x) = x^{k-1}H(x)/(k-1)!$ is a fundamental solution for the differential equation of order $k \geq 1$.

On this context, it may be important to consider the so-called singular support of a distribution $T$, i.e., the set of points having no open neighborhood to which the restriction of $T$ is a $C^\infty$ function. Recall that a function $f$ defined in $\mathbb{R}^d \setminus \{0\}$ is called homogeneous of degree $n$ if $f(rx) = r^nf(x)$ for every $r > 0$ and almost every $x$ in $\mathbb{R}^d \setminus \{0\}$. Thus, also related is the concept of homogeneous distributions $T$ of degree $n$, i.e., a distribution in $\mathbb{R}^d \setminus \{0\}$ such that $\langle T, \varphi \rangle = r^n\langle T, \varphi_r \rangle$, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^d \setminus \{0\})$ and $r > 0$, where $\varphi_r(x) = r^d\varphi(rx)$. Hence, our interest is on homogeneous fundamental solutions $E$ with the singular support $\{0\}$. For instance, the reader may be interested in checking the book Hörmander [68] and Schwartz [112].
Now going back to our problem, for order $k = 1$, $E_{12}(x) = \ln(|x|)/(2\pi)$ and $E_{13}(x) = 1/(4\pi|x|)$, solve $\Delta E = \delta$ in $\mathbb{R}^d$, for dimension $d = 2$ and $d = 3$. As an example, note that $E(x) = \exp(-r|x|)/(4\pi|x|)$ solves the elliptic PDE $(r - \Delta)E = \delta$ in $\mathbb{R}^3$. For a dimension $d \geq 3$ the expression $E_{1d}(x) = c_d|x|^{2-d}/(2-d)$, where $1/c_d = 2\pi^{d/2}/\Gamma(d/2)$ is the area of the unit sphere in $\mathbb{R}^d$, solves $(-\Delta)E = \delta$ in $\mathbb{R}^d$, for any dimension $d \geq 3$.

To check this, use polar coordinates to write the Laplacian operator $\Delta$ in dimension 2 as

$$\Delta f = \frac{1}{r}(r \partial_r f) + \frac{1}{r^2} \partial^2_{\theta} f = \frac{1}{r} \partial_r f + \partial^2_{\theta} f,$$

where $x = r \cos \theta$, $y = r \sin \theta$, and $r^2 = x^2 + y^2$, and use spherical coordinates to write the Laplacian operator $\Delta$ in dimension 3 as

$$\Delta f = \frac{1}{\rho^2} \partial_{\rho}(\rho^2 \partial_{\rho} f) + \frac{1}{\rho^2 \sin \theta} \partial_{\theta}(\sin \theta \partial_{\theta} f) + \frac{1}{\rho^2 \sin^2 \theta} \partial^2_{\phi} f,$$

where $x^2 + y^2 + z^2 = \rho^2$, $x^2 + y^2 = r^2$, and $z = r \cos \phi$. Thus dimension $d \geq 2$, in general,

$$\Delta f = \partial^2_{|x|} f + \frac{(d-1)}{|x|} \partial_{|x|} f + \frac{1}{|x|^2} \Delta_{S^{d-1}} f,$$

where $\Delta_{S^{d-1}}$ is the Laplace-Beltrami (or spherical Laplacian) operator on the $(d-1)$-sphere $S^{d-1} = \{x \in \mathbb{R}^d : |x| = 1\}$, and again the radial term can be written as $|x|^{1-d} \partial_{|x|}(|x|^{d-1} \partial_{|x|} f)$. All this means that the Laplacian $\Delta$ becomes $\partial^2_r + \frac{(d-1)}{r} \partial_r$ for homogeneous distributions $T = T(r)$, with $r = |x| > 0$, i.e., the one dimensional Euler ODE $t^2 y'' + (d-1)ty' = 0$, $t > 0$, with characteristic equation $m^2 + (d-2)m = 0$, which has the roots $m = 0$ and $m = 2 - d$, distinct only when $d \geq 3$. Hence, this ODE has $y(t) = c_1 + c_2 t^{2-d}$ (or $c_1 + c_2 \ln t$ if $d = 2$) as the general solution. Therefore, the proposed expression of $E$ yields $E(r) = r^{2-k-d}(a_k \ln r + b_k)$ satisfied $a_k = 0$ for $k = 1$ (i.e., $\Delta$) and $d \geq 3$.

If $k = 2$ then on homogeneous functions $f$ the Laplacian square becomes

$$\Delta^2 f = |x|^{1-d} \partial_{|x|}(|x|^{d-1} \partial_{|x|}(|x|^{1-d} \partial_{|x|}(|x|^{d-1} \partial_{|x|} f)))$$

and the above argument could be developed. However, let us consider only the case $k = 1$, i.e., $E_{12}(x) = \ln(|x|)/(2\pi)$ and $E_{1d}(x) = c_{d-1}|x|^{2-d}/(2-d)$, $d \geq 3$. In both cases, an integration by parts yields

$$\langle \partial_i E, \varphi \rangle = -\langle E, \partial_i \varphi \rangle = \lim_{\varepsilon \to 0} \int_{|x| \geq \varepsilon} E(x) \partial_i \varphi(x) \, dx =$$

$$= \int_{\mathbb{R}^d} E(x) \partial_i \varphi(x) \, dx + \lim_{\varepsilon \to 0} \int_{|x| = \varepsilon} E(x) \varphi(x)(x_i/|x|) \, dx',$$

where the surface integral in $dx'$ is comparable to $\varepsilon$ if $d \geq 3$ and to $\varepsilon \ln \varepsilon$ if $d = 2$, so that it vanishes as $\varepsilon \to 0$. This proves that the distribution $\partial_i E(x)$
can be identified with the locally integrable function $c_d x_i |x|^{-d}$. Next, apply the divergence theorem to the last term of the equality

$$
\langle \Delta E, \varphi \rangle = \langle E, \Delta \varphi \rangle = \lim_{\varepsilon \to 0} \int_{|x| \geq \varepsilon} (E(x) \Delta \varphi(x) - \varphi(x) \Delta E(x)) \, dx
$$

to obtain

$$
\langle \Delta E, \varphi \rangle = \lim_{\varepsilon \to 0} \int_{|x| \geq \varepsilon} \nabla \cdot (E \nabla \varphi - \varphi \nabla E) \, dx = -\lim_{\varepsilon \to 0} \int_{|x| = \varepsilon} (E \nabla \varphi - \varphi \nabla E) \cdot \left( \frac{x}{|x|} \right) \, dx' = \varphi(0),
$$

where the normalization constant $c_d$ is determined.

A similar argument can be used with the expression

$$
E(x,t) = (4\pi t)^{d/2} \exp \left( -\frac{|x|^2}{4t} \right), \quad t > 0,
$$

and $E(x,t) = 0$ for $t < 0$ to show that $E$ is locally integrable in $(x,t)$ belonging to $\mathbb{R}^{d+1}$, that $E$ belongs to $C^\infty(\mathbb{R}^{d+1} \setminus \{0\})$, and that $(\partial_t - \Delta) E = \delta$.

Finally, if $T$ is a distribution with a compact support then $E \ast T$ is defined and

$$
E \ast (\Delta^k T) = (\Delta^k E) \ast T = \delta \ast T = T.
$$

Therefore, if $\Delta^k T$ is a distribution of order $n$ (i.e., it belongs to the dual space of $C^n$) with a compact support and $2k - d - 1 \geq n$ then $T$ is the distribution associated to the function $x \mapsto \langle \Delta^k T, E(x \cdot \cdot) \rangle$, and thus $T$ belongs to $C^n$.

An application of this last assertion is the following: First, if $\Omega$ is an open convex subset of $\mathbb{R}^d$ then an element $T$ in $\mathcal{D}'(\Omega)$ is called convex if $\tau_h T + \tau_{-h} T - 2T \geq 0$ for every $h$ in $\mathbb{R}^d$ with $|h|$ sufficiently smaller [i.e., $\tau_h$ is the translation operator $\langle \tau_h T, \varphi \rangle = \langle T, \tau_{-h} \varphi \rangle$, and $\tau_h \varphi(x) = \varphi(x + h)$, and $\langle T, \tau_{-h} \varphi + \tau_h \varphi - 2 \varphi \rangle \geq 0$, for any $\varphi$ in $\mathcal{D}(\Omega)$ and $|h|$ smaller than the distance form the support of $\varphi$ to the boundary $\partial \Omega$]. Now, to prove that a distribution $T$ in $\Omega$ is convex if and only if the Hessian $D^2 T$ is nonnegative definite, i.e., $\langle h \cdot D^2 T h, \varphi \rangle = \langle T, h \cdot D^2 \varphi h \rangle \geq 0$, for every $h$ in $\mathbb{R}^d$ and any $\varphi \geq 0$ in $\mathcal{D}(\Omega)$. Indeed, referring to Exercises 3.12 and 3.13, the only missing point is to show that in the above statement, the distribution $T$ is necessarily a function. To this purpose, multiply $T$ by a cutting test function to see that $T$ may be assumed to have a compact support. Since $\Delta T$ is a locally finite signed measure, i.e., belongs to $C^0$, the previous argument shows that indeed, the distribution $T$ can be identified to a continuous function, i.e., the only convex distributions are indeed continuous functions.

Exercise 3.25. Verify the correctness of the following examples of convolutions:
a.- Riesz potentials: $R_\alpha$, $0 < \alpha < d$, and for any $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$,
\[
(-\Delta)^{-\alpha/2}\varphi(x) = R_\alpha \ast \varphi(x) = C_{\alpha,d} \int_{\mathbb{R}^d} |x - y|^{-(d-\alpha)} \varphi(y)dy,
\]
where $C_{\alpha,d} = \Gamma((d - \alpha)/2)/[2^\alpha \pi^{d/2}\Gamma(\alpha/2)]$ is a normalizing constant.

b.- Calderon-Zygmund integro-differential operator, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$,
\[
(-\Delta)^{1/2}\varphi(x) = C_d \sum_{i=1}^d \lim_{\varepsilon \to 0} \int_{|x| \geq \varepsilon} (x_i - y_i)|x - y|^{-d-1} \partial_i \varphi(y)dy,
\]
where $C_d = \Gamma((d + 1)/2)\pi^{-(d+1)/2}$ is again a normalizing constant. Note the singular integral and recall that the limit is called the principal value of the integral.

c.- The Newtonian potential for $d \geq 3$ is defined by
\[
(-\Delta)^{-1}\varphi(x) = (N \ast \varphi)(x) = \frac{1}{(d - 2)\omega_d} \int_{\mathbb{R}^d} |x - y|^{2-d} \varphi(y)dy,
\]
where $\omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of the unit sphere. For $d = 2$ ($d = 1$) we use the kernel $(1/2\pi)\ln(|x - y|) (|x|/2)$. *If $\Delta = \partial_1^2 + \cdots + \partial_d^2$ is the usual Laplacian then verify that $\Delta(N \ast \varphi)(x) = 0$ for every $x$ in $\mathbb{R}^d$.

d.- Double layer potential, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^{d-1})$, with $x = (x', x_d)$
\[
N \ast (\varphi(x') \otimes \delta'(x_d)) = \frac{1}{\omega_d} \int_{\mathbb{R}^{d-1}} x_d \varphi(y')(|x' - y'|^2 + x_d^2)^{-d/2}dy'.
\]
*Verify that $u(x', x_d) = 2N \ast (\varphi(x') \otimes \delta'(x_d))$, which is called the Poisson integral formula, yields a solution of the Dirichlet problem $\Delta u = 0$ in $\mathbb{R}^d$ and $u(\cdot, 0) = \varphi$ in $\mathbb{R}^{d-1}$.

e.- Single layer potential, for any $\varphi$ in $\mathcal{D}(\mathbb{R}^{d-1})$, with $x = (x', x_d)$
\[
N \ast (\varphi(x') \otimes \delta(x_d)) = \frac{1}{(d - 2)\omega_d} \int_{\mathbb{R}^{d-1}} \varphi(y')(|x' - y'|^2 + x_d^2)^{1-d/2}dy',
\]
*Verify that the $\partial_d$ of the single layer potential is equal to double layer potential.

Questions marked with * could not be so simple. The reader may want to check the book by Stein [113] for a detail account of Singular Integrals. 

**Proof.** (a) Let us consider Riesz potentials. The kernel is the function $k(x) = C_{\alpha,d}|x|^{-d+\alpha}$ with $x$ in $\mathbb{R}^d$ and $0 < \alpha < d$. Because $\alpha > 0$, the kernel $R_\alpha = k$ is a locally integrable function. Indeed, by spherical coordinates, a direct computation shows that the integral of $k$ is actually equal to 1, which determines the value of the constant $C_{\alpha,d}$, namely,
\[
C_{\alpha,d} \int_{|y-x| \leq a} |x - y|^{-(d-\alpha)}dy = C_{\alpha,d} \int_0^a r^{-d+\alpha}\omega_d r^{d-1}dr = a^\alpha,
\]
and
\[
C_{\alpha,d} \int_{|y| \leq a} |x - y|^{-(d-\alpha)}dy \leq (3a)^\alpha 1_{\{x \in B_{2a}\}} + (|x| - a)^{-d+\alpha}a^d 1_{\{x \not\in B_{2a}\}},
\]
for any \( a > 0 \), where \( \omega_d \) is the surface measure (area) of the unit sphere in \( \mathbb{R}^d \). Thus, this function \( k \) belongs to \( L^1_{\text{loc}}(\mathbb{R}^d) \) and so it can be identified with a distribution in \( \mathbb{R}^d \), i.e., an element of \( \mathcal{D}'(\mathbb{R}^d) \). Therefore, if \( \varphi \) is a bounded function with a compact support in \( \mathbb{R}^d \) then the convolution expression yields a bounded function \( x \mapsto (R_\alpha \ast \varphi)(x) \) which vanishes like \( |x|^{-d+\alpha} \) as \( |x| \to \infty \).

If we take \( \alpha \leq 0 \) then the kernel \( k \) is no more a distribution, since it is not locally integrable. On the other hand, if \( \alpha \geq d \) then the kernel \( k \) is at least a continuous function and the interest in the convolutions properties is limited. The notation \((-\Delta)^{-\alpha/2} \varphi \) come from the fact that the Laplacian \( \Delta = \partial_1^2 + \cdots + \partial_d^2 \) and the Fourier transform \( \mathcal{F} \) (as discussed later, see Chapter 5) for tempered distributions \( \mathcal{S}(\mathbb{R}^d) \) enjoy the relation

\[ \mathcal{F}(\Delta \varphi)(x) = 4\pi^2|x|^2 \mathcal{F}(\varphi)(x), \quad \forall x \in \mathbb{R}^d, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d), \]

and so, with this tool, most of the formal calculations become valid in this sense. In particular, the Fourier transform of the function \( x \mapsto C_{\alpha,d}|x|^{-d+\alpha} \) is \( \xi \mapsto (2\pi|\xi|)^{-\alpha} \).

Recall Young inequality for the convolution, see Proposition B.65,

\[ \|f \ast g\|_r \leq \|f\|_p\|g\|_q, \quad 1 \leq p, q, r \leq \infty, \quad 1/p + 1/q - 1/r = 1, \]

where \( \cdot \| \) denotes the norm in \( L^p(\mathbb{R}^d) \), and now, express the kernel \( k(x) = C(\alpha,d)|x|^{-d+\alpha} \) as

\[ k = k_1 + k_\infty, \quad \text{with} \quad k_1(x) = k(x) \mathbb{1}_{|x| \leq 1}, \quad \forall x \in \mathbb{R}^d, \]

then \( k \ast f = k_1 \ast f + k_\infty \ast f \). Since \( k_1 \) belongs to \( L^1(\mathbb{R}^d) \) the first convolution operator \( f \mapsto k_1 \ast f \) maps \( L^p(\mathbb{R}^d) \) into itself, i.e., \( \|k_1 \ast f\|_p \leq \|k_1\|_1\|f\|_p \), for every \( f \) in \( L^p(\mathbb{R}^d) \). Similarly, it is clear that \( k_\infty \) belongs to \( L^q(\mathbb{R}^d) \), for any \( 1 \leq q < \infty \) such that \((-d+\alpha)q < -d\), which is equivalent to \( \alpha < d(1/p - 1/r) \), with the notation of Young inequality. This, for any \( 1 \leq p < \infty \), we can find \( 1 < r < \infty \) such that \((-d+\alpha)q < -d \) with \( 1/q = 1 - 1/p + 1/r \), i.e., \( \|k_\infty \ast f\|_r \leq \|k_\infty\|_q\|f\|_p \). Therefore, the expression defining the convolution \( (k \ast f)(x) \) is meaningful for almost every \( x \) in \( \mathbb{R}^d \) as an absolutely convergence integral, for any \( f \) in \( L^p(\mathbb{R}^d) \), \( 1 \leq p < \infty \). More effort is needed to show the Hardy-Littlewood-Sobolev inequality, namely, for any \( 1 < p < q < \infty, 1/q = 1/p - \alpha/d \), there exists a constant \( C_{p,q} \) such that

\[ \|k \ast f\|_q \leq C_{p,q}\|f\|_p, \quad \forall f \in L^p(\mathbb{R}^d), \]

For instance, the interested reader may take a look at the classic book by Stein [113, Section V.1, pp. 117–121].

(b) Previously, the Riesz potentials \( (R_\alpha \ast \varphi)(x) \) are defined for any test function \( \varphi \) in \( \mathcal{D}'(\mathbb{R}^d) \) without any further considerations, as a convergent integral for every \( x \) in \( \mathbb{R}^d \). Now, the Calderon-Zygmund integro-differential operator is defined as the singular integral

\[ (-\Delta)^{1/2} \varphi(x) = C_d \sum_{i=1}^d \int_{\mathbb{R}^d} (x_i - y_i)|x - y|^{-d-1}\partial_i \varphi(y)dy, \]
with $C_d = \Gamma((d+1)/2)\pi^{-(d+1)/2}$, which needs some previous analysis. Indeed, even for a test function $\varphi$, the above integral is not absolutely convergent due to the kernel $k(x) = \sum x_i |x|^{-d-1}$, which is not locally integrable, and so, it is not directly interpreted as a distribution. The technique of the principal value can be used to check that

$$(-\Delta)^{1/2} \varphi(x) = C_d \sum_{i=1}^{d} \lim_{\varepsilon \to 0} \int_{|x| \geq \varepsilon} (x_i - y_i)|x - y|^{-d-1} \partial_i \varphi(y) dy =$$

$$= C_d \sum_{i=1}^{d} \int_{\mathbb{R}^d} (x_i - y_i)|x - y|^{-d-1} [\partial_i \varphi(y) - \partial_i \varphi(x)] dy.$$ 

Indeed, because the kernel $k$ satisfies $k(x) = -k(-x)$ (i.e., it is an odd function) and integrable outside its only singularity (i.e., the origin $x = 0$) we have

$$\int_{|x| \geq \varepsilon} (x_i - y_i)|x - y|^{-d-1} \partial_i \varphi(x) dy = 0,$$

and now, the singularity at $y = x$ of the functions

$$y \mapsto (x_i - y_i)|x - y|^{-d-1} [\partial_i \varphi(y) - \partial_i \varphi(x)],$$

for $i = 1, \ldots, d$, are integrable, actually of order $-d + 1$.

The notation $(-\Delta)^{1/2}$ come from the fact that the Laplacian and the Fourier transform mentioned above. The assertion that the Fourier transform of the singular kernel $x \mapsto C_d x_i |x|^{-d-1}$ is the function $\xi \mapsto i \xi_i / |\xi|$, combined with the formula $\mathfrak{F}(\partial_i \varphi)(\xi) = -2\pi i \xi_i \mathfrak{F}(\varphi)(\xi)$, shows that the Fourier transform (symbol) of the Calderon-Zygmund integro-differential operator is indeed $(2\pi |\xi|)$.

It is clear that to study these operators, the derivative $\partial_i \varphi$ can be replaced by $\varphi$ (which is referred to as the Riesz transform), and then consider a composition with the partial differential operator $\partial_i$. Thus, the simplest example of this type of operators is the Hilbert transform

$$H\varphi(x) = \frac{1}{\pi} \lim_{\varepsilon \to 0} \int_{|y| \geq \varepsilon} \frac{\varphi(x - y)}{y} dy = \frac{1}{\pi} \int_{\mathbb{R}} \frac{\varphi(x - y) - \varphi(x)}{y} dy.$$ 

The Hilbert and Riesz transforms, and in general convolution (or non) with Calderon-Zygmund kernels define bounded operators in from $L^p$ into itself, any $1 < p < \infty$. Certainly, this generalizes the Fourier transform technique used to handle the case $L^2$. The interested reader may take a look at the classic book by Stein [113, Section III.1, pp. 54–60] and references therein.

(c) For $d \geq 3$, the Newtonian potential has a kernel $N(x) = k(x) = |x|^2 - d/(d-2)\omega_d$, with $\omega_d = 2\pi^{d/2}/\Gamma(d/2)$; while for $d = 2$ (or $d = 1$) the kernel is given by $(1/2\pi) \ln(|x - y|)$ (or $|x|/2$). Thus, there are specific properties that may change with the dimension $d$. Usually dimension $d \geq 3$ is assumed and the cases $d = 2$ or $d = 1$ are studied separately or left for the reader. In any case, the kernel is locally integrable and everything has a clear meaning in the sense
of distributions. It appears very naturally when trying to solve the PDE (partial differential equation) \(-\Delta u = f\) in \(\mathbb{R}^d\) (Poisson equation), i.e., find a function \(u\) for a given function \(f\). By means of the Fourier transform, this PDE becomes

\[4\pi^2|\xi|^2\hat{\mathcal{F}}(u)(\xi) = \hat{\mathcal{F}}(f)(\xi),\]

which is solved by writing \(u = \hat{\mathcal{F}}^{-1}((2\pi|\xi|)^{-2}) * f\), i.e., a solution \(u\) is the convolution with the kernel \(k\) which is the Fourier inverse of the function \(\xi \mapsto (2\pi|\xi|)^{-2}\), and computations show that this kernel is as above. Based on the fact that the Fourier transform is a homeomorphism on the space of tempered distributions, all this argument is valid either on the space of test function rapidly decreasing \(\mathcal{S}(\mathbb{R}^d)\) or on its dual space \(\mathcal{S}'(\mathbb{R}^d)\), under some extra conditions on \(f\), i.e., for any \(f\) in either \(\mathcal{S}(\mathbb{R}^d)\) or \(\mathcal{S}'(\mathbb{R}^d)\) such that the function \(\xi \mapsto (2\pi|\xi|)^{-2}\mathcal{F}(f)(\xi)\) belongs to either \(\mathcal{S}(\mathbb{R}^d)\) or \(\mathcal{S}'(\mathbb{R}^d)\), the Newtonian potential \(u = (-\Delta)^{-1}f\) belongs to either \(\mathcal{S}(\mathbb{R}^d)\) or \(\mathcal{S}'(\mathbb{R}^d)\) and it satisfies Poisson equation \(-\Delta u = f\). Note that for \(f = 0\), the solutions of the equation \(-\Delta u = 0\) are the so-called harmonic functions, so that the uniqueness questions regarding the Poisson equation is not a simple task.

The Newtonian kernel \(N(x)\) is a smooth function outside of the origin, and a routine calculation shows that \(\Delta N(x) = 0\) for any \(x \neq 0\). Thus, the difficulty is when taking a second derivative inside the integral sign, e.g. can we calculate the limit

\[
\lim_{h_i \to 0} \int_{\mathbb{R}^d} [\partial_i N(x + h - y) - \partial_i N(x - y)] f(y) dy = \int_{\mathbb{R}^d} N(y) \partial_i^2 f(x - y) dy.
\]

Note that the derivative kernel \(\partial_i N\) is locally integrable, but the second derivative kernel \(\partial_i^2 N\) is singular in the sense that it cannot be regarded as a distribution corresponding to a locally integrable function. In general, the singularity in the integral defining the convolution with the kernel \(\partial_i^2 N\) is non locally integrable and should be removed by some ‘cancellation’ property, i.e.,

\[
\int_{|x-y| \geq \epsilon} \partial_{ij}^2 N(x - y) dy = 0, \quad \forall x \in \mathbb{R}^d,
\]

and if \(f\) is a H"older continuous function, namely, \(|f(x) - f(y)| \leq C|x - y|^{\alpha}\), for every \(x, y\) in \(\mathbb{R}^d\) and for some \(\alpha\) in \((0, 1)\), then

\[
\lim_{\epsilon \to 0} \int_{|x-y| \geq \epsilon} \partial_{ij}^2 N(x - y) f(y) dy = \int_{\mathbb{R}^d} \partial_{ij}^2 N(x - y) (f(y) - f(x)) dy,
\]

which is a local property as \(|x - y|\) is small, i.e., \(\partial_{ij}^2 N(x - y)\) has a non-locally integrable singularity of order \(|x - y|^{-d}\) and with a H"older continuous function \(f\), this becomes an integrable singularity of order \(|x - y|^{-d+\alpha}\). Therefore, to actually show that if \(f\) is H"older continuous (and some boundedness or integrability assumptions as \(|y| \to \infty\)) then the Newtonian potential \(u = N * f\) is twice-continuously differentiable with H"older continuous second derivatives and \(-\Delta u = f\) in \(\mathbb{R}^d\).

Moreover, if \(f\) is a continuous function with a compact support then the
expression

\[(N \ast f)(x) = \frac{1}{(d-2)\omega_d} \int_{\mathbb{R}^d} |x - y|^{2-d} f(y) dy = \]

\[= \frac{1}{(d-2)\omega_d} \int_{\mathbb{R}^d} |y|^{2-d} f(x - y) dy,\]

is a smooth function for every \(x\) outside the support of \(f\). Actually, the equality

\[(N \ast f)(x) - N(x) \int_{\mathbb{R}^d} f(y) dy = \]

\[= \frac{1}{(d-2)\omega_d} \int_{\mathbb{R}^d} \left[ |x - y|^{2-d} - |x|^{2-d} \right] f(y) dy,\]

and more work (and assumptions, in particular, that \(f\) is radial symmetric) shows that the integral on the right-hand side vanishes, i.e., \(N \ast f\) is equal to a constant times the Newtonian kernel \(N\), which is interpreted in physics as follows: the potential energy of a small mass outside a much larger spherically symmetric mass distribution is the same as if all of the mass of the larger object were concentrated at its center. The interested reader may take a look at the books DiBenedetto [30], Evans [42], Hellwig [67], among other textbooks.

(D) In the double layer potential, the notation \(x = (x', x_d)\) means that the actual kernel \(k(x)\) is written as \(k(x' - y', x_d) = x_d(|x' - y'|^2 + x_d^2)^{-d/2}/\omega_d\) and considered only for \(x_d \neq 0\), usually for \(x_d > 0\). This is not a singular integral, but it becomes singular when \(x_n \to 0\), either from the right or from the left. If the starting point is the Newtonian kernel \(N(x) = |x|^{2-d}/(d-2)\omega_d\) then \(k(x) = \partial_d N(x', x_d)\), which justify the notation \((k \ast \varphi)(x', x_d) = N \ast (\varphi(x') \otimes \delta'(x_d))\) in the sense of distributions, i.e., if \(\varphi\) is a test function in \(\mathbb{R}^d\) then \(\delta_d = \delta'(x_d)\varphi\) means the distribution \(\varphi \mapsto \varphi(x', x_d)\) and \(\delta'_d = \delta'(x_d)\varphi\) means the distribution \(\varphi \mapsto \partial_d \varphi(x', x_d)\), both as acting only on the variable \(x_d\). Therefore for a distribution \(\Phi\) on \(\mathbb{R}^{d-1}\) with compact support, and identifying \(N\) with a distribution given through a locally integrable kernel \(N\), this reduces to \((\delta'_d N) \ast \Phi = N \ast (\Phi \otimes \delta'_d)\), i.e., due to the convolution, the action of \(\delta'_d\) on \(\varphi\) is regarded as acting on \(N\) to produce \(\delta'_d N = \partial_d N = k\), the kernel used in the double layer potential convolution.

If \(u\) denotes the double layer potential corresponding to a continuous function \(f\) (with some boundedness or integrability assumptions as \(|y| \to \infty\) in \(\mathbb{R}^{d-1}\)), then \(u(x', x_d)\) is a smooth function for any \(x'\) in \(\mathbb{R}^{d-1}\) and satisfies \(-\Delta' u(x', x_d) = 0\), for any \(x_d \neq 0\), where \(\Delta'\) is the Laplacian operator in the variable \(x'\) of \(\mathbb{R}^{d-1}\), moreover \(2u(x', x_d) \to \pm f(x')\) as \(x_d \to 0\), depending on the side whether \(x_d > 0\) or \(x_d < 0\). To show this fact, calculations begin with a Hölder continuous function \(f\) and the property

\[\frac{2}{\omega_d} \int_{\mathbb{R}^{d-1}} x_d(|x' - y'|^2 + x_d^2)^{-d/2} dy' = 1,\]

which can be proved by means of the Fourier transform. Among other books, the interested reader may check Stein [113, Section III.2, pp. 60–68] and references therein.
Actually the rule is as follows: if $\Omega$ is a domain in $\mathbb{R}^d$ with a smooth boundary $\partial \Omega$ and for a point $a$ is a point in the boundary $\partial \Omega$ and $n(y)$ denotes the exterior unit normal vector on the boundary at $y$, then the limits

$$u^+(a',a_d) = \lim_{x \to a, x \in \Omega} \int_{\partial \Omega} (n(y) \cdot \nabla N(x-y)) f(y')d\sigma(y)$$

and $u^-(a',a_d)$ (when $x \to a$ is kept outside of $\Omega$) exist and the following jump relation

$$u^\pm(a',a_d) = \pm \frac{1}{2} f(a) + \int_{\partial \Omega} (n(y) \cdot \nabla N(a-y)) f(y')d\sigma(y)$$

holds true, where $d\sigma(y)$ is the area measure on the $(d-1)$-dimensional manifold $\partial \Omega$ and nabla $\nabla$ is the gradient operator. The interested reader may take a look at the textbook by DiBenedetto [30, Chapter III, pp. 116–160].

(e) It is clear that the kernel for the single layer potential is $k = N$, the Newtonian potential integrated in $\mathbb{R}^{d-1}$. Thus the convolution expression defining the single layer potential $N \ast \varphi(x') \otimes \delta(x_d)$ is a better singularity then the one defining the double layer potential, but we are interested in the derivative of this single layer potential, i.e., similarly, the justification for the notation for the single layer potential is the equality $(\delta_d N) \ast \Phi = N \ast (\Phi \otimes \delta_d)$, and with this notation it should be clear that $\partial_d (\delta_d N) = \delta_d' N$. If $u$ denotes the single layer potential corresponding to a continuous function $f$ (with some boundedness or integrability assumptions as $|y| \to \infty$) in $\mathbb{R}^{d-1}$, then $u(x',x_d)$ is a smooth function for any $x'$ in $\mathbb{R}^{d-1}$ and satisfies $-\Delta' u(x',x_d) = 0$, for any $x_d \neq 0$, where $\Delta'$ is the Laplacian operator in the variable $x'$ of $\mathbb{R}^{d-1}$, moreover $2\partial_d u(x',x_d) \to \pm f(x')$ as $x_d \to 0$, depending on the side whether $x_d > 0$ or $x_d < 0$.

Certainly, all this is related to the fundamental solutions and the Green functions corresponding to elliptic PDE in the half-space $\mathbb{R}^{d+}_+ = \{(x',x_d) \in \mathbb{R}^d : x_d > 0\}$. Key tools in this analysis are first the Green identity: if $u, v$ are in $C^2(\bar{\Omega})$ for a smooth open domain $\Omega \subset \mathbb{R}^d$ then the divergence theorem yields

$$\int_{\Omega} v(x) \Delta u(x) dx = - \int_{\Omega} \nabla v(x) \cdot \nabla u(x) dx + \int_{\partial \Omega} v(x)n(x) \cdot \nabla u(x) d\sigma(x),$$

$$\int_{\Omega} [v(x) \Delta u(x) - u(x) \Delta v(x)] dx =$$

$$= \int_{\partial \Omega} [v(x)n(x) \cdot \nabla u(x) - u(x)n(x) \cdot \nabla v(x)] d\sigma(x),$$

and by approximation, these equalities hold true for any $u, v$ in $C^2(\Omega) \cap C^1(\bar{\Omega})$ such that $\Delta u$ and $\Delta v$ are in $L^\infty(\Omega)$. Next, Stokes identity: if $u$ in $C^2(\Omega) \cap C^1(\bar{\Omega})$ such that $\Delta u$ is in $L^1(\Omega)$ then for any $x$ in $\Omega$, we have

$$u(x) = \frac{1}{(d-2)\omega_d} \int_{\Omega} |x-y|^{2-d} \Delta u(x) dx +$$

$$+ \frac{1}{(d-2)\omega_d} \int_{\partial \Omega} [|x-y|^{2-d}n(y) \cdot \nabla u(y) - u(y)n(y) \cdot \nabla |x-y|^{2-d}] d\sigma(y),$$
whenever $d \geq 3$, while the Newtonian kernel $|x - y|^{2-d}/(d-2)\omega_d$ becomes $\ln|x - y|/2\pi$ when $d = 2$. Note that this is an 'implicit' representation formula for smooth function, but not more work is necessary to construct the actual Green function. For instance, the interested reader may check the textbook by DiBenedetto [30, Chapter II, pp. 55–115] regarding the Laplace equation.

Exercise 3.26. Let $\{T_k\}$ be a sequence of distribution converging to 0 in $\mathcal{D}'(\mathbb{R}^d)$, and let $S$ be another distribution. Prove if either (a) $S$ has a compact support or (b) the supports of $\{T_k\}$ are contained in a fixed compact set, then $S \ast T_k \to 0$ in $\mathcal{D}'(\mathbb{R}^d)$.

Proof. Because the strong and the weak* topologies on the dual space of distribution $\mathcal{D}'(\mathbb{R}^d)$ coincides, the fact that $T_k \to 0$ translates into $\langle T_k, \varphi \rangle \to 0$, for every $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$. By definition

$$\langle T_k \ast S, \varphi \rangle = \langle T_k, x, \langle S_y, \varphi(x + y) \rangle \rangle = \langle S_x, \langle T_k, y, \varphi(x + y) \rangle \rangle = \langle T_k \otimes S, \varphi^\oplus \rangle,$$

where $\varphi^\oplus(x, y) = \varphi(x + y)$ and $\varphi$ is any element in $\mathcal{D}(\mathbb{R}^d)$.

Thus, if $S$ has a compact support then there exists a test function $\varphi_0$ such that $S = \varphi_0 S$ and

$$\langle T_k \ast S, \varphi \rangle = \langle T_k, x, \langle S_y, \varphi_0(y)\varphi(x + y) \rangle \rangle,$$

which implies that the function $\psi_\varphi : x \mapsto \langle S_y, \varphi_0(y)\varphi(x + y) \rangle$ is a test function, for every $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$, i.e., $\psi_\varphi$ belongs to $\mathcal{D}(\mathbb{R}^d)$. Hence, $T_k \to 0$ yields $\langle T, \psi_\varphi \rangle \to 0$ as desired.

Similarly, if the supports of $\{T_k\}$ are contained in a fixed compact set then there exists a test function $\varphi_0$ such that $T_k = \varphi_0 T_k$ and

$$\langle T_k \ast S, \varphi \rangle = \langle T_k, x, \varphi_0(x)\langle S_y, \varphi(x + y) \rangle \rangle,$$

which implies again that the function $\psi_\varphi : x \mapsto \varphi_0(x)\langle S_y, \varphi(x + y) \rangle$ is a test function, for every $\varphi$ in $\mathcal{D}(\mathbb{R}^d)$, and the conclusion follows as above.

(3.3.3) Local Structure

(3.3.4) Recap on Inductive Limits
Exercises - Chapter (4) 
Introduction to Sobolev Spaces

(4.1) Density and Extension
(4.1.1) Regularity on the Domain
(4.1.2) Lipschitz Transformation

(4.2) Imbedding and Compactness
(4.2.1) Some Typical Estimates
(4.2.2) General Imbedding

(4.3) Traces on the Boundary
(4.3.1) In Half-space
(4.3.2) In a Smooth Domain
(4.3.3) Spaces on the Boundary

(4.4) Fractional Order Spaces
(4.4.1) Discussion and Definition
(4.4.2) Basic Properties
Exercises - Chapter (5)
Basic Fourier Transform

(5.1) Smooth Functions

(5.2) Tempered Distributions

Exercise 5.1. First, prove that Fourier transform commute with the tensor product, i.e., if the $T_1$ and $T_2$ are two tempered distributions then $\hat{T_1} \otimes S_2 = \hat{T_1} \otimes \hat{T_2}$. Secondly, for the convolution of two distributions, prove that if $T$ belongs to $S'(\mathbb{R}^d)$ and $S$ belongs to $\mathcal{E}'(\mathbb{R}^d)$ then the convolution $T * S$ belongs to $S'(\mathbb{R}^d)$, the Fourier transform $\hat{S}$ is identified with a smooth function, namely, $\xi \mapsto \langle S, e^{-2\pi i \xi \cdot} \rangle$, and $\hat{T} * \hat{S} = \hat{T S}$.

Proof. The fact that Fourier transform commute with the tensor product is based on the property of the exponential and the translation, and the fact that the tensor product $T_1 \otimes T_2$ is initially defined on the product space $\mathcal{D}(\mathbb{R}^{d_1}) \otimes \mathcal{D}(\mathbb{R}^{d_2})$ and then uniquely extended to the space $\mathcal{D}(\mathbb{R}^{d_1+d_2})$, i.e., the equalities

$$\langle \hat{T_1} \otimes \hat{T_2}, \varphi_1 \otimes \varphi_2 \rangle = \langle T_1 \otimes T_2, \hat{\varphi_1} \otimes \hat{\varphi_2} \rangle$$

and $\hat{\varphi_1} \otimes \hat{\varphi_2} = \hat{\varphi_1} \otimes \hat{\varphi_2}$, yield the desired conclusion.

Recall that the convolution of two distributions $T$ and $S$ is defined

$$\langle T * S, \varphi \rangle = \langle T_x, \langle S_y, \varphi(x + y) \rangle \rangle = \langle S_x, \langle T_y, \varphi(x + y) \rangle \rangle$$

for any $\varphi$ in element in $\mathcal{D}(\mathbb{R}^d)$, as long as this make sense, i.e., one of the distribution should have a compact support. Thus, under the assumptions of the second part of this of this exercise, we must establish that for every element $\varphi$ in $S(\mathbb{R}^d)$, the function $x \mapsto \langle S, \varphi(x + \cdot) \rangle$ belongs to $S(\mathbb{R}^d)$.

Since $S$ is distribution with a compact support, $S$ is an element of $\mathcal{E}'(\mathbb{R}^d)$ and therefore, there is an index $k$, a compact set $K \subset \mathbb{R}^d$, and a constant $C$ such that

$$|\langle S, \varphi \rangle| \leq C \sup_{y \in K, |\alpha| \leq k} \{ |\partial^\alpha \varphi(y)| \}, \quad \forall \varphi \in \mathcal{D}(\mathbb{R}^d),$$
which means that
\[
(1 + |x|^2)^{n/2} |\langle S, \varphi(x + \cdot) \rangle| \leq C \sup_{y \in K} \{ (1 + |x|^2)^{n/2} (1 + |x + y|^2)^{-n/2} \} \times \\
\times \sup_{y \in K, |\alpha| \leq k} \{ (1 + |x + y|^2)^{n/2} |\partial^\alpha \varphi(x + y)| \},
\]
for every element \(\varphi\) in \(\mathcal{D}(\mathbb{R}^d)\). Moreover, because
\[
\frac{(1 + |x|^2)^{n/2}}{(1 + |x + y|^2)^{n/2}} \leq 2^{n/2} (1 + |y|^2)^{n/2}, \quad \forall x, y \in \mathbb{R}^d,
\]
we deduce
\[
\sup_{x \in \mathbb{R}^d, |\alpha| \leq n} \{ (1 + |x|^2)^{n/2} |\langle S, \partial^\alpha \varphi(x + \cdot) \rangle| \} \leq \\
\leq C_{k,K} \sup_{x \in \mathbb{R}^d, |\alpha| \leq k+n} \{ (1 + |x|^2)^{n/2} |\partial^\alpha \varphi(x)| \},
\]
with
\[
C_{k,K} = C \sup_{y \in K} \{ 2^{n/2} (1 + |y|^2)^{n/2} \},
\]
and for every \(\varphi\) in \(\mathcal{D}(\mathbb{R}^d)\). This effectively establishes that the function \(x \mapsto \langle S, \varphi(x + \cdot) \rangle\) belong to \(\mathcal{S}(\mathbb{R}^d)\), \(\varphi\) in \(\mathcal{D}(\mathbb{R}^d)\), i.e., the convolution \(T \ast S\) belongs to \(\mathcal{S}'(\mathbb{R}^d)\).

Now, since \(S\) belongs to \(\mathcal{E}'(\mathbb{R}^d)\) and the function \((\xi, x) \mapsto e^{-2\pi i (\xi \cdot x)}\) is smooth, by definition of the Fourier transform
\[
\langle \hat{S}, \varphi \rangle = \langle S, \hat{\varphi} \rangle = \int_{\mathbb{R}^d} \varphi(x) \langle S_\xi, e^{-2\pi i (\xi \cdot x)} \rangle dx,
\]
i.e., the distribution \(\hat{S}\) is identified with the function \(\xi \mapsto \langle S, e^{-2\pi i \xi \cdot \cdot} \rangle\). Next, the equalities
\[
\langle \hat{T \ast S}, \varphi \rangle = \langle T \ast S, \hat{\varphi} \rangle = \langle T_x, \langle S_y, \hat{\varphi}(x + y) \rangle \rangle
\]
and
\[
\langle S_y, \hat{\varphi}(x + y) \rangle = \int_{\mathbb{R}^d} \varphi(y) \langle S_\xi, e^{-2\pi i (\xi + x) \cdot y} \rangle dy = \\
= \int_{\mathbb{R}^d} \varphi(y) e^{-2\pi i (x \cdot y)} \langle S_\xi, e^{-2\pi i (\xi \cdot y)} \rangle dy,
\]
yield
\[
\langle T_x, \langle S_y, \hat{\varphi}(x + y) \rangle \rangle = \int_{\mathbb{R}^d} \varphi(y) \langle T_x, e^{-2\pi i (x \cdot y)} \rangle \langle S_\xi, e^{-2\pi i (\xi \cdot y)} \rangle dy,
\]
which is clearly valid for any \( \varphi \) in \( \mathcal{D}(\mathbb{R}^d) \), under the assumption that the tempered distribution \( T \) has a compact support, i.e., we have proved the equality \( \widehat{T \ast S} = \hat{T} \hat{S} \) as long as both distributions \( T \) and \( S \) are in \( \mathcal{E}'(\mathbb{R}^d) \).

Finally, if \( k_\varepsilon \) is a smooth kernel with compact support then \( T \ast k_\varepsilon \rightarrow T \) in \( \mathcal{S}'(\mathbb{R}^d) \) and the continuity of the Fourier transform allows us to complete proof.

(5.3) Integrable Functions

(5.4) Periodic Functions

(5.5) Fourier Multiplier
Exercises - Chapter (6) 
Besov and Sobolev Spaces

(6.1) Hilbert Sobolev Spaces
(6.1.1) In the Whole Space
(6.1.2) In Continuous Domains
(6.1.3) Trace Operator

(6.2) Riesz and Bessel Potentials
(6.2.1) Initial Discussion
(6.2.2) Bessel Kernel and Potentials
(6.2.3) Fundamental Solutions

(6.3) Besov and Sobolev Relations
(6.3.1) Besov spaces
(6.3.2) Bessel Potential Spaces
(6.3.3) Traces on Besov Spaces
Appendix B

Measure and Integration

This ‘background’ chapter is not an integral part of this book, it is included only by convenience for the reader. In the following chapters, it is assumed that the reader is somehow familiar with measure theory and integral, e.g., by having taken a first course in real analysis with some topics taken from a typical textbook such as, DiBenedetto [31], Dshalalow [35], Dudley [36], Folland [44], Jones [70], Pollard [102], Royden [108], Stein and Shakarchi [115], Taylor [124], Wheeden and Zygmund [133], or many others.

Therefore, this Chapter is a summary of the essential material discussed in the book [89]. Our objective is to describe three independent ways for constructing measures, namely, the outer approach (or Caratheodory’s arguments) in Section 2, the inner approach (or compact technique) in Section 3, and finally, the Lebesgue integral is presented in Section 4. Then, in Section 4, a quick discussion (mainly definition) of the integral and some complements in Section 5.

B.1 Classes of Sets

Let $\Omega$ be a nonempty set and $2^\Omega$ be the parts of $\Omega$, i.e., set of all subsets of $\Omega$. Clearly, if $\Omega$ has $n$ elements then $2^\Omega$ has $2^n$ elements, but our interest is when $\Omega$ has an infinite number of elements, for instance if $\Omega$ is countable infinite (i.e., it is in a one-to-one relation with the positive integers) then $2^\Omega$ has the cardinality of the continuum. A class (collection or family or system) of sets is a subset of $2^\Omega$, that by convenience, we assume it contains the empty set. Note that $\emptyset \subset \Omega$ and $\emptyset, \Omega \in 2^\Omega$. Typical operations between two elements $A$ and $B$ in $2^\Omega$ are the intersection $A \cap B$, the union $A \cup B$, the difference $A \setminus B$ and the complement $A^c = \Omega \setminus A$. The union and the intersection can be extended to any number of sets, e.g., if $A_i \in 2^\Omega$ for $i$ in some sets of indexes $I$ then we have $\bigcup_{i \in I} A_i$ and $\bigcap_{i \in I} A_i$. Sometimes, to simplify notation we write $A + B$ or $\sum_{i \in I} A_i$ (for disjoint unions) to express the fact that $A + B = A \cup B$ with $A \cap B = \emptyset$ or $\sum_{i \in I} A_i = \bigcup_{i \in I} A_i$ with $A_i \cap A_j = \emptyset$ if $i \neq j$.
B.1.1 First Properties

Definition B.1. Given classes $\mathcal{P}, \mathcal{L}, \mathcal{R}$ and $\mathcal{A}$ of subsets of $\Omega$, each containing $\emptyset$, we say that

- $\mathcal{P}$ is a $\pi$-class if $A, B \in \mathcal{P}$ implies $A \cap B \in \mathcal{P}$,
- $\mathcal{L}$ is a $\ell$-class (or additive class) if (a) $A, B \in \mathcal{L}$ with $A \cap B = \emptyset$ implies $A \cup B \in \mathcal{L}$ and (b) $A, B \in \mathcal{L}$ with $A \subset B$ implies $B \setminus A \in \mathcal{L}$,
- $\mathcal{R}$ is a ring if $A, B \in \mathcal{R}$ implies (a) $A \setminus B \in \mathcal{R}$ and (b) $A \cup B \in \mathcal{R}$,
- $\mathcal{A}$ is algebra if (a) $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$ and (b) $A, B \in \mathcal{A}$ implies $A \cup B \in \mathcal{A}$.

Finally, a $\pi$-class $\mathcal{S}$ is called (1) a semi-ring if $A, B \in \mathcal{S}$ with $A \subset B$ implies $B \setminus A = \sum_{i=1}^n C_i$ with $C_i \in \mathcal{S}$, (2) a semi-algebra if $A \in \mathcal{S}$ implies $A^c = \sum_{i=1}^n C_i$ with $C_i \in \mathcal{S}$, and (3) a lattice if $A, B \in \mathcal{S}$ implies $A \cup B \in \mathcal{S}$. \hfill $\Box$

From the definitions, it is clear that any interception of $\pi$-classes, $\ell$-classes, lattices, rings or algebras is again a $\pi$-class, an $\ell$-class, a lattice, a ring or an algebra. Therefore, given any subset $\mathcal{G}$ of $2^\Omega$ we may define the $\pi$-class, $\ell$-class, lattice, ring or algebra generated by $\mathcal{G}$, e.g., the algebra $\mathcal{A}(\mathcal{G})$ generated by $\mathcal{G}$ is indeed the intersection of all algebras containing $\mathcal{G}$.

A semi-ring of interest for us is the class $S$ of intervals of the form $(a, b]$, with $a, b$ real numbers. For instance, an carefully discussion on semi-rings can be found in Dudley [36, Section 3.2, pp. 94–101]. Another point to remember is that if $\mathcal{K} = \ell(\mathcal{P})$ is the smallest $\ell$-class containing a given $\pi$-class $\mathcal{P}$ then $\mathcal{K}$ is also the ring generated by $\mathcal{P}$. Moreover, if $\Omega \in \mathcal{K}$ then $\mathcal{K}$ is the smallest algebra containing $\mathcal{P}$.

Definition B.2. A $\sigma$-algebra (or $\sigma$-field) $\mathcal{A}$ is a class containing $\emptyset$ which is stable under the (formation of) complements and countable unions, i.e., (a) if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$ and (b) if $A_i \in \mathcal{A}$, $i = 1, 2, \ldots$ then $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$. Similarly, a $\sigma$-ring $\mathcal{A}$ is a non-empty class stable under differences and countable unions, i.e., (c) if $A, B \in \mathcal{R}$ then $A \setminus B \in \mathcal{R}$ and (b) as above. \hfill $\Box$

The classes mostly used are the $\sigma$-algebras. Any ring (or algebra) with a finite number of elements is a $\sigma$-ring (or $\sigma$-algebra). It is relatively simple to generate (and identify) a $\pi$-class, an $\ell$-class, a ring or an algebra, this is not the same for the $\sigma$-classes, because transfinite induction is involved.

The concept of monotone classes is used to clarify the distinction between algebras (or rings) and $\sigma$-algebras (or $\sigma$-rings). A monotone class (of subset of $\Omega$) is a subset $\mathcal{M}$ of $2^\Omega$ stable under countable monotone unions and intersections, i.e., (a) $A_i \in \mathcal{M}$, $A_i \subset A_{i+1}$, $i = 1, 2, \ldots$ then $\bigcup_{i=1}^\infty A_i \in \mathcal{M}$ and (b) $A_i \in \mathcal{M}$, $A_i \supset A_{i+1}$, $i = 1, 2, \ldots$ then $\bigcap_{i=1}^\infty A_i \in \mathcal{M}$.

Proposition B.3. Let $\mathcal{K} = \mathcal{M}(\mathcal{R})$ be the smallest monotone class containing a given ring $\mathcal{R}$. Then $\mathcal{K}$ is also the $\sigma$-ring generated by $\mathcal{R}$. Moreover, if $\Omega \in \mathcal{K}$ then $\mathcal{K}$ is the smallest $\sigma$-algebra containing $\mathcal{R}$.

Proof. For every $K \in \mathcal{K}$ define the class of sets $\Phi_K = \{A \in \mathcal{K} : A \setminus K, K \setminus A, A \cup K \in \mathcal{K}\}$. Clearly, (a) $A \in \Phi_K$ if and only if $K \in \Phi_A$, and (b) the relations
(\cup_i A_i) \setminus K = \cup_i (A_i \setminus K), (\cap_i A_i) \setminus K = \cap_i (A_i \setminus K), K \setminus (\cup_i A_i) = \cap_i (K \setminus A_i), K \setminus (\cap_i A_i) = \cup_i (K \setminus A_i), (\cup_i A_i) \cup K = \cup_i (A_i \cup K) \text{ and } (\cap_i A_i) \cup K = \cap_i (A_i \cup K) \text{ imply that } \Phi_K \text{ is a monotone class for any fixed } K.

In particular, if } K = R \in \mathcal{R} \text{ then } A \in \mathcal{R} \implies A \in \Phi_R. \text{ Thus } \mathcal{R} \subset \Phi_R \text{ and because } K \text{ is the smallest monotone class containing } \mathcal{R} \text{ we have } \mathcal{K} \subset \Phi_R. \text{ This is } K \in \mathcal{K} \implies K \in \Phi_R, \text{ or equivalently } R \in \Phi_K, \text{ for every } R \in \mathcal{R}. \text{ Hence } \mathcal{R} \subset \Phi_K \text{ and again, because } K \text{ is the smallest monotone class containing } \mathcal{R} \text{ we have } \mathcal{K} \subset \Phi_K, \text{ but this time for every } K \in \mathcal{K}. \text{ This proves that for any } A,K \in \mathcal{K} \text{ we have } A \setminus K, K \setminus A, A \cup K \in \mathcal{K}, \text{ i.e., } \mathcal{K} \text{ is a ring.}

Finally, we conclude by noting that a } \sigma\text{-ring is a } \sigma\text{-algebra if and only if it contains } \Omega. \blacksquare

- Remark B.4. From Propositions B.3 follows that if } \mathcal{P} \text{ is a } \pi\text{-class then } \mathcal{M}(\mathcal{P}) \text{ is the smallest } \sigma\text{-algebra containing } \mathcal{P}. \blacksquare

The notation } \sigma(\mathcal{K}) \text{ means the smallest } \sigma\text{-algebra containing a given class } \mathcal{K}, \text{ or the } \sigma\text{-algebra generated by } \mathcal{K}. \text{ It is clear that if } \mathcal{K} \text{ is finite then } \sigma(\mathcal{K}) \text{ is also finite.}

Let } \mathcal{R} \text{ be the union of all } \sigma\text{-rings } \mathcal{R}(\mathcal{E}_c) \text{ generated by a countable subclass } \mathcal{E}_c \text{ of a given class } \mathcal{E} \text{ in } 2^\Omega \text{ containing the empty set. Since } \mathcal{R} \text{ is indeed a } \sigma\text{-ring, we have } \mathcal{R} = \mathcal{R}(\mathcal{E}), \text{ the } \sigma\text{-ring generated by the whole class } \mathcal{E}. \text{ Thus, for a given } A \text{ in } \mathcal{R}(\mathcal{E}) \text{ there exists a countable subclass } \mathcal{E}_c \text{ (depending on } A) \text{ such that } A \text{ belongs to } \mathcal{R}(\mathcal{E}_c). \text{ If we can keep the same countable subclass for every set } A \text{ then the } \sigma\text{-ring } \mathcal{R}(\mathcal{E}) \text{ is called separable. Moreover, we say that a } \sigma\text{-algebra } \mathcal{F} \text{ is countable generated or separable if there exists a countable class } \mathcal{K} \text{ such that } \mathcal{F} = \sigma(\mathcal{K}).

Frequently, the previous Propositions are combined in the so-called argument of monotone class as follows. A } \lambda\text{-class (or } \sigma\text{-additive class) is a subset } \mathcal{D} \text{ of } 2^\Omega \text{ stable under the formation of countable monotone unions, monotone differences and it contains } \Omega, \text{ i.e., (a) } A_i \in \mathcal{D}, A_i \subset A_{i+1}, i = 1,2,\ldots \text{ then } \bigcup_{i=1}^\infty A_i \in \mathcal{D}, \text{ (b) if } A,B \in \mathcal{D} \text{ with } A \subset B \text{ then } B \setminus A \in \mathcal{D} \text{ and (c) } \Omega \in \mathcal{D}. \text{ From the equality } A + B = (A^c \setminus B)^c \text{ we deduce that a } \lambda\text{-class is stable under the formation of countable disjoint unions.}

Proposition B.5 (monotone argument). Let } \mathcal{D} \text{ be a } \lambda\text{-class and } \mathcal{P} \text{ be a } \pi\text{-class. Then } \mathcal{D} \text{ is a } \sigma\text{-algebra if and only if } \mathcal{D} \text{ is also stable under finite intersections. Moreover, if } \mathcal{P} \subset \mathcal{D} \text{ then } \sigma(\mathcal{P}) \subset \mathcal{D}.

Proof. To verify the first part, because } \Omega \in \mathcal{D} \text{ we remark that } \mathcal{D} \text{ is stable under complement. Next, we note that any countable union } A = \cup_i A_i \text{ can be expressed as } A = \cup_i B_i, \text{ with } B_n = \bigcup_{i=1}^n A_i = (\bigcap_{i=1}^n A_i^c)^c \text{ which satisfy } B_i \subset B_{i+1}, \text{ for every } i. \text{ So, if } \mathcal{D} \text{ is also a } \pi\text{-class then } B_n \text{ belongs to } \mathcal{D}, \text{ and } \mathcal{D} \text{ is stable under countable union, i.e., a } \mathcal{D} \text{ is indeed a } \sigma\text{-algebra.}

For the second part, if } \lambda(\mathcal{P}) \text{ denotes the smallest } \lambda\text{-class containing } \mathcal{P} \text{ then, for every } E \in \lambda(\mathcal{P}), \text{ define the class of sets } \Phi_E = \{A \in \lambda(\mathcal{P}) : A \cap E \in \lambda(\mathcal{P})\}. \text{ An argument similar to those of Propositions B.3 proves that } \Phi_E = \lambda(\mathcal{P}) \text{ is a } \sigma\text{-algebra, i.e., } \lambda(\mathcal{P}) = \sigma(\mathcal{P}). \text{ Therefore } \sigma(\mathcal{P}) \subset \mathcal{D}. \blacksquare
• **Remark B.6.** Recall the distributive formula: Given a family \( \{F_{i,j} : i \in I_j, j \in J\} \) of subsets of \( \Omega \), verify the distributive formula

\[
\bigcup_{j \in J} \bigcap_{i \in I_j} F_{i,j} = \bigcap_{k \in K} \bigcup_{j \in J} F_{k,j} \quad \text{and} \quad \bigcap_{j \in J} \bigcup_{i \in I_j} F_{i,j} = \bigcup_{k \in K} \bigcap_{j \in J} F_{k,j},
\]

where \( K = \prod_{j \in J} I_j \), i.e., \( \{i_j : j \in J\} \), and \( F_{k,j} = F_{i,j} \). It is clear that if \( J \) is finite and each \( I_j \) is countable then \( K \) is a countable set, however, if for instance, \( I_j = \{0, 1\} \) for every \( j \) in an infinite set of indexes \( J \) then \( K = \{0, 1\}^J \) is not a countable set of indexes.

• **Remark B.7.** Recalling that \( \sum \) denotes disjoint union of sets, for a given semi-ring (or ring or algebra) \( \mathcal{E} \subset 2^X \), we consider the class \( \mathcal{F} = \{ \sum_{k=1}^\infty E_k : E_k \in \mathcal{E} \} \) of subsets in \( 2^X \). First, if \( A = \bigcup_{i=1}^\infty A_i \) with \( A_i \subset A_{i+1} \) and \( A_i \) in \( \mathcal{E} \) then \( A = \bigcup_{i=1}^\infty B_i \), with \( B_1 = A_1, B_2 = A_2 \setminus B_1, \ldots, B_n = A_n \setminus B_{n-1} \), and because \( \mathcal{E} \) is a semi-ring, each \( B_i \) is a finite disjoint union of elements in \( \mathcal{E} \), i.e., \( A \) is a countable disjoint union of sets in \( \mathcal{E} \), which proves that \( \mathcal{F} = \{ \sum_{k=1}^\infty E_k : E_k \in \mathcal{E} \} \). Now, if \( F_j = \bigcup_i E_{i,j} \) then \( F = \bigcup_j F_j = \bigcup_{i,j} E_{i,j} \) is a countable union of sets in \( \mathcal{E} \) and therefore, \( F \) belongs to \( \mathcal{F} \), i.e., \( \mathcal{F} \) is stable under countable unions. However, the distributive formula of Remark B.6 can only be used to show that \( \mathcal{F} \) is stable under finite intersections, since \( \bigcap_{j=1}^\infty \bigcap_{i=1}^\infty E_{i,j} = \sum_{k \in K} \bigcap_{j=1}^\infty E_{k,j} \), where \( K = I^J \), \( E_{k,j} = E_{i,j} \), but \( K \) is not a countable set of indexes. Thus, if \( A = \bigcup_{i=1}^\infty A_i \) and \( B = \bigcup_{j=1}^\infty B_j \) with \( A_i \) and \( B_j \) in \( \mathcal{E} \) then \( A \setminus B = \bigcup_{i=1}^\infty \bigcap_{j=1}^\infty (A_i \setminus B_j) \), where each difference \( A_i \setminus B_j \) is a finite disjoint union of elements in \( \mathcal{E} \), but \( \bigcap_{j=1}^\infty (A_i \setminus B_j) \) is not necessarily in \( \mathcal{F} \), i.e., \( \mathcal{F} \) may not be stable neither under countable intersection nor under differences. Therefore \( \mathcal{F} \), which is stable under countable unions and finite intersections, may be strictly smaller than the \( \sigma \)-ring generated by \( \mathcal{E} \).

Given a non empty set \( \Omega \) (called space) with a \( \sigma \)-algebra \( \mathcal{F} \), the couple \( (\Omega, \mathcal{F}) \) is called a measurable space and each element in \( \mathcal{F} \) is called a measurable set. Moreover, the measurable space is said to be separable if \( \mathcal{F} \) is countable generated, i.e., if there exists a countable class \( K \) such that \( \sigma(K) = \mathcal{F} \). An atom of a \( \sigma \)-algebra \( \mathcal{F} \) is a set \( F \) in \( \mathcal{F} \) such that any other subset \( E \subset F \) with \( E \) in \( \mathcal{F} \) is either the empty set, \( E = \emptyset \), or the whole \( F, E = F \). Thus, a \( \sigma \)-algebra separates points (i.e., for any \( x \neq y \) in \( \Omega \) there exist two sets \( A \) and \( B \) in \( \mathcal{F} \) such that \( x \in A, y \in B \) and \( A \cap B = \emptyset \) if and only if the only atoms of \( \mathcal{F} \) are the singletons (i.e., sets of just one point, \( \{x\} \) in \( \mathcal{F} \)).

### 1.1.2 Topology Included

Recall that a topology on \( \Omega \) is a class \( \mathcal{T} \subset 2^\Omega \) with the following properties: (1) \( \emptyset, \Omega \in \mathcal{T} \), (2) if \( U, V \in \mathcal{T} \) then \( U \cap V \in \mathcal{T} \) (stable under finite intersections) and (3) if \( U_i \in \mathcal{T} \) for an arbitrary set of indexes \( i \in I \) then \( \bigcup_{i \in I} U_i \in \mathcal{T} \) (stable under arbitrary unions). Every element of \( \mathcal{T} \) is called open and the complement of an open set is called closed. A basis for a topology \( \mathcal{T} \) is a class \( \mathcal{B} \subset \mathcal{T} \) such that for any point \( x \in \Omega \) and any open set \( U \) containing \( x \) there exists an element...
$V \in \mathcal{T}$ such that $x \in V \subset U$, i.e., any open set can be written as a union of open sets in $\mathcal{T}$. Clearly, if $\mathcal{T}$ is known then also $\mathcal{F}$ is known as the smallest class satisfying (1), (2), (3) and containing $\mathcal{T}$. Moreover, a class $\mathcal{F}$ containing $\emptyset$ and such that $\bigcup \{ \mathcal{V} \in \mathcal{F} \}$ = $\Omega$ is called a sub-basis and the smallest class satisfying (1), (2), (3) and containing $\mathcal{F}$ is called the weakest topology generated by $\mathcal{F}$ (note that the class constructed as finite intersections of elements in a sub-basis forms a basis). A space $\Omega$ with a topology $\mathcal{T}$ having a countable basis $\mathcal{F}$ is commonly used. If the topology $\mathcal{T}$ is induced by a metric then the existence of a countable basis $\mathcal{F}$ is obtained by assuming that the space $\Omega$ is separable, i.e., there exists a countable dense set.

Given a family of spaces $\Omega_i$ with a topology $\mathcal{T}_i$ for $i$ in some arbitrary family of indexes $I$, the product topology $\mathcal{T} = \prod_{i \in I} \mathcal{T}_i$ (also denoted by $\otimes_i \mathcal{T}_i$) on the Cartesian product space $\Omega = \prod_{i \in I} \Omega_i$ is generated by the basis $\mathcal{F}$ of open cylindrical sets, i.e., sets of the form $\prod_{i \in I} U_i$, with $U_i \in \mathcal{T}_i$ and $U_i = \Omega_i$ except for a finite number of indexes $i$. Certainly, it suffice to take $U_i$ in some basis $\mathcal{T}_i$ to get a basis $\mathcal{F}$, and therefore, if the index $I$ is countable and each space $\Omega_i$ has a countable basis then so does the (countable!) product space $\Omega$. Recall Tychonoff’s Theorem which states that any (Cartesian) product of compact (Hausdorff) topological spaces is again a compact (Hausdorff) topological space with the product topology.

On a topological space $(\Omega, \mathcal{T})$ we define the Borel $\sigma$-algebra $\mathcal{B} = \mathcal{B}(\Omega)$ as the $\sigma$-algebra generated by the topology $\mathcal{T}$. If the space $\Omega$ has a countable basis $\mathcal{F}$, then $\mathcal{B}$ is also generated by $\mathcal{F}$. However, if the topological space does not have a countable basis then we may have open sets which are not necessarily in the $\sigma$-algebra generated by a basis. The couple $(\Omega, \mathcal{B})$ is called a Borel space, and any element of $\mathcal{B}$ is called a Borel set.

Similar to the product topology, if $\{(\Omega_i, \mathcal{F}_i) : i \in I\}$ is a family of measurable spaces then the product $\sigma$-algebra on the product space $\Omega = \prod_{i \in I} \Omega_i$ is the $\sigma$-algebra $\mathcal{F} = \prod_{i \in I} \mathcal{F}_i$ (also denoted by $\otimes_i \mathcal{F}_i$) generated by all sets of form $\prod_{i \in I} A_i$, where $A_i \in \mathcal{F}_i$, $i \in I$ and $A_i = \Omega_i$, $i \notin J$ with $J \subset I$, finite. However, only if $I$ is finite or countable, we can ensure that the product $\sigma$-algebra $\prod_{i \in I} \mathcal{F}_i$ is also generated by all sets of form $\prod_{i \in I} A_i$, where $A_i \in \mathcal{F}_i$, $i \in I$. For a finite number of factors, we write $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n$. Sometimes, the notation $\mathcal{F} = \otimes_{i \in I} \mathcal{F}_i$ is used (i.e., with $\otimes$ replacing $\times$), to distinguish from the Cartesian product which is rarely used for classes of sets).

**Proposition B.8.** Let $\Omega$ be a topological space such that every open set is a countable union of closed sets. Then the Borel $\sigma$-algebra $\mathcal{B}(\Omega)$ is the smallest class stable under countable unions and intersections which contains all closed sets.

**Proof.** Let $\mathcal{B}_0$ be the smallest class stable under countable unions and intersections which contains all closed sets. Since every open set is a countable union of closed sets, we deduce that $\mathcal{B}_0$ contains all open sets. Define $\Phi = \{ B \in \mathcal{B}(\Omega) : B \in \mathcal{B}_0 \}$. It is clear that $\Phi$ is stable under countable unions and intersections, and it contains all closed sets. The minimal character of $\mathcal{B}_0$ implies
that $\Phi = B_0$, and because $\Phi$ is also stable under the formation of complement, we deduce that $B_0$ is a $\sigma$-algebra, i.e., $B_0 = B(\Omega)$. \hfill \Box$

For instance, if $d$ is a metric on $\Omega$ then any closed $C$ can be written as $\{x \in \Omega : d(x, C) < 1/n\}$, i.e., as a countable intersection of open sets, and by taking complement, any open set can be written as a countable union of closed sets. In this case, Proposition B.8 proves that the Borel $\sigma$-algebra $B(\Omega)$ is the smallest class stable under countable unions and intersections which contains all closed (or open) sets.

On a topological space $\Omega$ we define the classes $F_\sigma$ (and $G_\delta$) as the countable unions of closed (intersections of open) sets. Thus, any countable unions of sets in $F_\sigma$ is again in $F_\sigma$ and any countable intersections of sets in $G_\delta$ is again in $G_\delta$. In particular, if the singletons (sets of only one point) are closed then any countable set is an $F_\sigma$. However, we can show (with a so-called category argument) that the set of rational numbers is not a $G_\delta$ in $\mathbb{R} = \Omega$.

In $\mathbb{R}$, we may argue directly that any open interval is a countable (disjoint) union of open intervals, and any open interval $(a, b)$ can be written as the countable union $\bigcup_{n=1}^{\infty} [a + 1/n, b - 1/n]$ of closed sets, an in particular, we show that any open set (in $\mathbb{R}$) is an $F_\sigma$. In a metric space $(\Omega, d)$, a closed set $F$ can be written as $F = \bigcap_{n=1}^{\infty} F_n$, with $F_n = \{x \in \Omega : d(x, F) < 1/n\}$, which proves that any closed set is a $G_\delta$, and by taking the complement, any open set in a metric space is a $F_\sigma$.

Certainly, we can iterate these definitions to get the classes $F_{\sigma\delta}$ (and $G_{\delta\sigma}$) as countable intersections (unions) of sets in $F_\sigma$ ($G_\delta$), and further, $F_{\sigma\delta\sigma}$, $G_{\delta\sigma\delta}$, etc. Any of these classes are family of Borel sets, but in general, not every Borel set belongs necessarily to one of those classes.

### B.1.3 Measurable Functions

Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be two measurable spaces. A function $f : \Omega \to E$ is called measurable if $f^{-1}(B) = \{\omega : f(\omega) \in B\}$ belong to $\mathcal{F}$ for any $B$ in $\mathcal{E}$. Since $\mathcal{A} = \{A \in \mathcal{E} : f^{-1}(A) \in \mathcal{F}\}$ is a $\sigma$-algebra, we deduce that if $\mathcal{E} = \sigma(\mathcal{K})$ then for $f$ to be measurable it suffices that $K \in \mathcal{K}$ implies $f^{-1}(K) \in \mathcal{F}$.

The particular case where $E$ is a Lusin space (i.e., $E$ is homeomorphic to a Borel subset of a compact metrizable space or equivalently, $E$ is a one-to-one continuous image of a Polish space) and $\mathcal{E} = B(E)$ (its Borel $\sigma$-algebra) is sufficiently general to accommodate all situations of interest, for instance a complete metrizable space or a Borel set $E \subset \mathbb{R}^d$ is a typical example. Recalling that a function $f$ is continuous if and only if $f^{-1}(U)$ is open in $\Omega$ for any open set $U$ in $E$, we obtain that any continuous function is measurable (whenever any open set in $\Omega$ belongs to $\mathcal{F}$).

Suppose that $E$ is a topological space where every open set $O$ can be written as countable union of open sets with closure contained in $O$, i.e., $O = \bigcup_i O_i$, for a sequence of open set $\{O_i : i = 1, 2, \ldots\}$ satisfying $\overline{O_i} \subset O$, e.g., a metric space. If $\{f_n\}$ is a sequence of measurable functions with values in $E$ such that $f_n(x) \to f(x)$ for every $x \in \Omega$, then $f$ is also measurable. Indeed, it suffices
to write \( f^{-1}(O) = \bigcup_i \bigcap_k \bigcap_{n \geq k} f^{-1}(F_i) \) for any open set \( O = \bigcup_i O_i \) in \( E \), with \( O_i \) open sets and \( F_i = \overline{O_i} \) closed sets. Similarly, if \( d \) is a complete metric on \( E \) and \( C \) is the subset of \( \Omega \) where \( \{f_n(x)\} \) converges then the expression \( C = \bigcap_k \bigcup_m \bigcap_{n \geq m} \{ x \in \Omega : d(f_n(x), f_m(x)) < 1/k \} \) shows that \( C \) is a measurable set, and therefore, the limit \( f(x) = \lim_n f_n(x) \) for \( x \in C \) can be extended to a measurable function defined on the whole \( \Omega \).

The composition of measurable functions is clearly measurable and so, in particular, if \( E \) is a vector (algebra) topological space (i.e., the sum, scalar multiplication and product are continuous operations and \( E \) is endowed with its Borel \( \sigma \)-algebra) then \( cf + g \) (\( fg \)) is measurable for any scalar \( c \) and any measurable functions \( f \) and \( g \). Thus, the class of measurable functions \( \mathcal{L}^0 = \mathcal{L}^0(\Omega, \mathcal{F}; E) \) is a vector space if \( E \) is so. Note that if \( E \) is not separable then distinct notions of measurability may appear and a deeper analysis is necessary.

Sometimes we use measurable functions with values in either \( (-\infty, +\infty) \) or \( [-\infty, +\infty) \) or \( \mathbb{R} = [-\infty, +\infty] \), i.e., extended real values. In this case, we have to specify how to handle the symbols \(-\infty\) and \(+\infty\). The corresponding Borel \( \sigma \)-algebra is obtained by simply adding the extra symbols, e.g., \( B \in \mathcal{B}(\mathbb{R}) \) if and only if \( B \cap \mathbb{R} \in \mathcal{B}(\mathbb{R}) \). For a sequence \( \{f_n\} \) of functions taking values in \( [-\infty, +\infty) \) or \( \mathbb{R} \), the function \( f(x) = \inf_n f_n(x) \) is measurable if each \( f_n \) is so, and similarly with the sup, \( \liminf \) and \( \limsup \). Essentially, all countable operation preserves measurability. However, if \( \{f_i : i \in I\} \) is a family of real-valued measurable functions with an infinite non countable index \( I \) such that \( f_i \leq C \) for some constant \( C \) and for every \( i \in I \) then the real-valued function \( f(x) = \sup \{f_i(x) : i \in I\} \) is not necessarily measurable.

Let \( \{f_i : i \in I\} \) be a family of functions \( f_i : \Omega \to E_i \), where \( (E_i, \mathcal{E}_i) \) is measurable space. We denote by \( \sigma(\{f_i : i \in I\}) \) the \( \sigma \)-algebra generated by the class of sets \( \{f_i^{-1}(B_i) : B_i \in \mathcal{E}_i, i \in I\} \), which is the smallest \( \sigma \)-algebra in \( \Omega \) such that every \( f_i \) is measurable. It is clear that if \( f_i \) is \( \mathcal{F} \)-measurable for each \( i \) then \( \sigma(\{f_i : i \in I\}) \subset \mathcal{F} \). Moreover, if \( \mathcal{F}_i = \sigma(f_i) \) is the \( \sigma \)-algebra generated by \( \{f_i^{-1}(B_i) : B_i \in \mathcal{E}_i\} \), a fixed \( f_i \), then \( \sigma(\bigcup_{i \in I} \mathcal{F}_i) = \sigma(f_i : i \in I) \), where \( \sigma(\bigcup_{i \in I} \mathcal{F}_i) \) is the smallest \( \sigma \)-algebra containing every \( \mathcal{F}_i \). A typical example of this construction is the case where \( \Omega = \prod_{i \in I} \Omega_i \), \( E_i = \Omega_i \), \( \mathcal{E} = \mathcal{F} \) and \( f_i = \pi_i \) are the projections, i.e., \( \pi_i : \Omega \to \Omega_i \), \( \pi_i(\omega) = \omega_i \) for any \( \omega = (\omega_i : i \in I) \).

It is easy to verify that the product \( \sigma \)-algebra \( \mathcal{F} = \prod_{i \in I} \mathcal{F}_i \) as defined in the previous section satisfies \( \mathcal{F} = \sigma(\{\pi_i : i \in I\}) \).

It should be clear that our main example is the Borel line \( (\mathbb{R}, \mathcal{B}(\mathbb{R})) \). The space \( \mathbb{R} \) has a nice topology, in particular, it is a complete separable metric space (i.e., a Polish space). Even if the \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}) \) has the cardinality of the continuum, and so it is much smaller than \( 2^{\mathbb{R}} \), most of the sets (in \( \mathbb{R} \)) we encounter are Borel set and most functions are Borel function. This is to say \( \mathcal{B}(\mathbb{R}) \) has a reasonable size with respect to the space \( \mathbb{R} \).

Certainly, the same remarks apply to \( (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \), which can be also viewed as a product space. We this in mind, let us consider the following examples:

(1) The space \( \mathbb{R}^\infty \) or \( \mathbb{R}^\mathbb{N} \) with \( \mathbb{N} = \{1,2,\ldots\} \) is the space of all sequences of real numbers. For instance, the family of (open) cylinder sets of the form
The space \( R^\sigma \) topology is the family of open cylinder sets of the form \( C^\sigma \) coordinate and we get a Borel set. In this case, again, the size of the Borel \( i \) for any \( \text{size of} \) is also generated by all sets of the form \( B^\sigma \). The Borel \( \sigma \)-algebra \( B(\mathbb{R}^\infty) \) is equal to the product \( \sigma \)-algebra \( B^\infty(\mathbb{R}) \), which is also generated by all sets of the form \( B_1 \times \cdots \times B_n \times \cdots \), with \( B_i \in B(\mathbb{R}) \) for any \( i = 1, 2, \ldots \), i.e., we can impose any kind of Borel constraint on each coordinate and we get a Borel set. In this case, again, the size of the Borel \( \sigma \)-algebra \( B^\infty(\mathbb{R}) \) is a reasonable, with respect to the space \( \mathbb{R}^\infty \).

(2) The space \( \mathbb{R}^T \), where \( T \) is an infinite uncountable set (e.g., an interval in \( \mathbb{R} \)), is the space of all real-valued function defined on \( T \). A basis for the product topology is the family of open cylinder sets of the form \( C = \prod_{t \in T} (a_t, b_t) \), with \( a_t < b_t \) for every \( t \in T \), and \( -a_t = b_t = +\infty \) for every \( t \) except a finite number. Again, a sequence in \( \mathbb{R}^T \) (i.e., a sequence of real-valued functions defined on \( T \)) converges if each coordinate converges, i.e., the pointwise convergence, and the topology becomes complicate. Moreover, the Borel \( \sigma \)-algebra \( B(\mathbb{R}^T) \) is not equal to the product \( \sigma \)-algebra \( B^T(\mathbb{R}) \), which is generated by open (or Borel) cylinder sets as described in general early. It is not hard to show that elements in \( B^T(\mathbb{R}) \) have the form \( B \times \mathbb{R}^{T \setminus S} \) (disregarding the order of indexes), with \( B \in B^S(\mathbb{R}) \) for some countable subset \( S \subset T \). This means that (product) Borel sets in \( \mathbb{R}^T \) allow only a countable number of Borel constraint on each coordinate, and for instance, we deduce the unpleasant conclusion that the set of continuous functions is not a Borel set. In this sense, the product Borel \( \sigma \)-algebra \( B^T(\mathbb{R}) \) is (too) small relative to the (too) big space \( \mathbb{R}^T \). The Borel \( \sigma \)-algebra \( B(\mathbb{R}^T) \) is larger, but attached to the pointwise convergence, which create other serious complications.

(3) Usually, for a domain we mean a connected set which is the closure of its interior. Thus, the set \( C(D) \) of all real-valued bounded continuous functions defined on a domain \( D \subset \mathbb{R}^d \) with the uniform convergence is a good example of a Banach (complete normed) space. If \( D \) is bounded then the space \( C(D) \) is separable, a very important property for the construction of the Borel \( \sigma \)-algebra. When \( D \) is unbounded (e.g., \( D = \mathbb{R}^d \)), we prefer to use the locally uniform convergence. Actually, this is also the case when considering continuous functions on an open set \( \mathcal{O} \subset \mathbb{R}^d \). As discussed later Chapters, this space has a nice topology, referred to as locally convex topological vector spaces. For instance, as in the case of \( \mathbb{R}^\infty \), we may choose a increasing sequence \( \{K_i\} \) of compact subsets of \( \mathbb{R}^d \) such that either \( \mathbb{R}^d = \bigcup_i K_i \) or \( \mathcal{O} = \bigcup_i K_i \) to define a metric

\[
d(f, g) = \sum_{i=1}^{\infty} \frac{2^{-i} \|f - g\|_i}{1 + \|f - g\|_i}, \quad \forall f, g \in C(\mathbb{R}^d) \text{ or } C(\mathcal{O}),
\]
where $\| \cdot \|_n$ is the supremum norm within $K_n$. Thus, $C(\mathbb{R}^d)$ and $C(O)$ are complete separable metric spaces under the locally uniform convergence topology. Actually, if $X$ is a locally compact space, we may consider the space $C_0(X)$ of real-valued continuous functions with compact support. Then, besides the Borel $\sigma$-algebra on $X$, we may consider smaller $\sigma$-algebras which make all functions in $C_0(X)$ measurable, i.e., the Baire $\sigma$-algebra on $X$. If $X$ is a locally compact Polish space (e.g., $X$ is a domain or an open set in $\mathbb{R}^d$) both $\sigma$-algebra coincide, but this is not the case in general.

(4) As mentioned early, a Polish space $\Omega$ is a complete separable metric space, i.e., the topology of the space $\Omega$ is also generated by a basis composed of open balls $B(x,r) = \{y \in \Omega : d(y,x) < r\}$ for $x$ in some countable dense set of $\Omega$, $r$ any positive rational and there exists some metric (equivalent to $d$) which makes $\Omega$ complete. For instance, $\Omega$ is a closed subset of $\mathbb{R}$ with the induced or relative topology; or a more elaborated example $\Omega$ is the space of real-valued continuous functions defined on some locally compact space with the locally uniform convergence. Since, for any closed set $F \subset \Omega$ the function $d(x,F) = \inf\{d(x,y) : y \in F\}$ is continuous, we deduce that the Borel $\sigma$-algebra $B(\Omega)$ in a Polish space is the smallest $\sigma$-algebra for which every real-valued continuous function defined on $\Omega$ is measurable. This fact is not granted for a general topological space and give rise to the Baire $\sigma$-algebra. To study stochastic processes we use the so-called canonical sample space $D([0,\infty])$ of cad-lag real-valued functions, i.e., functions $\omega : [0,\infty] \to \mathbb{R}$ which are right-continuous with left limit. This space is a Polish space with a suitable topology and metric.

### B.1.4 Some Tools

Let us consider real-valued measurable functions defined on $(\Omega, \mathcal{F})$. A measurable function $\varphi$ taking a finite number of values is called a simple function, i.e., if $\varphi$ takes only the values $a_1, \ldots, a_n$ then $\varphi(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(x)$, where $A_i = f^{-1}\{\{a_i\}\}$ and $\mathbb{1}_{A}(x) = \mathbb{1}_{\{x \in A\}}$ is the characteristic function of the set $A$ (or indicator of the condition $x \in A$). Thus $\varphi$ is a simple function if there exist a finite number of measurable sets $B_1, \ldots, B_n$ and values $b_1, \ldots, b_n$ such that $\varphi(x) = \sum_{i=1}^{n} b_i \mathbb{1}_{B_i}(x)$, for every $x \in \Omega$; and this presentation is by no means unique. It is not so hard to show that $f$ is a simple function if and only if $f^{-1}(B(\mathbb{R}))$ is a finite sub $\sigma$-algebra of $\mathcal{F}$.

The set of simple functions form an algebra and a lattice, i.e., if $\varphi$ and $\psi$ are simple functions so are the their sum $\varphi + \psi$, their product $\varphi \psi$, their max $\varphi \vee \psi$, and their min $\varphi \wedge \psi$. A key point used later is the following approximation result.

**Proposition B.9.** If $(\Omega, \mathcal{F})$ is a measurable space and $f : \Omega \to [0,\infty]$ is measurable, then there exists a sequence of simple functions $\{f_n\}$ such that $0 \leq f_1 \leq \ldots \leq f_n \leq \ldots \leq f$, $f_n \to f$ pointwise in $\Omega$, and $f_n \to f$ uniformly on every set where $f$ is bounded.

**Proof.** Take $n$ and define $F^k_n = f^{-1}([k2^{-n}, (k+1)2^{-n}])$ and $F_n = f^{-1}([2^n, \infty])$, for every $k$ between 1 and $2^{2n} - 1$,
Because $f$ is measurable $F_n^k, F_n \in \mathcal{F}$. Now, set
\[ f_n(x) = 2^n \mathbb{1}_{F_n}(x) + \sum_{k=1}^{2^n-1} k 2^{-n} \mathbb{1}_{F_n^k}(x), \quad \forall x \in \Omega. \]
By construction we have $f_n \leq f_{n+1}$ for any $n$, and $0 \leq f - f_n \leq 2^{-n}$ on the set where $f \leq 2^n$. Hence, conclusion follows. \hfill \Box

If we apply the above arguments by components or coordinates then the previous approximation result remains true for a measurable function with values in $[0, \infty]^d$.

**Corollary B.10.** Let $\mathcal{G} \subset 2^\Omega$ be $\pi$-class and $\mathbb{V}$ be a set of real-valued functions defined on $\Omega$ with the following properties: (1) $\mathbb{1}_\Omega \in \mathbb{V}$ and $\mathbb{1}_A \in \mathbb{V}$, for every $A \in \mathcal{G}$, (2) if $u, v \in \mathbb{V}$ then $\alpha u + \beta v \in \mathbb{V}$ for every $\alpha, \beta \in \mathbb{R}$, (3) if $\{v_n\}$ is a monotone increasing convergent sequence of functions in $\mathbb{V}$, i.e., $v_n \leq v_{n+1}$, $\forall n$ and $v_n(x) \to v(x)$, finite $\forall x \in \Omega$, then $v \in \mathbb{V}$. Then $\mathbb{V}$ contains all $\sigma(\mathcal{G})$ measurable functions.

**Proof.** Let $\mathcal{A}$ be the class of $A \subset \Omega$ such that $\mathbb{1}_A \in \mathbb{V}$. Since $\mathbb{1}_{A \triangle B} = \mathbb{1}_A - \mathbb{1}_B$ if $A \supset B$ and $\mathbb{V}$ is a vector space, the class $\mathcal{A}$ is stable under monotone differences. Moreover, $\mathcal{A}$ is stable under monotone countable unions because $\mathbb{V}$ is stable under the monotone increasing pointwise convergence. Hence $\mathcal{A}$ is a $\lambda$-class containing $\mathcal{G}$, and invoking Proposition B.5, we deduce $\sigma(\mathcal{G}) \subset \mathcal{A}$. Now, writing any measurable function $f = f^+ - f^-$ and applying Proposition B.9, we conclude. \hfill \Box

- **Remark B.11.** If $\{g_i : i \in I\}$ is a family of measurable functions then the $\sigma$-algebra $\mathcal{G} = \sigma(g_i : i \in I)$ generated by this family is countable dependent in the following sense: For any set $A$ in $\mathcal{G}$ there exists a countable subset of indexes $J$ of $I$ such that $A$ is also measurable with respect to $\sigma(g_i : i \in J)$. Indeed, to check this, observe that the class of sets having the above property forms a $\sigma$-algebra. Thus, if $h$ is a measurable function on $(\Omega, \mathcal{G})$ assuming only a finite number (or countable) of values (i.e., a simple function) then there exist a measurable function $k$ and a countable subset $J$ of $I$ such that $h = k(g_i : i \in J)$, i.e., $k$ is independent of the coordinates $i$ in $I \setminus J$. Indeed, such a function $h$ has the form $h = \sum_n a_n \mathbb{1}_{A_n}$ for some sequence $\{A_n\}$ of disjoint measurable sets and some sequence $\{a_n\}$ of values. Each $A_n$ is measurable with respect to $\sigma(g_i : i \in J_n)$ for some countable subset of indexes $J_n$ of $I$, and so, $h$ is measurable with respect to $\sigma(g_i : i \in J)$ for the measurable subset $J = \bigcup_n J_n$ of $I$. Therefore, the function $k$ can be taken measurable with respect to $\sigma(g_i : i \in J)$.

Given a measurable space $(\Omega, \mathcal{F})$, we may not necessarily know if a singleton is measurable, i.e. $\{\omega\} \in \mathcal{F}$. However, we define the *atoms* of $\mathcal{F}$ as elements $A \in \mathcal{F}$ such that $A \neq \emptyset$, and any $B \subset A$ with $B \in \mathcal{F}$ results $B = \emptyset$ or $B = A$. We can show that any measurable function must be constant on every atom, and in general, the family (possible uncountable) of all atoms (of $\mathcal{F}$) may not generate $\mathcal{F}$. For instance, the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^d)$ contains all singletons, but,
any uncountable Borel set with an uncountable complement does not belong to the \( \sigma \)-algebra generated by \( \{ x \} \) with \( x \in \mathbb{R}^d \).

Perhaps, some readers could benefice from taking a quick look at Al-Gwaiz and Elsanousi [5, Chapter 10, pp. 349–392], for a discussion on preliminaries of the Lebesgue measure in \( \mathbb{R} \).

## B.2 Caratheodory’s Arguments

It is rather simple to define a finitely additive measure. For instance the Jordan-Riemann measure \( m \) in \( \mathbb{R} \), namely, \( A \subset 2^\mathbb{R} \) is the algebra of sets that can be written as a disjoint finite union of intervals (closed, open, semi-open, bounded, unbounded), say a generic interval different from \( \mathbb{R} \) is denoted by \( I \) and has the form \((a, b), [a, b], [a, b)\) or \((a, b]\) with \( a \leq b, a, b \in [-\infty, +\infty] \), and \( m(I) = b - a \), \( m(\mathbb{R}) = \infty \) and finally for \( A = \bigcup_{i=1}^n I_i \) with \( I_i \cap I_j = \emptyset \) if \( i \neq j \) where we define \( m(A) = \sum_{i=1}^n m(I_i) \). A more difficult step is to show the \( \sigma \)-additivity and to extend the definition of \( m \) to a \( \sigma \)-algebra (the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}) \) in the case of the Jordan-Riemann measure).

### B.2.1 Caratheodory’s construction

**Definition B.12.** A function \( \mu^* : 2^\Omega \rightarrow [0, \infty] \) is called an outer measure (or exterior measure) on \( \Omega \) if (1) \( \mu^*(\emptyset) = 0 \), (2) \( A \subset B \) implies \( \mu^*(A) \leq \mu^*(B) \) (monotone or isotone), and (3) \( \mu^*(\bigcup_{n=1}^\infty A_n) \leq \sum_{n=1}^\infty \mu^*(A_n) \) (sub \( \sigma \)-additive). Next a subset \( A \subset \Omega \) is said to be \( \mu^* \)-measurable if \( \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \), for every \( E \subset \Omega \), i.e., \( \mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \), in view of the sub-additivity.

**Theorem B.13.** If \( \mu^* \) is an outer measure on \( \Omega \) and \( \mathcal{F} \) is the class of all \( \mu^* \)-measurable sets then \( \mathcal{F} \) is a \( \sigma \)-algebra and the restriction \( \mu \) of \( \mu^* \) to \( \mathcal{F} \) is a complete measure.

**Proof.** First, because the definition of \( \mu^* \)-measurability is symmetric in \( A \) and \( A^c \), the class \( \mathcal{F} \) is stable under the formation of complement. Next, if \( A, B \in \mathcal{F} \) and \( A \subset \Omega \), by the subadditivity we have

\[
\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A \cap B) + \\
+ \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c) \geq \\
\geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c).
\]

Hence \( A \cup B \in \mathcal{F} \), i.e., the class \( \mathcal{F} \) is an algebra. Moreover, if \( A, B \in \mathcal{F} \) and \( A \cap B = \emptyset \) then

\[
\mu^*(A \cup B) = \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \cap A^c) = \mu^*(A) + \mu^*(B),
\]

i.e., \( \mu^* \) is finitely additive on \( \mathcal{F} \).
To show that $\mathcal{F}$ is a $\sigma$-algebra we have to prove only that $\mathcal{F}$ is stable under countably disjoint unions. Thus, for any sequence $\{A_j\}$ of disjoint sets in $\mathcal{F}$, define $B_n = \bigcup_{j=1}^{n} A_j$ and $B = \bigcup_{j=1}^{\infty} A_j$ to get

$$
\mu^*(E \cap B_n) = \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) =
\mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}), \quad \forall E \subset \Omega,
$$

and by induction, this yields $\mu^*(E \cap B_n) = \sum_{j=1}^{n} \mu^*(E \cap A_j)$. Therefore

$$
\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \sum_{j=1}^{n} \mu^*(E \cap A_j) + \mu^*(E \cap B^c),
$$

and as $n \to \infty$ we obtain

$$
\mu^*(E) \geq \sum_{j=1}^{\infty} \mu^*(E \cap A_j) + \mu^*(E \cap B^c) \geq \mu^*\left( \bigcup_{j=1}^{\infty} (E \cap A_j) \right) + \mu^*(E \cap B^c) \geq \mu^*(E),
$$
i.e., all the above inequalities becomes equalities. Hence $B \in \mathcal{F}$, and by taking $E = B$ we have $\mu^*(B) = \sum_{j=1}^{\infty} \mu^*(A_j)$, i.e., $\mu^*$ is countably additive on $\mathcal{F}$.

Finally, if $\mu^*(A) = 0$ then we have

$$
\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \leq \mu^*(E), \quad \forall E \subset \Omega,
$$
i.e., $A \in \mathcal{F}$, and $\mu = \mu^*|_{\mathcal{F}}$ is a complete measure. \hfill \square

At this point, we need to discuss how we obtain an outer measure.

**Proposition B.14.** Let $\mathcal{E} \subset 2^\Omega$ and $\mu : \mathcal{E} \to [0, +\infty]$ be such that $\emptyset \in \mathcal{E}$, $\mu(\emptyset) = 0$ and $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$, for some sequence $\{\Omega_n\}$ in $\mathcal{E}$. Define

$$
\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(E_n) : E_n \in \mathcal{E}, \ A \subset \bigcup_{n=1}^{\infty} E_n \right\}, \quad \forall A \subset \Omega. \quad (B.1)
$$

Then $\mu^*$ is an outer measure on $\Omega$. Moreover, if a set $A \subset \Omega$ satisfies $\mu(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$, for every $E$ in $\mathcal{E}$ with $\mu(E) < \infty$ then $A$ is $\mu^*$-measurable.

**Proof.** Since $\emptyset \in \mathcal{E}$ and $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$, with $\Omega_n \in \mathcal{E}$, the set function $\mu^*$ is defined for every $A \in 2^\Omega$ and $\mu^*(\emptyset) = 0$. If $A \subset B$ then any time we cover $B$ with elements in $\mathcal{E}$ also we cover $A$, and so the infimum satisfies $\mu^*(A) \leq \mu^*(B)$.

To check the sub $\sigma$-additivity, let $\{A_n\}$ a sequence in $2^\Omega$. The definition of infimum ensures that for every $\varepsilon > 0$ and $n$ there is a sequence $\{E_j^n\}$ such that

$$
A_n \subset \bigcup_{j=1}^{\infty} E_j^n \quad \text{and} \quad \sum_{j=1}^{\infty} \mu(E_j^n) \leq \mu^*(A_n) + 2^{-n}\varepsilon.
$$
Since ε is arbitrary, definition (B.1) yields a μ* sub σ-additivity, i.e., μ* is an outer measure.

Finally, pick a set $A \subset \Omega$ satisfying $\mu(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$, for every $E$ in $\mathcal{E}$ with $\mu(E) < \infty$ (note that for $\mu(E) = \infty$ the inequality is trivially satisfied). Pick any set $F \subset \Omega$ and a sequence $\{E_n\} \subset \mathcal{E}$ covering $F$. Since $\bigcap_{n}(E_n \cap A) = F \cap A$ and $\bigcap_{n}(E_n \cap A^c) = F \cap A^c$, the sub σ-additivity of μ* implies

$$
\sum_{n} \mu(E_n) \geq \sum_{n} \mu^*(E_n \cap A) + \sum_{n} \mu^*(E_n \cap A^c) \geq \mu^*(F \cap A) + \sum_{n} \mu^*(F \cap A^c),
$$

and by taking the infimum over all covers we deduce $\mu(F) \geq \mu^*(F \cap A) + \mu^*(F \cap A^c)$, which means that $A$ is μ*-measurable.

\openbullet **Remark** B.15. Recall the notation $\sum_n E_n$ to indicate a disjoint union, i.e., $\sum_n E_n = \bigcup_n E_n$ with $E_n \cap E_m = \emptyset$ if $n \neq m$. Assume that the class $\mathcal{E}$ is a semi-ring and μ is additive on $\mathcal{E}$, i.e., $E = \sum_{i=1}^{n} E_i$, $E$ and $E_i$ belong to $\mathcal{E}$ yield $\mu(E) = \sum_{i=1}^{n} \mu(E_i)$. Then the outer measure μ* induced by μ by means of (B.1) satisfies

$$
\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(E_n) : E_n \in \mathcal{E}, A \subset \bigcup_{n=1}^{\infty} E_n \right\}, \quad \forall A \subset \Omega.
$$

Indeed, if $\{E_n : n \geq 1\} \subset \mathcal{E}$ is a covering of $A$ then define $E'_1 = E_1$, $E'_2 = E_2 \setminus E_1$, and by induction

$$
E'_n = (E_n \setminus E_{n-1}) \cup (E_n \setminus E_{n-2}) \cup \cdots (E_n \setminus E_1).
$$

Because the class $\mathcal{E}$ is a semi-ring, we can write each $E'_n$ as a disjoint union of sets in $\mathcal{E}$, i.e., $E'_n = \sum_{i=1}^{k_n} E''_{n,i}$. The additivity of μ implies that $\mu(E_n) \geq \sum_{i=1}^{k_n} \mu(E''_{n,i})$. Hence, $\{E_{n,i} : i = 1, \ldots, k_n, n \geq 1\} \subset \mathcal{E}$ is a countable cover of $A$ satisfying

$$
A \subset \sum_{n} \sum_{i=1}^{k_n} E''_{n,i} \quad \text{and} \quad \sum_{n} \mu(E_n) \geq \sum_{n} \sum_{i=1}^{k_n} \mu(E''_{n,i}),
$$

which complete the proof. \openbullet
Theorem B.16. If $\mu$ is a measure on an algebra $\mathcal{E}$ and $\mu^*$ is defined by (B.1) then (a) $\mu^*|_{\mathcal{E}} = \mu$ and (b) every set in $\mathcal{A} = \sigma(\mathcal{E})$ is $\mu^*$-measurable and $\mu = \mu^*|_{\mathcal{A}}$ is a measure. Moreover, if $\mu$ is $\sigma$-finite (i.e., there exists $\{A_n\} \subset \mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n = \Omega$ with $\mu(A_n) < \infty$) then $\mu$ is uniquely determinate on $\mathcal{A}$, i.e., if $\nu$ is another measure on $\mathcal{A}$ such that $\nu|_{\mathcal{E}} = \mu$ then $\nu = \mu$.

Proof. To show (a), take a generic element $E \in \mathcal{E}$ and for any countable cover $\{E_n\} \subset \mathcal{E}$ define $F_n = E \cap (E_n \setminus \bigcup_{i=1}^{n-1} E_i)$ to satisfy $F_n \in \mathcal{E}$, $E = \bigcup_{n=1}^{\infty} F_n$, $F_n \cap F_m = \emptyset$ for $n \neq m$ and $F_n \subset E_n$. Hence $\mu(E) = \sum_{n=1}^{\infty} \mu(F_n) \leq \sum_{n=1}^{\infty} \mu(E_n)$, and since the cover is arbitrary, we deduce $\mu(E) \leq \mu^*(E)$. On the other hand, choosing $E_1 = E$ and $E_i = \emptyset$ for $i \geq 2$ we get $\mu^*(E) \leq \mu(E) + 0$, i.e., $\mu(E) = \mu^*(E)$ for every $E \in \mathcal{E}$.

To establish (b), we need to show that every set $E \in \mathcal{E}$ is $\mu^*$-measurable. Thus, take any $F \subset \Omega$ and $\varepsilon > 0$ and by definition of $\mu^*(F)$, there exists a countable cover $\{F_n\} \subset \mathcal{E}$ of $F$ such that $\mu^*(F) + \varepsilon \geq \sum_{n=1}^{\infty} \mu(F_n)$. Since $\{F_n \cap E\}$ and $\{F_n \cap E^c\}$ cover $F \cap E$ and $F \cap E^c$, the additivity of $\mu$ on $\mathcal{E}$ implies

$$\mu^*(F) + \varepsilon \geq \sum_{n=1}^{\infty} \left( \mu(F_n \cap E) + \mu(F_n \cap E^c) \right) \geq \mu^*(F \cap E) + \mu^*(F \cap E^c),$$

and because $\varepsilon$ is arbitrary, the set $E$ is $\mu^*$-measurable. Next, by means of Theorem B.13, $\mu$ induces an outer measure $\mu^*$. In turn, $\mu^*$ yields a measure $\bar{\mu}$ on the $\sigma$-algebra $\mathcal{A}^*$ of $\mu^*$-measurable sets. Since $\mathcal{E} \subset \mathcal{A}^*$ we deduce that $\sigma(\mathcal{E}) = \mathcal{A} \subset \mathcal{A}^*$. Moreover, by (a), $\mu^*|_{\mathcal{E}} = \mu$.

Let us prove that the extension to $\mathcal{A}$ is unique. Suppose that $\nu$ is another measure such that $\nu|_{\mathcal{E}} = \mu$. For any $A \in \mathcal{A}$ and any sequence $\{E_i\} \subset \mathcal{E}$ with $A \subset \bigcup_{i=1}^{\infty} E_i$ we have $\nu(A) \leq \sum_{i=1}^{\infty} \nu(E_i) = \sum_{i=1}^{\infty} \mu(E_i)$, which yields $\nu(A) \leq \bar{\mu}(A)$. Setting $E = \bigcup_{i=1}^{\infty} E_i$, we get $\nu(E \setminus A) \leq \bar{\mu}(E \setminus A)$ and

$$\nu(E) = \lim_{n \to \infty} \nu \left( \bigcup_{i=1}^{n} E_i \right) = \lim_{n \to \infty} \mu \left( \bigcup_{i=1}^{n} E_i \right) = \bar{\mu}(E).$$

If $\bar{\mu}(A) < +\infty$, for any $\varepsilon > 0$ we can choose a cover $\{E_i\}$ such that $\bar{\mu}(E) < \bar{\mu}(A) + \varepsilon$, i.e., $\bar{\mu}(E \setminus A) < \varepsilon$. Then

$$\bar{\mu}(A) \leq \bar{\mu}(E) = \nu(E) = \nu(A) + \nu(E \setminus A) \leq \nu(A) + \bar{\mu}(E \setminus A) \leq \nu(A) + \varepsilon,$$

and because $\varepsilon$ is arbitrary, we have $\bar{\mu}(A) = \nu(A)$. Finally, if $\bar{\mu}$ is $\sigma$-finite then $\Omega = \bigcup_{n=1}^{\infty} A_n$, with $\bar{\mu}(A_n) < +\infty$, and we may assume that $A_n \cap A_m = \emptyset$ for $n \neq m$. Hence for any $A \in \mathcal{A}$ we have

$$\bar{\mu}(A) = \sum_{n=1}^{\infty} \bar{\mu}(A \cap A_n) = \sum_{n=1}^{\infty} \nu(A \cap A_n) = \nu(A),$$

i.e., $\nu = \bar{\mu}$.

□
• Remark B.17. If $\mathcal{E}$ is a $\pi$-class (i.e., closed under finite intersections and contains the empty set $\emptyset$) and $\mu$ is a (nonnegative) set function defined on $\mathcal{E}$ then we say that $\mu$ is additive on $\mathcal{E}$ if for every $\varepsilon > 0$ and every $F$ and $E$ in $\mathcal{E}$ there exists a sequence (possible finite) $\{E_n\} \subset \mathcal{E}$ such that $F \setminus E \subset \bigcup_n E_n$ and $\mu(F + \varepsilon) = \mu(F) + \sum_n \mu(F \cap E_n)$. Similarly, we say that $\mu$ is a pre-measure on $\mathcal{E}$ if (a) $\mu(\emptyset) = 0$, (b) $E \subset F, \ E$ and $F$ in $\mathcal{E}$ implies $\mu(E) \leq \mu(F)$ (i.e., monotone on $\mathcal{E}$), (c) $E \subset \bigcup_n E_n$, $E$ and $E_n$ in $\mathcal{E}$ implies $\mu(E) \leq \sum_n \mu(E_n)$ (i.e., sub $\sigma$-additive on $\mathcal{E}$). Now, remark that in the proof of the precedent Theorem B.16, we have also proved that (1) if $\mathcal{E}$ is a $\pi$-class and the initial set function $\mu$ is additive then any set in the $\sigma$-algebra generated by $\mathcal{E}$ is $\mu$-measurable; and (2) if the initial set function $\mu$ is a pre-measure then $\mu^* = \mu$ on $\mathcal{E}$. In particular, if the initial set function $\mu$ can be extended to a measure on the $\sigma$-algebra $\mathcal{A} = \sigma(\mathcal{E})$ generated by a class $\mathcal{E}$ (satisfying the assumptions of Proposition B.14) then $\mu = \mu^*$ on $\mathcal{E}$ (but not necessarily on $\mathcal{A}$); and moreover, if $\mathcal{E}$ is a $\pi$-class then any set in $\mathcal{A}$ is $\mu^*$-measurable. □

B.2.2 From a Semi-Ring

If the initial set function $\mu$ is a finitely additive measure on a ring $\mathcal{E}$ then we can define the outer measure $\mu^*$, for any $A \subset \Omega$, by

either $\mu^*(A) = \inf \left\{ \lim_n \mu(E_n) : E_n \in \mathcal{E}, A \subset \bigcup_{n=1}^{\infty} E_n, E_n \subset E_{n+1} \right\}$,

or $\mu^*(A) = \inf \left\{ \sum_n \mu(E_n) : E_n \in \mathcal{E}, A \subset \bigcup_{n=1}^{\infty} E_n \right\}$,

instead of using (B.1). Actually, the last expression with coverings in the form of disjoint unions remains valid for a semi-ring $\mathcal{E}$. Similarly, if $\mu$ is a measure on a $\sigma$-algebra $\mathcal{A}$ then

$\mu^*(A) = \inf \left\{ \mu(E) : E \in \mathcal{E}, A \subset E, \forall A \subset \Omega, \right\}$,

yields an outer measure. Denoting by $\mathcal{A}^*$ the $\sigma$-algebra of all $\mu^*$-measurable sets, we have a complete measure $(\bar{\mu}, \mathcal{A}^*)$ by taking $\bar{\mu} = \mu^*|_{\mathcal{A}^*}$, which is an extension of $(\mu, \mathcal{A})$, and a set $N \subset \Omega$ is negligible if and only if $\mu^*(N) = 0$.

Recall the algebra $\mathcal{A}$ (ring) generated by a $\mathcal{S}$ semi-algebra (semi-ring) is the class of finite disjoint unions, i.e., $A \in \mathcal{A}$ if and only if $A = \bigcup_{i=1}^{n} A_i$ for some $A_i \in \mathcal{S}$.

Proposition B.18. Let $\mathcal{E}$ be a semi-ring and $\mu : \mathcal{E} \to [0, \infty)$ be a $\sigma$-additive finite-valued set function. Then $\mu$ can be uniquely extended to $\sigma$-additive set function on the $\sigma$-ring $\mathcal{R}$ generated by $\mathcal{E}$. Moreover, a further unique extension of the measure $\mu$ to the $\sigma$-ring $\mathcal{R}$ of all ($\sigma$-finite) $\mu^*$-measurable sets is also possible. In particular, if there exists sequence $\{E_n\} \subset \mathcal{E}$ such that $\Omega = \bigcup_{n=1}^{\infty} E_n$ and $\mu(E_n) < \infty$, then $\mu$ can be uniquely extended to a measure on the $\sigma$-algebra $\mathcal{A}$ generated by $\mathcal{E}$. Furthermore, a set $A \subset \Omega$ is $\mu^*$-measurable if and only if $\mu(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$, for every $E \in \mathcal{E}$.
Proof. If \( \mathcal{R}_0 \) is the ring generated by \( \mathcal{E} \) then, recalling that any set in \( \mathcal{R}_0 \) can be written as a finite disjoint union of elements in \( \mathcal{E} \), we extend the definition of \( \mu \) to \( \mathcal{R}_0 \),

\[
\mu(A) = \sum_{i=1}^{n} \mu(E_i), \quad A = \bigcup_{i=1}^{n} E_i, \quad E_i \cap E_j = \emptyset \text{ if } i \neq j.
\]

Because there is only a finite sum (or disjoint union), we deduce that \( \mu \) remains \( \sigma \)-additive on the ring \( \mathcal{R}_0 \). At this point, we revise the proof of Theorem B.16 (remarking that \( F_n \cap E^c = F_n \setminus E \in \mathcal{R}_0 \), for every \( F_n, E \in \mathcal{R}_0 \)) to check that the algebra generated by \( \mathcal{E} \) can be replaced by the ring \( \mathcal{R}_0 \) and the results remain valid. Hence, \( \mu \) has a unique extension to \( \sigma \)-ring \( \mathcal{R} \) generated by \( \mathcal{E} \).

It is clear that if \( \Omega = \bigcup_{n=1}^{\infty} E_n \) and \( \mu(E_n) < \infty \), for every \( n \), then the \( \sigma \)-ring \( \mathcal{R} \) is indeed the \( \sigma \)-algebra \( \mathcal{A} = \sigma(\mathcal{E}) \).

Finally, Theorem B.16 also ensure a unique extension to the \( \sigma \)-ring \( \bar{\mathcal{R}} \) of all \( \mu^* \)-measurable sets. Because the initial set function \( \mu \) assume only finite values, all set in \( \sigma \)-ring \( \bar{\mathcal{R}} \) are \( \sigma \)-finite. In any case, the uniqueness of the extension is only warranty on the \( \sigma \)-ring \( \bar{\mathcal{R}} \) of all \( \sigma \)-finite \( \mu^* \)-measurable sets.

It is also clear that because \( \mu^* = \mu \) on \( \mathcal{E} \), Proposition B.14 yields the stated characterization of a \( \mu^* \)-measurable set in term of sets in the semi-ring \( \mathcal{E} \).  

- **Remark B.19.** In the statement of Proposition B.18, we may initially assume \( \mu : \mathcal{E} \to [0, \infty] \) and define \( \mathcal{E}_0 = \{ E \in \mathcal{E} : \mu(E) < \infty \} \), which is again a semi-ring, i.e., \( \mathcal{R} \) is the \( \sigma \)-ring generated by \( \mathcal{E}_0 \). In general, a subset \( A \) of \( \Omega \) is called \( \sigma \)-finite relative to a set function \( \mu \) defined on a class \( \mathcal{E} \subset 2^\Omega \) if there exists a sequence \( \{E_n\} \) in \( \mathcal{E} \) such that \( A \subset \bigcup_n E_n \) and \( \mu(E_n) < \infty \), for every \( n \). Thus \( \mathcal{R} \) is the \( \sigma \)-ring generated by the \( \sigma \)-finite sets in \( \mathcal{E} \) relative to \( \mu \). Therefore, if the initial class \( \mathcal{E} \) is a semi-algebra then we may be forced to define the semi-ring \( \mathcal{E}_0 \) as above, which may not be a semi-algebra.

- **Remark B.20.** The reader can verify that only the finitely additive character (instead of the \( \sigma \)-additivity) of the set function \( \mu \) is used to prove that any set in \( \mathcal{E} \) is \( \mu^* \)-measurable, that \( \mu \leq \mu^* \) on \( \mathcal{E} \) and that a set \( A \subset \Omega \) is \( \mu^* \)-measurable if and only if \( \mu(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \), for every \( E \) in \( \mathcal{E} \). However, to check that \( \mu = \mu^* \) on \( \mathcal{E} \) the \( \sigma \)-additivity is involved. Sometimes, a function defined on a (semi-)ring is called *content* if it is additive and *pre-measure* if it is \( \sigma \)-additive. In this context, finitely additive on a semi-ring \( \mathcal{E} \) means that \( \mu(E) = \sum_{i<n} \mu(E_i) \) whenever \( E = \sum_{i<n} E_i \) with all sets in \( \mathcal{E} \), just the case of two sets may not be sufficient.

- **Remark B.21.** Recall that if \( \mathcal{S}_i \) is a semi-ring (semi-algebra) in a measure space \( (\Omega_i, \mathcal{F}_i, \mu_i) \), for \( i = 1, 2 \), then \( \mathcal{S} = \{ S_1 \times S_2 : S_i \in \mathcal{S}_i, i = 1, 2 \} \), is a semi-ring (semi-algebra). Thus the product expression \( \mu(S_1 \times S_2) = \mu_1(S_1)\mu_2(S_2) \) defines an additive measure on \( \mathcal{S} \) (or in Cartesian product \( \mathcal{F}_1 \times \mathcal{F}_2 \)), which can be extended to the product \( \sigma \)-algebra \( \mathcal{F} = \sigma(S_1) \otimes \sigma(S_2) \), by the Caratheodory’s extension Theorem B.16. However, to verify that \( \mu^* = \mu \) on the semi-ring \( \mathcal{S} \), we need to check that \( \mu = \mu_1 \times \mu_2 \) is indeed \( \sigma \)-additive on \( \mathcal{S} \). Actually, this
will be address later by either the construction of the integral or a discussion on inner measures (more tools are needed to prove this fact).

Summing up, the construction of a \((\sigma\text{-finite})\) measure on a \(\sigma\text{-algebra} A\) begins with a \(\sigma\text{-additive (set) function defined on a semi-ring } E\), which generates \(A\). Actually, the \(\sigma\text{-algebra } A^* \) of all \(\mu^*\)-measurable sets is usually strictly larger than \(A = \sigma(E)\). Usually, the passage of a finitely additive measure defined on an algebra to a \(\sigma\text{-additive measure on the generated } \sigma\text{-algebra is called Hopf’s extension theorem, e.g., see Richardson [106, Section 2.4, pp. 24–30].}

\textbf{Remark B.22.} The uniqueness argument can be restated as following: If \(E \subset 2^\Omega\) is a \(\pi\text{-class} \) and \(\mu\) and \(\nu\) are two measures on \(A = \sigma(E)\) such that (1) \(\mu = \nu\) on \(E\) and (2) there exists a monotone increasing sequence \(\{E_n\}\) of elements in \(E\) satisfying \(\Omega = \bigcup_n E_n\) and \(\mu(E_n) = \nu(E_n) < \infty\) for every \(n\), then \(\mu = \nu\) on \(A\). This assertion (and previous statements) about the unique extension of a measure \(\mu\) initially defined on a \(\pi\text{-class} \) requires the \(\sigma\text{-finite property of } \mu \) with respect to \(E\). In general, the assumption \(\mu(E) < \infty\) (for every \(E \in E\)) yields a unique measure on the \(\sigma\text{-ring } R\) generated by \(E\), see also Remark B.19. Indeed, a monotone argument shows that the class of sets in \(R\) which are included in a countable union of sets in \(E\) is indeed the whole \(\sigma\text{-ring } R\). Hence, we deduce that \(\mu = \nu\) on \(R\).

For instance, the reader may take a look at Taylor [122, Chapter 4, 177–225], and many other textbooks.

\section{B.3 Inner Measure Approach}

In a way analogous to the outer measure in Section B.2 (using the Caratheodory splitting method), we may develop the inner measure construction. However, this section is not referred to for the typical Lebesgue measure defined in the next section, it could be only used later, when topology is involved. Begin with

\subsection{B.3.1 Inner Measures}

\textbf{Definition B.23.} A function \(\mu_* : 2^\Omega \to [0, \infty]\) is called an \textit{inner measure} (or \textit{interior measure}) on \(\Omega\) if (1) \(\mu_*(\emptyset) = 0\), (2) \(A \subset B\) implies \(\mu_*(A) \leq \mu_*(B)\) (monotone or isotone), and (3) \(A \cap B = \emptyset\) implies \(\mu_*(A \cup B) \geq \mu_*(A) + \mu_*(B)\) (super-additive). Next a subset \(A \subset \Omega\) is said to be \(\mu_*\text{-measurable} \) if \(\mu_*(E) = \mu_*(E \cap A) + \mu_*(E \cap A^c)\), for every \(E \subset \Omega\), i.e., \(\mu_*(E) \leq \mu_*(E \cap A) + \mu_*(E \cap A^c)\), in view of the super-additivity.

Note that by induction, the monotony and super-additivity of \(\mu_*\) implies \(\mu_*\left(\sum_{i=1}^\infty A_i\right) \geq \mu_*\left(\sum_{i=1}^n A_i\right) \geq \sum_{i=1}^n \mu_*(A_i)\), and as \(n \to \infty\), we deduce a property that could be called super \(\sigma\text{-additivity}. \) It is also clear that the sets \(\emptyset\) and \(\Omega\) are \(\mu_*\)-measurable.

\textbf{Proposition B.24.} If \(\mu_*\) is an inner measure on \(\Omega\) and \(A\) is the class of all \(\mu_*\text{-measurable sets} \) then \(A\) is an algebra and the restriction \(\mu\) of \(\mu_*\) to \(A\) is a complete finitely additive measure.
Proof. First, because the definition of $\mu_*$-measurability is symmetric in $A$ and $A^c$, the class $\mathcal{A}$ is stable under the formation of complement. Next, for any $A, B \in \mathcal{A}$ and $E \subset \Omega$, the equality

$$(E \cap A^c \cap B) \cup (E \cap A \cap B^c) \cup (E \cap A^c \cap B^c) = E \cap (A \cap B)^c$$

and the super-additivity of $\mu_*$ imply

$$
\mu_*(E) = \mu_*(E \cap A) + \mu_*(E \cap A^c) = \mu_*(E \cap A \cap B) + \\
+ \mu_*(E \cap A \cap B^c) + \mu_*(E \cap A^c \cap B) + \mu_*(E \cap A^c \cap B^c) \leq \\
\mu_*(E \cap (A \cap B)) + \mu_*(E \cap (A \cap B)^c).
$$

Hence $A \cap B \in \mathcal{A}$, i.e., the class $\mathcal{A}$ is an algebra. Moreover, if $A, B \in \mathcal{A}$ and $A \cap B = \emptyset$ then

$$
\mu_*(A \cup B) = \mu_*((A \cup B) \cap A) + \mu_*((A \cup B) \cap A^c) = \mu_*(A) + \mu_*(B),
$$

i.e., $\mu_*$ is finitely additive on $\mathcal{A}$.

Finally, if $\mu_*(A) = 0$ and $B \subset A$ with $A$ in $\mathcal{A}$ then the monotony of $\mu_*$ implies

$$
\mu_*(E) \leq \mu_*(E \cap A) + \mu_*(E \cap A^c) = \\
= \mu_*(E \cap A^c) \leq \mu_*(E \cap B) + \mu_*(E \cap B^c), \quad \forall E \subset \Omega,
$$

i.e., $B \in \mathcal{A}$, and $\mu = \mu_*|_{\mathcal{A}}$ is a complete finitely additive measure. \qed

The essential properties of an inner measure are captured by the expression

$$
\mu_*(A) = \sup \{ \mu_*(B) : B \subset A, \ \mu_*(B) < \infty \}, \quad \forall A \in 2^\Omega. \quad (B.2)
$$

Indeed, any set function $\mu_*$ with $\mu_*(\emptyset) = 0$ satisfying the sup representation (B.2) is monotone, super-additive, and semi-finite (i.e., for every set $A$ with $\mu_*(A) = \infty$ there is a sequence $\{A_n\}$ such that $A_n \subset A$ and $\mu_*(A_n) \to \infty$). Conversely, any semi-finite inner measure $\mu_*$ satisfies (B.2).

**B.3.2 Inner Construction**

Similarly to the previous sections, our intension is to construct an inner measure $\mu_*$ (such that its restriction to the $\mu_*$-measurable sets is a measure) out of a finite-valued set $\mu : \mathcal{K} \to [0, \infty)$ defined on a $\pi$-class $\mathcal{K}$ with $\mu(\emptyset) = 0$. A good candidate is the following sup expression

$$
\mu_*(A) = \sup \left\{ \sum_{i=1}^{n} \mu_*(K_i) : \sum_{i=1}^{n} K_i \subset A, \ K_i \in \mathcal{K} \right\}, \quad \forall A \in 2^\Omega. \quad (B.3)
$$

Due to the supremum, there is not need to allow infinite series of sets inside $A$, but because $\mathcal{K}$ is only a $\pi$-class, a finite union is needed. Moreover, contrary to the case of a semi-ring, additivity on a $\pi$-class is almost meaningless and
replaced with the so-called $\mathcal{K}$-tightness, i.e., for every $K$ and $K'$ in $\mathcal{K}$ with $K' \subset K$ we have $\mu(K) = \mu(K') + \mu_*(K \setminus K')$, in other words, (a) $\mu$ is monotone (i.e., $\mu(K') \leq \mu(K)$) if $K$ and $K'$ in $\mathcal{K}$ with $K' \subset K$ and (b) for every $\varepsilon > 0$ there exists a finite sequence of disjoint sets $\{K_i : i < n\} \subset \mathcal{K}$ such that $\sum_{i<n} \mu(K_i) \subset K \setminus K'$ and $\mu(K) \leq \mu(K') + \varepsilon + \sum_{i<n} \mu(K_i)$. An important role is played by the lattice $\bar{\mathcal{K}}$ generated by $\mathcal{K}$ (i.e., the class of finite unions of sets in $\mathcal{K}$) and the class

$$ F \in \mathcal{K}_F \quad \text{iff} \quad K \in \mathcal{K} \text{ implies } F \cap K \in \bar{\mathcal{K}}. $$

(B.4)

In this context, if any decreasing sequence $\{\bar{K}_n\}$ of finite disjoint unions of sets in $\mathcal{K}$, $\bar{K}_n = \sum_{i<m_n} K_{n,i}$, with $\bigcap_n \bar{K}_n = \emptyset$ satisfies $\sum_{i<m_n} \mu(K_{n,i}) \to 0$, then $\mu$ is called $\sigma$-smooth on $\mathcal{K}$ at $\emptyset$. We are ready to state the main result

**Theorem B.25.** Let $\mu$ be a finite-valued set defined on a $\pi$-class $\mathcal{K}$ with $\mu(\emptyset) = 0$. Then $\mu_*$ defined by (B.3) is an inner measure. Now, denote by $\mathcal{A}$ the algebra of $\mu_*$-measurable sets and assume that $\mu$ is $\mathcal{K}$-tight. Then

$$ A \in \mathcal{A} \quad \text{iff} \quad \mu(K) \leq \mu_*(K \cap A) + \mu_*(K \setminus A) \quad \forall K \in \mathcal{K}, $$

(B.5)

the algebra $\mathcal{A}$ contains the class $\mathcal{K}_F$ defined by (B.4), and $\mu_*|_{\mathcal{K}} = \mu$. Moreover, if $\mu$ is $\sigma$-smooth on $\mathcal{K}$ at $\emptyset$ then $\mathcal{A}$ is a $\sigma$-algebra and $\mu_*$ is a semi-finite complete measure on $\mathcal{A}$, uniquely determined by $\mu$ on the $\mathcal{K}$, i.e., if $\nu$ is another semi-finite measure on a $\sigma$-algebra $\mathcal{F}$ with $\mathcal{K} \subset \mathcal{F} \subset \mathcal{A}$ such that $\nu|_{\mathcal{K}} = \mu$ then $\nu = \mu_*$ on $\mathcal{F}$.

**Proof.** If $E \subset F$ then the supremum defining $\mu_*(E)$ is taken over a larger family, so $\mu_*(E) \leq \mu_*(F)$. When $E \cap F = \emptyset$, each finite disjoint sequences $\{K_i\}$ and $\{K'_i\}$ with $\sum_i K_i \subset E$ and $\sum_i K'_i \subset F$ we can construct another finite disjoint sequence $\{K''_i\}$ with $\sum_i K''_i \subset E \cup F$ and $\sum_i \mu(K''_i) = \sum_i \mu(K_i) + \sum_i \mu(K'_i)$, which means that $\mu_*(E) + \mu_*(F) \leq \mu_*(E \cup F)$. This shows that $\mu_*$ is monotone and super-additive on $2^\Omega$, and thus (B.3) defines an inner measure $\mu_*$. Therefore, Proposition B.24 implies that $\mu_*$ is an additive set function (i.e., a finite additive measure) on algebra $\mathcal{A}$ of all $\mu_*$-measurable sets.

Let $A$ be a set satisfying $\mu(K) \leq \mu_*(K \cap A) + \mu_*(K \setminus A)$ for any $K$ in $\mathcal{K}$. Since $\mu_*$ is super-additive and monotone, if $\sum_{i=1}^n K_i = K \subset E$ with $K_i$ in $\mathcal{K}$ then

$$ \sum_{i=1}^n \mu(K_i) \leq \sum_{i=1}^n \mu_*(K_i \cap A) + \sum_{i=1}^n \mu_*(K_i \setminus A) \leq \mu_*(K \cap A) + \mu_*(K \setminus A) \leq \mu_*(E \cap A) + \mu_*(E \setminus A), $$

and taking the supremum over all finite disjoint sequences $\{K_i\}$ we deduce $\mu_*(E) \leq \mu_*(E \cap A) + \mu_*(E \setminus A)$. The reverse inequality follows from the super-additivity, and therefore, $A$ belongs to $\mathcal{A}$. This shows (B.5) as desired. The fact that $\mathcal{K}$ is stable under finite intersections was not used in the current (or the previous) paragraph, but it is needed for later arguments.
Step 1 (with tightness) From the definition of \( \mu_* \) follows that \( \mu(K) \leq \mu_*(K) \) for every \( K \) in \( \mathcal{K} \). Now, if \( K'' \) belongs to \( \mathcal{K} \) then apply the tightness property to any set \( K \) in \( \mathcal{K} \) and \( K' = K \cap K'' \) to get

\[
\mu(K) = \mu(K \cap K'') + \mu_*(K \setminus K'') \leq \mu_*(K \cap K'') + \mu_*(K \setminus K''),
\]

which implies, after invoking (B.5), that \( K'' \) belongs to the algebra \( \mathcal{A} \), i.e., \( \mathcal{K} \subset \mathcal{A} \). Moreover, if \( F \) belongs to \( \mathcal{K}_F \) then for any set \( K \) in \( \mathcal{K} \), the intersection \( K \cap F \) is a finite union of sets in \( \mathcal{K} \subset \mathcal{A} \). Thus \( K \setminus F = K \setminus (K \cap F) \) and \( K \cap F \) belong to the algebra \( \mathcal{A} \), and hence, the additivity of \( \mu_* \) yields

\[
\mu(K) \leq \mu_*(K) = \mu_*(K \cap F) + \mu_*(K \setminus F)
\]

which implies that \( F \) belongs to the algebra \( \mathcal{A} \), i.e., \( \mathcal{K}_F \subset \mathcal{A} \).

To show that \( \mu = \mu_* \) on \( \mathcal{K} \), pick \( K = \sum_{i=1}^n K_i \) with all sets in \( \mathcal{K} \) and use the tightness condition with \( K \) and \( K' = K_1 \) to obtain \( \mu(K) = \mu(K_1) + \mu_*(K \setminus K_1) \).

Since \( \mu \leq \mu_* \) on \( \mathcal{K} \), \( K \setminus K_1 = \bigcup_{i=2}^n K_i \) and \( \mu_* \) is additive on \( \mathcal{A} \supset \mathcal{K} \) we have \( \mu_*(K \setminus K_1) = \sum_{i=2}^n \mu(K_i) \), which yields \( \mu(K) \geq \sum_{i=1}^n \mu(K_i) \), the superadditivity of \( \mu \). Therefore, the sup defining \( \mu_*(K) \) is achieved for \( K \) and \( \mu(K) = \mu_*(K) \) for every \( K \) in \( \mathcal{K} \), which means that \( \mu = \mu_* \) is additive on \( \mathcal{K} \).

Step 2 (\( \sigma \)-smooth) Even if we suppose that \( \mu_* \) is monotone continuous from above on \( \mathcal{K} \) at \( \emptyset \) (i.e., \( \sigma \)-smooth), then \( \mu_* \) is \( \sigma \)-additive on the algebra \( \mathcal{A} \), and therefore, Caratheodory extension Theorem B.16 ensures that \( \mu_* \) can be extended to a measure on the \( \sigma \)-algebra generated by \( \mathcal{A} \), but a priori, the extension needs not to preserve the sup representation (B.3).

The next point is to show that \( \mathcal{A} \) is a \( \mu_* \)-complete \( \sigma \)-algebra, independent of the fact that Caratheodory extension of \( (\mu_*, \mathcal{A}) \) yields a complete measure \( (\overline{\mu}_*, \overline{\mathcal{A}}) \). Actually, the completeness of \( \mu_* \) comes from Proposition B.24.

Let us prove that \( \mu_* \) is \( \sigma \)-smooth on \( \mathcal{A} \) at \( \emptyset \), i.e., if \( \{A_n\} \subset \mathcal{A} \) is a decreasing sequence with \( \bigcap_n A_n = \emptyset \) and \( \mu_*(A_1) < \infty \) then \( \mu_*(A_n) \to 0 \). Indeed, the sup definition (B.3) of \( \mu_* \) ensures that for any \( \varepsilon > 0 \) and for any \( n \geq 1 \) there exist a finite disjoint union \( \tilde{K}_n \) of sets in \( \mathcal{K} \) such that \( \tilde{K}_n \subset A_n \) and \( \mu_*(A_n) - \varepsilon 2^{-n} < \mu_*(\tilde{K}_n) \). Define the decreasing sequence \( \{\tilde{K}_n'\} \) with \( \tilde{K}_n' = \bigcap_{i \leq n} \tilde{K}_i \) (which can be written as a finite disjoint union of sets in \( \mathcal{K} \)) and use the \( \sigma \)-smoothness property of \( \mu \) to obtain \( \mu_*(\tilde{K}_n') \to 0 \). Since the inclusion \( A_n \setminus \tilde{K}_n' \subset \bigcup_{i \leq n} (A_i \setminus \tilde{K}_i) \) yields

\[
\mu_*(A_n \setminus \tilde{K}_n') \leq \sum_{i \leq n} \mu_*(A_i \setminus \tilde{K}_i) \leq \sum_{i \leq n} \varepsilon 2^{-i} \leq \varepsilon,
\]

and \( \mu_*(A_n) = \mu(\tilde{K}_n') + \mu_*(A_n \setminus \tilde{K}_n') \), we deduce that \( \mu_*(A_n) \to 0 \), i.e., \( \mu_* \) is \( \sigma \)-smooth on \( \mathcal{A} \) at \( \emptyset \).

Step 3 (finishing) Now, to check that \( \mathcal{A} \) is a \( \sigma \)-algebra, we have to show only that \( \mathcal{A} \) is stable under the formation of countable intersections, i.e., if \( \{A_i, i \geq 1\} \) is a sequence of sets in \( \mathcal{A} \) then we should show that \( A = \bigcap_i A_i \) also belongs to \( \mathcal{A} \). For this purpose, from the sup definition (B.3) of \( \mu_* \) and because \( \mathcal{A} \) contains any finite union of sets in \( \mathcal{K} \), for any \( \varepsilon > 0 \) and for any set \( K \) in \( \mathcal{K} \)
there exist a set $A' \subset K \cap \mathcal{A}$ in $\mathcal{A}$ such that $\mu_*(K \cap A) - \varepsilon < \mu_*(A')$. Thus, define the decreasing sequence $\{B_n\}$ with $B_n = \bigcap_{i \leq n} A_i$ to have $\bigcap_n (K \cap B_n \cap A') = A'$ and to use the $\sigma$-smoothness of $\mu_*$ on $\mathcal{A}$ with the sequence $(K \cap B_n \cap A') \setminus A$. Hence $\lim_n \mu_*(K \cap B_n \cap A') = \mu_*(A')$, which yields

$$\lim_n \mu_*(K \cap B_n) \geq \lim_n \mu_*(K \cap B_n \cap A') = \mu_*(A') \geq \mu_*(K \cap A) - \varepsilon$$

and proves that $\lim_n \mu_*(K \cap B_n) = \mu_*(K \cap A)$. Recall that $B_n$ is in $\mathcal{A}$ to have

$$\mu(K) \leq \mu_*(K \cap B_n) + \mu_*(K \setminus B_n) \leq \mu_*(K \cap B_n) + \mu_*(K \setminus A),$$

and, after taking $n \to \infty$ and invoking the condition (B.5), to deduce that $A$ belongs to $\mathcal{A}$, i.e., $\mathcal{A}$ is a $\sigma$-algebra.

The final argument is to show that $\mu_*$ is $\sigma$-additive. Indeed, pick a sequence $\{A_n\} \subset \mathcal{A}$ with $A = \sum_n A_n$. If $A$ is a set in $\mathcal{A}$ with finite measure $\mu_*(A) < \infty$ then the $\sigma$-smoothness property of $\mu_*$ on $\mathcal{A}$ implies that $\mu_*(A \setminus \sum_{i < n} A_i) \to 0$, i.e., $\mu_*(A) = \sum_n \mu_*(A_n)$. If $\mu_*(A) = \infty$, the sup definition (B.3) ensures that there exists a sequence $\{A'_k\} \subset \mathcal{A}$ such that $A'_k \subset A$, $\mu_*(A'_k) < \infty$ and $\mu_*(A'_k) \to \infty$. Hence $\mu_*(A'_k) = \sum_n \mu_*(A'_k \cap A_n) \leq \sum_n \mu_*(A_n)$, and as $k \to \infty$ we deduce $\infty = \sum_n \mu_*(A_n)$, i.e., $\mu_*$ is $\sigma$-additive on the $\sigma$-algebra $\mathcal{A}$.

The uniqueness of $\mu_*$ is not really an issue, we have to show that if another semi-finite measure $\nu$ on a $\sigma$-algebra $\mathcal{F} \subset \mathcal{A}$ containing the class $\mathcal{K}$ and such $\nu = \mu$ on $\mathcal{K}$ then $\nu = \mu_*$ on $\mathcal{F}$. Indeed, they both agree on any set of finite measure, and for any set $F$ in $\mathcal{F}$ with infinite measure there exists a sequence $\{F_n\} \subset \mathcal{F}$ with $\nu(F_n) < \infty$, $F_n \subset F$ and $\lim_n \nu(F_n) = \nu(F)$, i.e., $\nu(F) = \mu_*(F)$ too.

Note that the $\sigma$-smoothness on $\mathcal{K}$ at $\emptyset$ and the $\mathcal{K}$-tightness assumptions are really conditions on the $\pi$-class $\mathcal{K}$ of all disjoint unions of sets in $\mathcal{K}$. Indeed, it is clear that if (a) $\mu$ is monotone on $\mathcal{K}$ and (b) $\mu$ is additive on $\mathcal{K}$ (i.e., $\mu(K) = \sum_{i < n} \mu(K_i)$ whenever $K = \sum_{i < n} K_i$ are sets in $\mathcal{K}$) then $\mu$ can be extended (in a unique way) to the $\pi$-class $\mathcal{K}$ preserving (a) and (b) by setting $\mu(\sum_{i < n} K_i) = \sum_{i < n} \mu(K_i)$. Therefore, $\mathcal{K}$-tightness translates into three properties: (a), (b) and (c) for every $K \supset K'$ sets in $\mathcal{K}$ (could be in $\mathcal{K}$) and every $\varepsilon > 0$ there exists $\tilde{K} \subset K \setminus K'$ in $\mathcal{K}$ such that $\mu(K) \leq \mu(K') + \varepsilon + \mu(\tilde{K})$. Similarly, $\sigma$-smoothness on $\mathcal{K}$ at $\emptyset$ translates into one condition: any decreasing sequence $\{\tilde{K}_n\}$ of sets in $\mathcal{K}$ such that $\bigcap_n \tilde{K}_n = \emptyset$ satisfies $\mu(\tilde{K}_n) \to 0$. With this in mind, there is not loss of generality if in Theorem B.25 we assume that the $\pi$-class $\mathcal{K}$ is also stable under the formation of finite disjoint unions.

**Remark** B.26. If the class $\mathcal{K}$ contains the empty set $\emptyset$, but it is not necessarily stable under finite intersections, then the sup-expression (B.3) defines an inner measure $\mu_*$. Hence, Proposition B.24 proves that $\mu_*$ is a finitely additive set function on the algebra $\mathcal{A}$ of $\mu_*$-measurable sets. Moreover, if a subset $A$ of $\Omega$ satisfies $\mu(K) \leq \mu_*(K \cap A) + \mu_*(K \setminus A)$ for any $K$ in $\mathcal{K}$ then $A$ belongs to $\mathcal{A}$. However, it is not affirmed that $\mathcal{K} \subset \mathcal{A}$. 

---

[B.3. Inner Measure Approach] 339

Remark B.27. If the finite-valued set function \( \mu \) defined on the \( \pi \)-class \( K \) can be extended to a (finitely) additive set function \( \bar{\mu} \) define on the semi-ring \( S \) generated by \( K \) then \( \mu \) is necessarily \( K \)-tight. Indeed, first recall that \( \bar{\mu} \) is additive on the semi-ring \( S \) if (by definition) \( \bar{\mu}(S) = \sum_{i<n} \bar{\mu}(S_i) \), for any finite sequence \( \{S_i, i < n\} \) of disjoint sets in \( S \) with \( S = \sum_{i<n} S_i \) also in \( S \). Thus, if \( K \supset K' \) are sets in \( K \) then \( K \setminus K' \) is a finite disjoint unions of sets in \( S \), i.e., \( K \setminus K' = \sum_{i<n} S_i \), and the additivity of \( \bar{\mu} \) implies \( \mu(K) = \mu(K') + \sum_{i<n} \bar{\mu}(S_i) \). Hence, this yields the following monotone property: if \( \sum_{i<n} S_i \subset S \) with all sets in \( S \) then \( \sum_{i<n} \bar{\mu}(S_i) \leq \bar{\mu}(S) \), and as a consequence, \( \mu(K) = \mu(K') + \sum_{i<n} \mu^*(S_i) \), i.e., \( \mu \) is \( K \)-tight. Therefore, Proposition B.18 on Caratheodory extension from a semi-ring and the previous Theorem B.25 can be combined to show a \( \sigma \)-additive set function \( \mu \) defined on a semi-ring \( S \) can be extended to an inner measure by means of the sup expression (B.3) with \( K \) replaced by \( S \). In this case \( \mu^* \leq \mu^* \) in \( 2^\Omega \), and \( \mu^* = \mu^* \) on the completion of the \( \sigma \)-algebra generated by \( S \). □

The interested reader may check the books by Halmos [63, Section III.14, pp. 58–62] and Pollard [102, Appendix A, pp. 289–300]. For instance, the books by Cohn [28], Bogachev [19] and Mattila [87] could be used for even further details.

### B.4 Examples and Convergence

First the prototype Lebesgue measure is presented and then some quick discussed on convergence in measure is necessary.

#### B.4.1 Lebesgue Measures

Three approaches for the construction of measures have been described, first the outer measure, which begins with almost no assumptions (Caratheodory’s construction Theorem B.13 and Proposition B.14), but they are really useful under the semi-ring condition of Proposition B.18. Next, the inner measure, which begins from a \( \pi \)-class (Proposition B.24 and Theorem B.25), but it is mainly used in conjunction with topological spaces. Finally, there is another more geometric approach, the so-called Hausdorff construction, which is not discussed here. Certainly, all three can be used to construct the Lebesgue measure in \( \mathbb{R}^d \).

As we have seen early, the Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}^d) \) can be generated by the class \( \mathcal{I}_d \) of all \( d \)-dimensional intervals as

\[ [a,b] = \{ x \in \mathbb{R}^d : a_i < x_i \leq b_i, \ i = 1, \ldots, d \}, \quad \forall a, b \in \mathbb{R}^d, \ a \leq b, \]

in the sense that \( a_i \leq b_i \) for every \( i \). The class \( \mathcal{I}_d \) is a semi-ring in \( \mathbb{R}^d \) and clearly, we can cover the whole space with an increasing sequence of intervals in \( \mathcal{I}_d \). Sometimes, we prefer to use a semi-algebra of \( d \)-intervals, e.g., adding the cases \( ]-\infty,b_i[ \) or \( ]a_i, +\infty[ \) for \( a_i, b_i \in \mathbb{R} \), among others, under the convention that \( 0\infty = 0 \) in the product formula below. Therefore, to define a \( \sigma \)-finite measure on \( \mathcal{B}(\mathbb{R}^d) \) (via the Caratheodory’s construction Proposition B.18) we
need only to know its (nonnegative real) values and to show that it is σ-additive on \( \mathcal{I}_d \).

**Proposition B.28.** The Lebesgue measure \( m \), defined by

\[
m([a, b]) = \prod_{i=1}^{d} (b_i - a_i), \quad \forall a, b \in \mathbb{R}^d,
\]

is σ-additive on \( \mathcal{I}_d \).

**Proof.** Using the fact that for any two intervals \([a, b]\) and \([c, d]\) in \( \mathcal{I}_d \) such that \([a, b] \cap [c, d] = \emptyset\) and \([a, b] \cup [c, d]\) belongs to \( \mathcal{I}_d \) there exists exactly one coordinate \( j \) such that \([a_j, b_j] \cup [c_j, d_j] = [a_j \wedge c_j, b_j \vee d_j]\) and \([a_i, b_i] = [c_i, d_i]\) for any \( i \neq j \), it is relatively simple to check that the above definition produces an additive measure, and to show the σ-additivity, we use the character locally compact of \( \mathbb{R}^d \). Indeed, let \( I, I_n \in \mathcal{I}_d \) be such that \( I = \sum_{n=1}^{\infty} I_n \), and for any \( \varepsilon > 0 \) define

\[
J_n = J_n(\varepsilon) = \{x \in \mathbb{R}^d : a_{n,i} < x_i \leq b_{n,i} + 2^{-n} \varepsilon\},
\]

for \( I_n = [a_n, b_n] \). It is clear that there is a constant \( c > 0 \) such that \( b_{n,i} - a_{n,i} \leq c \), for every \( n, i \), which yields the estimate

\[
0 \leq \sum_{n=1}^{\infty} (m(J_n) - m(I_n)) \leq C \varepsilon, \quad \forall \varepsilon \in (0, 1], \tag{B.7}
\]

for a suitable constant \( C = C(c, d) \) depending only on \( c \) and the dimension \( d \). Similarly, if \( I = [a, b] \) and \( I_\varepsilon = \{x \in \mathbb{R}^d : a_i + \varepsilon < x_i \leq b_i\} \), then \( m(I_\varepsilon) \rightarrow m(I) \), as \( \varepsilon \) decreases to 0.

Now, the interiors \( \{J_n(\varepsilon)\} \) constitute a sequence of open sets which cover the (compact) closure \( \overline{I}_\varepsilon \), and therefore, there exists a finite subcover, namely \( J_{n_1}(\varepsilon), \ldots, J_{n_k}(\varepsilon) \). Hence, \( J_{n_1}(\varepsilon), \ldots, J_{n_k}(\varepsilon) \) will cover \( I_\varepsilon \), and in view of the sub-additivity we deduce

\[
m(I_\varepsilon) \leq \sum_{i=1}^{k} m(J_{n_i}) \leq \sum_{n=1}^{\infty} m(J_{n}) \leq C \varepsilon + \sum_{n=1}^{\infty} m(I_n).
\]

Because \( \varepsilon > 0 \) is arbitrary, we get \( m(I) \leq \sum_{n=1}^{\infty} m(I_n) \).

Finally, since \( I \supset \sum_{n=1}^{k} I_n \), the additivity implies \( m(I) \geq \sum_{n=1}^{k} m(I_n) \), and as \( k \rightarrow \infty \) we conclude.

Usually, the measure \( m \) (or sometimes denotes by \( \ell \) or \( \ell_d \) to make explicit the dimension \( d \)) considered on the Borel σ-algebra is called Lebesgue-Borel measure and its extension (or completion) to the σ-algebra \( \mathcal{L} \) of all \( m^* \)-measurable sets is called the Lebesgue measure.

• **Remark B.29.** A direct consequences of this construction is the following list of properties for the (outer) Lebesgue measure \( (\ell^*) \ell \) on \( (\mathbb{R}^d, \mathcal{L}) \):

[Preli

\begin{align}
&\text{[Preliminary]} \quad \text{Menaldi} \quad \text{November 11, 2016}
\end{align}
(1) (a) any Borel set is measurable and that the boundary $\partial I$ of any semi-open (semi-close) $d$-interval $I$ in the semi-ring $\mathcal{I}_d$ has Lebesgue measure zero; (b) for any subset $A$ of $\mathbb{R}^d$ and any $\varepsilon > 0$ there is an open set $O$ containing $A$ such that $\ell^*(A) + \varepsilon \geq \ell(O)$, and also there is a countable intersection of open sets $G$ containing $A$ such that $\ell^*(A) = \ell(G)$.

(2) (a) any measurable set $A$ with $\ell(A) < \infty$ and any $\varepsilon > 0$ there exits an open set $O$ with $\ell(O) < \infty$ and a compact set $K$ such that $K \subset A \subset O$ and $\ell(O \setminus K) < \varepsilon$; (b) for every measurable set $A \subset \mathbb{R}^d$ and any $\varepsilon > 0$ there exits a closed set $C$ and an open set $O$ such that $C \subset A \subset O$ and $\ell(O \setminus C) < \varepsilon$. Moreover, if $\mathcal{G}_\sigma$ denotes the class of countable unions of closed sets in $\mathbb{R}^d$ and $\mathcal{G}_\delta$ denotes the class of countable intersections of open sets in $\mathbb{R}^d$ then (c) for any measurable set $A$ there exits a set $G$ in $\mathcal{G}_\delta$ and a set $F$ in $\mathcal{G}_\sigma$ such that $F \subset A \subset G$ and $\ell(G \setminus F) = 0$.

(3) If $\mathcal{I}_d$ denotes the class of open bounded $d$-intervals in $\mathbb{R}^d$ and the hyper-volume set function $m$, i.e., of the form $I = (a_1, b_1) \times \cdots \times (a_d, b_d)$, with $a_i \leq b_i$ in $\mathbb{R}$, $i = 1, \ldots, d$, and $m(I) = (b_1 - a_1) \cdots (b_d - a_d)$. Even if $\mathcal{I}_d$ is not a semi-ring, the outer measure

$$m^*(A) = \inf \left\{ \sum_{n=1}^{\infty} m(I_n) : I_n \in \mathcal{I}_d, A \subset \bigcup_{n=1}^{\infty} I_n \right\}, \quad \forall A \subset \mathbb{R}^d,$$

can certainly be defined, and Caratheodory's construction Proposition B.14 yields the same Lebesgue measure as defined by means of Proposition B.28.

Also, the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^d)$ can be generated by the class $\mathcal{K}_d$ of all $d$-dimensional compact intervals as

$$[a, b] = \{ x \in \mathbb{R}^d : a_i \leq x_i \leq b_i, \ i = 1, \ldots, d \}, \quad \forall a, b \in \mathbb{R}^d, a \leq b,$$
in the sense that $a_i \leq b_i$ for every $i$. This $\mathcal{K}_d$ is a $\pi$-class and Theorem B.25) can be applied with

$$m([a, b]) = \prod_{i=1}^{d} (b_i - a_i), \quad \forall a, b \in \mathbb{R}^d,$$

provided $m$ is additive or tight (B.5), which can be shown with argument similar to those used in Proposition B.28.

The reader interested in the Lebesgue measure on $\mathbb{R}^d$ may check the books either Gordon [56, Chapters 1 and 2, pp. 1-27] or Jones [70, Chapters 1 and 2, pp. 1-63] where a systematic approach of the Lebesgue measure and measurability is considered, in either a one dimensional or multi-dimensional settings.

From the definition of the Lebesgue measure, we can check that $m$ is invariant under translations, i.e., for a given $h \in \mathbb{R}^d$ we have that $E$ measurable implies $E + h = \{ x \in \mathbb{R}^d : x - h \in E \}$ measurable and $m(E + h) = m(E)$. We will see later that the same is true for a rotation, i.e. if $r$ is an orthogonal $d$-dimensional matrix and $E$ is measurable then $r(E) = \{ x \in \mathbb{R}^d : r(x) \in E \}$ is measurable and $m(r(E)) = m(E)$. Moreover, we have
Theorem B.30 (invariance). Let $T$ be an affine transformation from $\mathbb{R}^d$ into itself with the linear part represented by a $d$-square matrix, also denoted by $T$. Then for every $A \subset \mathbb{R}^d$ we have $m^*(T(A)) = |\det(T)|m^*(A)$, where $\det(T)$ is the determinant of the matrix $T$ and $m^*$ is the Lebesgue outer measure on $\mathbb{R}^d$.

Proof. First the translation part of the affine transformation has already been considered, so only the linear part has to be discussed. Secondly, recall that an elementary matrix $E$ produces one of the following row operations (1) interchange rows, (2) multiply a row by a non zero scalar, (3) replace a row by that row minus a multiple of another. Next, any invertible matrix can be expressed as a finite product of elementary matrices. In particular, the case of a dilation $x \mapsto rx$ with $r$ integers, and select those $d$-intervals by bisecting the edges of the original $d$-intervals, the resulting $d$-intervals have the form $[a_1/2, a_1/2 + 1/2] \times \cdots \times [a_d/2, a_d/2 + 1/2]$, with $a_i$ integers. Now, select those $d$-intervals inside $A$. By continuing this procedure, we have $A = \bigcup_{k=1}^{\infty} J_k$ where the $J_k$ are disjoint $d$-intervals and each of them is a translation of a dilation of the reference $d$-interval $J$, i.e., $J_k = t_k + r_k J$. As mentioned before, translation does not modify the measure and a $(r_k)$ dilation amplify the measure (by a factor of $|r_k|^d$), i.e., $m(J_k) = |r_k|^d m(J)$. Since $T(J_k) = T(t_k) + r_k(T(J))$, the previous argument shows that $m(T(J_k)) = |\det(T)| m(J_k)$. Hence, by the $\sigma$-additivity $m(T(A)) = m(A)$.

Finally, if $T$ is not invertible then $\det(T) = 0$ and the dimension of $T(\mathbb{R}^d)$ is strictly less than $d$. As mentioned early, any hyperplane perpendicular to any axis, e.g., $\pi = \{x \in \mathbb{R}^d : x_1 = 0\}$ has measure zero, and then for any invertible linear transformation (in particular orthogonal) $S$ we have $m(S(\pi)) = |\det(S)| m(\pi) = 0$, i.e., any hyperplane has measure zero. In particular, we have $m(T(\mathbb{R}^d)) = 0$. \qed
• **Remark B.31.** As a consequence of Theorem B.30, for any given affine transformation $T$ from $\mathbb{R}^d$ into itself, we deduce that $T(E)$ is $m^*$-measurable if and only if $E$ is $m^*$-measurable. Note that the situation is far more complicated for an affine transformation $T : \mathbb{R}^d \to \mathbb{R}^n$ and we use the Lebesgue (outer) measure $(m^*)$ $m$ on $\mathbb{R}^d$ and $\mathbb{R}^n$, with $d \neq n$, see later sections on Hausdorff measure. □

• **Remark B.32.** Recall that the diameter of a set $A$ in Euclidean space $\mathbb{R}^d$ is defined as $d(A) = \sup \{|x - y| : x, y \in A\}$. If a set $A$ is contained in a ball of diameter $d(A)$ then the monotony of the Lebesgue outer measure $m^*$ in $\mathbb{R}^d$ implies

$$m^*(A) \leq c_d (d(A)/2)^d, \quad \forall A \subset \mathbb{R}^d,$$

where $c_d$ is the volume of unit ball in $\mathbb{R}^d$, calculated later as

$$c_d = \pi^{-d/2} \Gamma(d/2 + 1), \quad \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt,$$

with $\Gamma(\cdot)$ is the Gamma function. Certainly, any set $A$ with diameter $d(A)$ is contained in a ball of radius $d(A)$, which yields the estimate $m^*(A) \leq c_d (d(A))^d$. However, an equilateral triangle $T$ in $\mathbb{R}^2$ is not contained in a ball of radius $d(A)/2$. For instance, a carefully discussion on the isodiametric inequality (B.9) can be found in Evans and Gariepy [43, Theorem 2.2.1, pp. 69-70] or in Stroock [118, Section 4.2, pp. 74-79]. □

### B.4.2 Convergence in Measure

For functions from a measure space into a topological space we may think of various modes of convergence. For instance, (1) $f_n \to f$ pointwise a.e. (almost everywhere) if there exists a set $N \in \mathcal{F}$ with $\mu(N) = 0$ such that $f(x) \to f(x)$ for every $x \in \Omega \setminus N$; or (2) $f_n \to f$ pointwise quasi-uniform (quasi-uniformly) if for every $\varepsilon > 0$ there exists a set $\Omega_\varepsilon \in \mathcal{F}$ with $\mu(\Omega \setminus \Omega_\varepsilon) \leq \varepsilon$ such that $f_n(x) \to f(x)$ uniformly in $\Omega_\varepsilon$. It is clear that (2) implies (1) and the converse is not necessarily true. Also we have

**Definition B.33.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(E, d)$ be a metric space. A sequence $\{f_n\}$, $f_n : \Omega \to E$, of measurable functions is a *Cauchy sequence in measure* (or *in probability* if $\mu(\Omega) = 1$) if for every $\varepsilon > 0$ there exists $n(\varepsilon)$ such that $\mu(\{x \in \Omega : d(f_n(x), f_m(x)) \geq \varepsilon\}) < \varepsilon$ for every $n, m \geq n(\varepsilon)$. Similarly, $f_n \to f$ in measure, if for every $\varepsilon > 0$ there exists $n(\varepsilon)$ such that $\mu(\{x \in \Omega : d(f_n(x), f(x)) \geq \varepsilon\}) < \varepsilon$ for every $n \geq n(\varepsilon)$. □

Note that we may use the distance $d(x, y) = |\arctan(x) - \arctan(y)|$, for any $x, y$ in $E = [-\infty, +\infty]$, when working with extended-valued measurable functions, i.e., the mapping $z \mapsto \arctan z$ transforms the problem into real-valued functions. It is clear that for any sequence $\{x_n\}$ of real numbers we have $x_n \to x$ if and only if $\arctan(x_n) \to \arctan(x)$, but the usual distance $(x, y) \to |x - y|$ and $d(x, y)$ are not equivalent in $\mathbb{R}$. Actually, consider the
sequence \( \{ f_n(x) = (x + 1/n)^2 \} \) on Lebesgue measure space \((\mathbb{R}, \mathcal{L}, \ell)\) and the limiting function \( f(x) = x^2 \) to check that

\[
\ell(\{ x \in \mathbb{R} : |f_n(x) - f(x)| \geq \varepsilon \}) = \ell(\{ x \in \mathbb{R} : |x + (1/n)| \geq n\varepsilon \}) = \infty,
\]

for every \( \varepsilon > 0 \) and \( n \geq 1 \), i.e., \( f_n \) does not converge in measure to \( f \). However, if \( g_n(x) = \arctan(f_n(x)) \) and \( g(x) = \arctan(x^2) \) then \( |g_n(x) - g(x)| \leq 1/n \), i.e., \( g_n \) converges to \( g \) uniformly in \( \mathbb{R} \). Thus, on the Lebesgue measure space \((\mathbb{R}, \mathcal{L}, \ell)\), we have \( g_n \to g \) in measure, i.e., the convergence in measure depends not only on the topology given to \( \mathbb{R} \), but actually, on the metric used on it.

**Remark B.34.** It is simple to verify that if the sequence \( \{ f_n \} \), \( f_n : \Omega \to E \), of measurable functions is convergent (or Cauchy) in measure, \((Z,d_Z)\) is a metric space and \( \psi : E \to Z \) is a uniformly continuous function then the sequence \( \{ g_n \} \), \( g_n(x) = \psi(f_n(x)) \) is also convergent (or Cauchy) in measure. Thus, in particular, if \((E,|\cdot|_E)\) is a normed space then for any sequences \( \{ f_n \} \) and \( \{ g_n \} \) of \( E \)-valued measurable functions and any constants \( a \) and \( b \) we have \( af_n + bg_n \to af + bg \) in measure, whenever \( f_n \to f \) and \( g_n \to g \) in measure. Moreover, assuming that the sequence \( \{ g_n \} \) takes real (or complex) values, (a) if the sequences are also quasi-uniformly bounded, i.e., for any \( \varepsilon > 0 \) there exists a measurable set \( F \) with \( \mu(F) < \varepsilon \) such that the numerical series \( \{ |f_n(x)|_E \} \) and \( \{ |g_n(x)|_E \} \) are uniformly bounded for \( x \) in \( F^c \), then deduce that \( f_n g_n \to fg \) in measure. Furthermore, (b) if \( g_n(x)g(x) \neq 0 \) a.e. \( x \) and the sequences \( \{ f_n \} \) and \( \{ 1/g_n \} \) are also quasi-uniformly bounded then show that \( f_n / g_n \to f/g \) in measure. Finally, (c) verify that if the measure space \( \Omega \) has finite measure then the conditions on quasi-uniformly bounded are automatically satisfied.

For the particular case when \( E = \mathbb{R}^d \) the convergence in measure means

\[
\lim_n \mu(\{ x \in \Omega : |f_n(x) - f(x)| \geq \varepsilon \}) = 0, \quad \forall \varepsilon > 0,
\]

and if \( f_n(x) = 1_{\{|x| > n\}} \) then \( f_n(x) \to 0 \) for every \( x \) in \( \mathbb{R}^d \), but \( \ell(\{ x \in \mathbb{R}^d : |f_n(x)| \geq \varepsilon \}) = \infty \), with the Lebesgue measure \( \ell \), i.e., the pointwise almost everywhere convergence does not necessarily yields the convergence in measure. However, we have

**Theorem B.35.** Let \((\Omega,F,\mu)\) be a measure space, \( E \) be a complete metric space and \( \{ f_n \} \) be a Cauchy sequence in measure of measurable functions \( f_n : \Omega \to E \). Then there exist (1) a subsequence \( \{ f_{n_k} \} \) such that \( f_{n_k} \to f \) pointwise a.e. and (2) a measurable function \( f \) such that \( f_n \to f \) in measure. Moreover, if \( f_n \to g \) in measure then \( g = f \) a.e.

**Proof.** Given \( \varepsilon > 0 \) define \( X(\varepsilon,n,m) = \{ x \in \Omega : d(f_n(x),f_m(x)) \geq \varepsilon \} \) to see that for \( \varepsilon = 2^{-1} > 0 \) we can find \( n_1 \) such that \( \mu(X(\varepsilon,n_1,m)) < \varepsilon \) for every \( m \geq n_1 \). Next, for \( \varepsilon = 2^{-2} > 0 \) again, we can find \( n_2 > n_1 \) such that \( \mu(X(\varepsilon,n_2,m)) < \varepsilon \) for every \( m \geq n_2 \). By induction, we get \( n_k < n_{k+1} \) and \( A_k = X(2^{-k},n_k,n_{k+1}) \) with \( \mu(A_k) < 2^{-k} \), for every \( k \geq 1 \).
Now, if \( F_k = \bigcup_{i=k}^{\infty} A_i \) then \( \mu(F_k) \leq \sum_{i=k}^{\infty} 2^{-i} = 2^{1-k} \). On the other hand, if \( x \not\in F_k \) then for any \( i \geq j \geq k \) we have
\[
d(f_{n_j}(x), f_{n_i}(x)) \leq \sum_{r=j}^{i-1} d(f_{n_{r+1}}(x), f_{n_r}(x)) \leq \sum_{r=j}^{i-1} 2^{-r} \leq 2^{1-k},
\] (B.10)
i.e., \( \{ f_{n_i}(x) \} \) is a Cauchy sequence in \( E \), for every \( x \not\in F_k \).

Define \( F = \bigcap_k F_k \) to have \( \mu(F) \leq \mu(F_k) \), for every \( k \), i.e., \( \mu(F) = 0 \). If \( x \not\in F \) then \( x \) belongs to a finite number of \( F_k \) and therefore, because \( E \) is complete, there exists the limit of \( \{ f_{n_k}(x) \} \), which is called \( f(x) \). If \( x \in F \) we set \( f(x) = 0 \). Hence \( f_{n_k} \to f \) almost everywhere.

Let \( i \to \infty \) in (B.10) to have \( d(f_{n_k}(x), f(x)) \leq 2^{1-k} \) for every \( x \not\in F_k \). Since \( \mu(F_k) \leq 2^{1-k} \to 0 \), we deduce that \( f_{n_k} \to f \) in measure, and in view of the inclusion
\[
\{ x : d(f_n(x), f(x)) \geq \varepsilon \} \subseteq \{ x : d(f_n(x), f_{n_k}(x)) \geq \varepsilon/2 \} \cup \{ x : d(f_{n_k}(x), f(x)) \geq \varepsilon/2 \}, \quad \forall \varepsilon > 0,
\]
the whole sequence \( f_n \to f \) in measure. Moreover, in view of
\[
\{ x : d(f(x), g(x)) \geq \varepsilon \} \subseteq \{ x : d(f_n(x), g(x)) \geq \varepsilon/2 \} \cup \{ x : d(f_n(x), f(x)) \geq \varepsilon/2 \}, \quad \forall \varepsilon > 0,
\]
if \( f_n \to g \) in measure then \( f = g \) a.e.

\[ \square \]

Remark B.36. In a measure space \((\Omega, \mathcal{F}, \mu)\), take a measurable set \( A \in \mathcal{F} \) with \( 0 < \mu(A) \leq 1 \) and find a finite partition \( A = \bigcup_{i=1}^{k} A_{k,i} \) with \( 0 < \mu(A_{k,i}) \leq 1/k \), for every \( i \). If \( \{a_k\} \) and \( \{b_k\} \) are two sequences of real numbers then we construct a sequence of functions \( \{f_n\} \) as follows: the sequence of integers \( \{1, 2, 3, \ldots, 10, 11, \ldots\} \) is grouped as \( \{(1); (2, 3); (4, 5, 6); (7, 8, 9, 10); \ldots\} \) where the \( k \) group has exactly \( k \) elements, i.e., for any \( n = 1, 2, \ldots \), we select first \( k = 1, 2, \ldots \), such that \( (k - 1)k/2 < n \leq k(k + 1)/2 \) and we write (uniquely) \( n = (k - 1)k/2 + i \) with \( i = 1, 2, \ldots, k \) to define
\[
f_n(x) = \begin{cases} a_k & \text{if } x \in A \setminus A_{k,i}, \\ b_k & \text{if } x \in A_{k,i}. \end{cases}
\]
Assuming that \( a_k \to a \) as \( k \to \infty \) and \( |b_k - a| \geq c > 0 \) for any \( k \), we have \( \mu(\{|f_n - a| \geq \varepsilon\}) = \mu(A_{k,i}) \leq 1/k \leq 2/\sqrt{n} \) for every \( 0 < \varepsilon < c \), i.e., \( f_n \to f \) in measure with \( f(x) = a \) for every \( x \). However, for every \( x \in A \) there exist \( i, k \) such that \( x \in A_{k,i} \) and \( f_n(x) = b_k \), i.e., \( f_n(x) \) does not converge to \( f(x) \). Moreover, for any given \( b \leq a \leq \overline{b} \), we can choose \( b_k \) so that \( \liminf_n f_n(x) = b \) and \( \limsup_n f_n(x) = \overline{b} \), for every \( x \in A \).

\[ \square \]

Sometimes we begin with a known notion of convergence to define closed sets in a space \( X \). For instance, if we know that the “convergence \( x_n \to x \)” satisfies the following (Kuratowski) three axioms (1) uniqueness of the limit;
(2) for every $x$ in $X$, the constant sequence $\{x, x, \ldots\}$ converges to $x$; (3) given a sequence $\{x_n\}$ convergent to $x$, every subsequence $\{x_{n'}\} \subseteq \{x_n\}$ converges to the same limit $x$; then we can define the open sets in the topology $T$ as the complement of closed set, where a set $C$ is closed if for any sequence $\{x_n\}$ of point in $C$ such that $x_n \to x$ results $x$ in $C$. Next, knowing the topology $T$ we have the “convergence $x_n \xrightarrow{T} x$,” i.e., for any open set $O$ (element in $T$) with $x \in O$ there exists an index $N$ such that $x_n \in O$ for any $n \geq N$. Actually, this means that $x_n \xrightarrow{T} x$ if and only if for any subsequence $\{x_{n'}\}$ of $\{x_n\}$ there exists another subsequence $\{x_{n''}\} \subseteq \{x_{n'}\}$ such that $x_{n''} \to x$. Clearly, if $x_n \to x$ then $x_n \xrightarrow{T} x$. If the initial convergence $x_n \to x$ comes from a metric, then we can verify that $x_n \to x$ is equivalent to $x_n \xrightarrow{T} x$, but, in general, this could be false. For instance, let $(\Omega, \mathcal{F}, \mu)$ be a measure space with $\mu(\Omega) < \infty$, and consider the space $X$ of real-valued measurable functions (actually, equivalent classes of functions because we have identified functions almost everywhere equal), with the almost everywhere convergence $x_n(\omega) \to x(\omega)$ a.e. $\omega$. By means of Theorem B.35 we see that $x_n \xrightarrow{T} x$ if and only if $x_n \to x$ in measure.

**Remark B.37.** Assume that $(\Omega, \mathcal{F}, \mu)$ is a measure space, $(E, d)$ a metric space and $\{f_n\}$ a sequence of measurable functions $f_n : \Omega \to E$. It is relatively simple to show that if $\{f_n\}$ converges to some function $f$ pointwise quasi-uniform then $f_n \to f$ in measure. □

Recall the definition of Borel (outer) measure $\mu$ (e.g., the Lebesgue measure): for every set $F$ with finite outer measure $\mu^*(F) < \infty$ and any constant $\varepsilon > 0$ there exists an open set $O \supset F$ with $\mu(O \setminus F) < \varepsilon$. Now, let us compare the pointwise almost everywhere convergence with the pointwise uniform convergence and the convergence in measure. We have

**Theorem B.38 (Egorov).** If $\mu(\Omega) < \infty$ then pointwise almost everywhere convergence implies pointwise quasi-uniform convergence, i.e., if a sequence $\{f_n\}$ of measurable functions taking values in a metric space $(E, d)$ satisfies $f_n(x) \to f(x)$ a.e. in $x$, then for every $\varepsilon > 0$ there exists an index $n_\varepsilon$ and a set $F \in \mathcal{F}$ with $\mu(F) < \varepsilon$ such that $d(f_n(x), f(x)) < \varepsilon$ for every $n \geq n_\varepsilon$ and $x \in F^c = \Omega \setminus F$. Moreover, if $\mu$ is a Borel measure then $F = O$ is an open set of $\Omega$.

**Proof.** Even if this is not necessary, we first prove that assuming a finite measure, pointwise almost everywhere convergence implies convergence in measure. Indeed, given a sequence $\{f_n\}$ and a function $f$, define $X(\varepsilon, f_n, f) = \{x \in \Omega : d(f_n(x), f(x)) > \varepsilon\}$ to check that $f_n(x) \to f(x)$ if and only if $x \not\in F_\varepsilon = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} X(\varepsilon, f_k, f)$ for every $\varepsilon > 0$. Since $X(\varepsilon, f_n, f) \subset F_{\varepsilon,n} = \bigcap_{k=1}^{n} \bigcup_{i=k}^{\infty} X(\varepsilon, f_i, f)$, we have $\mu(X(\varepsilon, f_n, f)) \leq \mu(F_{\varepsilon,n})$, and therefore

$$\limsup_n \mu(X(\varepsilon, f_n, f)) \leq \lim_n \mu(F_{\varepsilon,n}), \quad \forall \varepsilon > 0.$$ 

If $f_n \to f$ pointwise almost everywhere then $\mu(F_\varepsilon) = 0$ for every $\varepsilon > 0$, and if also $\mu$ is a finite measure then $\mu(F_{\varepsilon,n}) \to \mu(F_\varepsilon) = 0$. 
To show the quasi-uniform convergence, let \( k, n \) be positive integers and set

\[
A_k(n) = \bigcup_{m=n}^{\infty} \{ x : d(f_m(x), f(x)) \geq 1/k \} = \bigcup_{m=n}^{\infty} X(1/k, f_m, f).
\]

It is clear that \( A_k(n) \supset A_k(n+1) \) for any \( k, n \), and the almost everywhere convergence implies that \( \mu(B_k) = 0 \) with \( \bigcap_{n=1}^{\infty} A_k(n) = B_k \). Since \( \mu(\Omega) < \infty \) we deduce \( \mu(A_k(n)) \to 0 \) as \( n \to \infty \). Hence, given \( \varepsilon > 0 \) and \( k \), choose \( n_k \) such that \( \mu(A_k(n_k)) < \varepsilon 2^{-k} \) and define \( F = \bigcup_{k=1}^{\infty} A_k(n_k) \). Thus \( \mu(F) < \varepsilon \), and \( d(f_n(x), f(x)) < 1/k \) for any \( n > n_k \) and \( x \notin F \). This yields \( f_n \to f \) uniformly on \( F^c \).

Finally, if \( \mu \) is a Borel measure then we conclude by choosing an open set \( O \supset F \) with \( \mu(O) < 2\varepsilon \).

As mentioned early, if the measure is not finite then pointwise almost everywhere convergence does not necessarily implies convergence in measure. The converse is also false. It should be clear (see Remark B.37) that quasi-uniform convergence implies the convergence in measure, so that Theorem B.38 also affirms that if the space has finite measure then pointwise almost everywhere convergence implies convergence in measure.

- **Remark** B.39. Another consequence of Egorov Theorem B.38 is the approximation of any measurable function by a sequence of continuous functions. Indeed, if \( \mu \) is a finite Borel measure on \( \Omega \) and \( f \) is \( \mu^* \)-measurable function with values in \( \mathbb{R}^d \) then there exists a sequence \( \{f_n\} \) of continuous functions such that \( f_n \to f \) almost everywhere, see Doob [34, Section V.16, pp. 70-71].

**B.4.3 Almost Measurable Functions**

For a given measure space \((\Omega, \mathcal{F}, \mu)\), we denote by \( \mathcal{L}^0 = \mathcal{L}^0(\Omega, \mathcal{F}; E) \) the space of measurable functions \( f : \Omega \to E \), where \( E \) is a measurable space. However, once a measure \( \mu \) is defined on \( \mathcal{F} \) and a measure space \((\Omega, \mathcal{F}; \mu)\) is constructed, we may complete the \( \sigma \)-algebra \( \mathcal{F} \) to get a complete measure space \((\Omega, \mathcal{F}^\mu, \mu)\) and to make use of \( \mathcal{L}^0(\Omega, \mathcal{F}^\mu; E) \), also denoted by \( \mathcal{L}^0(\Omega, \mu; E) \), instead of \( \mathcal{L}^0(\Omega, \mathcal{F}; E) \). If \( E \) is a vector space, to check that \( \mathcal{L}^0(\Omega, \mathcal{F}; E) \) is indeed a vector space we need to know that the sum and the scalar multiplication on \( E \) are Borel (or continuous) operations, e.g., when \( E \) is topological vector space or when \( E \) is separable metric space of real (or complex) functions.

Recall that the abbreviation a.e. means almost everywhere, i.e., there exists a set \( N \) (which can be assumed to be \( \mathcal{F} \)-measurable even if \( \mathcal{F} \) is not \( \mu \)-complete) such that the equality (or in general, the property stated) holds for any point \( \omega \) in \( \Omega \setminus N \). Thus, assuming that \( \mathcal{F} \) is complete with respect to \( \mu \) we have: (a) if \( f \) is measurable and \( f = g \) a.e. then \( g \) is also measurable; (b) if \( \{f_n\} \) are measurable and \( f_n \to f \) a.e. then \( f \) is also measurable. If \( \mathcal{F} \) is not necessarily \( \mu \)-complete then a function \( f \) measurable with respect to \( \mathcal{F}^\mu \), the \( \mu \)-completion of \( \mathcal{F} \), is called \( \mu \)-measurable. Now, if \( \varphi \) is a \( \mu \)-measurable simple function then by the definition of the completion \( \mathcal{F}^\mu \) there exists another \( \mathcal{F} \)-measurable simple
function $\psi$ such that $\varphi = \psi$ a.e., and since any measurable function is a pointwise limit of a sequence of simple functions, we conclude that for every $\mu$-measurable function $f$ there exists a $\mathcal{F}$-measurable function $g$ such that $f = g$ a.e.

Therefore, we are interested to study measurable functions defined (almost everywhere) outside of an unknown set of measure zero, i.e., $f : \Omega \setminus N \to E$ measurable with $\mu(N) = 0$. To go further in this analysis, we use $E = \mathbb{R}^n$, $n \geq 1$ or $\mathbb{R} = [-\infty, +\infty]$, or in general a (complete) metric (or Banach) space $E$ with its Borel $\sigma$-algebra $\mathcal{E}$. Clearly, the case $E = \mathbb{R}^n$, $n \geq 1$ is of main interest, as well as when $E$ is an infinite dimensional Banach space.

We endow $L^0(\Omega, \mathcal{F}; E)$ with the topology induced by convergence in measure. This topology does not separate points, so to have a Hausdorff space we are forced to consider equivalence class of functions under the relation $f \sim g$ if and only if there exists a set $N \in \mathcal{F}$ with $\mu(N) = 0$ and $f(\omega) = g(\omega)$ for every $\omega \in \Omega \setminus N$. Thus, the quotient space $L^0 = L^0/\sim$ or $L^0(\Omega, \mathcal{F}, \mu; E)$ becomes a Hausdorff topological space with the convergence in measure. Actually, we regard the elements of $L^0$ as measurable functions defined almost everywhere, so that even if $L^0(\Omega, \mathcal{F}; E)$ may not be equal to $L^0(\Omega, \mathcal{F}_\mu; E)$, we are really looking at $L^0 = L^0(\Omega, \mathcal{F}_\mu, \mu; E) = L^0(\Omega, \mu; E)$. Note that for the quotient space $L^0$ (where the elements are equivalence classes) we may omit the $\sigma$-algebra $\mathcal{F}$ from the notation, while for the initial space $L^0$ we may use the whole measure space $(\Omega, \mathcal{F}, \mu)$. Note that if $\Omega_0$ is a measurable subset in a measure space $(\Omega, \mathcal{F}, \mu)$ then we may define the restriction to $\Omega_0$, of $\mathcal{F}$ and $\mu$ to form the measure space $(\Omega_0, \mathcal{F}_0, \mu_0)$, and for instance, we may talk about functions measurable on $\Omega_0$.

**Definition B.40.** When the space $E$ is not separable, we need to modify the concept of measurability as follows: on a measure space $(\Omega, \mathcal{F}, \mu)$ a function with values in a Borel space $(E, \mathcal{E})$ is called measurable if (a) $f^{-1}(B)$ belongs to $\mathcal{F}$ for every $B$ in $\mathcal{E}$ and (b) $f(\Omega)$ is contained in a separable subspace of $E$. Also, functions measurable with respect to the completion $\mathcal{F}_\mu$ are called $\mu$-measurable. An equivalence class of $\mu$-measurable functions is called an almost measurable function, which is considered defined only almost everywhere, i.e., a function whose restriction to the complement of a null set is a measurable function. This space $L^0(\Omega, \mathcal{F}, \mu; E) = L^0(\Omega, \mathcal{F}_\mu, \mu; E)$ of $E$-valued measurable functions defined almost everywhere is denoted by $L^0(\Omega, \mu; E)$ and by $L^0$, when the meaning is clear from the context. Certainly, “equality” in $L^0$ means $\mu$-almost everywhere pointwise equality.

In most of the cases, $E$ is a metric space and $\mathcal{E}$ is its Borel $\sigma$-algebra. The imposition of a separable range $f(\Omega)$ is rather technical, but necessary most of the time. Most of the time, we have in mind the typical case of $E$ being a Polish space (mainly, the extended $\mathbb{R}^d$), so that this condition is always satisfied.

**Proposition B.41.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(E, d_E)$ be a metric space. If $f$ and $g$ are two almost measurable functions from $\Omega$ into $E$, we define

$$d_\mu(f, g) = \inf \{ r > 0 : \mu(\{ \omega \in \Omega : d_E(f(\omega), g(\omega)) > r \}) \leq r \}.$$
Then (1) the map \((f,g) \rightarrow d_\mu(f,g)\) is a metric on \(L^0 = L^0(\Omega, \mathcal{F}, \mu; E)\); (2) on has \(d_\mu(f_n, f) \rightarrow 0\) if and only if \(f_n \rightarrow f\) in measure; (3) the metric \(d_\mu\) is complete in \(L^0\) whenever \(d_E\) is complete in \(E\).

**Proof.** Note that to have \(d_\mu(f,g)\) fully define, we should contemplate the possibility of having \(\mu(\{\omega \in \Omega : d_E(f(\omega), g(\omega)) > r\}) = \infty\) for every \(r > 0\), in this case, we define \(d_\mu(f,g) = \infty\). Thus, to make a proper distance we could replace \(d_\mu(f,g)\) with \(d_\mu(f,g) \vee 1\), or equivalently re-define

\[
d_\mu(f,g) = \inf \{r \in (0,1] : \mu(\{\omega \in \Omega : d_E(f(\omega), g(\omega)) > r\}) \leq r\}
\]

with the understanding that \(\inf\{\emptyset\} = 1\).

First, we can check that \(d_\mu\) satisfies the triangular inequality and becomes a metric (or distance) in \(L^0\). Now, by definition, there exists a decreasing sequence \(r_n = r_n(f,g)\) such that \(r_n \rightarrow d_\mu(f,g)\) and \(\mu(\{\omega \in \Omega : d_E(f(\omega), g(\omega)) > r_n\}) \leq r_n\), the monotone continuity from below of the measure \(\mu\) shows that

\[
\mu(\{\omega \in \Omega : d_E(f(\omega), g(\omega)) > d_\mu(f,g)\}) \leq d_\mu(f,g),
\]

i.e., convergence in measure is given as the convergence in the metric \(d_\mu\). Finally, we conclude by applying Theorem B.35. \(\square\)

Consider \(S^0 = S^0(\Omega, \mathcal{F}; E) \subset L^0\) and \(S^0 = S^0(\Omega, \mu; E) \subset L^0\), the subspaces of all simple functions, (i.e., measurable functions assuming only a finite number of values). We may also consider \(S^0(\Omega, \mathcal{F}^\mu; E)\) if needed. Clearly, \(S^0\) is not closed (nor complete) in \(L^0\). For instance, if \(E\) is a separable metric space then for any element \(f\) in \(L^0(\Omega, \mu; E)\) there exists a sequence \(\{f_n\} \subset L^0(\Omega, \mu; E)\) and a null set \(N\) such that \(f_n\) is a measurable function assuming only a finite number of values (i.e., \(f_n\) is an almost everywhere simple function), and \(d_E(f_n(\omega), f(\omega))\) decreases to 0 as \(n \rightarrow \infty\) for every \(x \in \Omega \setminus N\). Hence, if \(\mu(\Omega) < \infty\) then \(f_n \rightarrow f\) in measure, i.e., \(d_\mu(f_n, f) \rightarrow 0\) as \(n \rightarrow \infty\).

Because it is desirable to approximate any function in \(L^0\) by a sequence of function in \(S^0\), we have modified a little the definition of measurable functions when \(E\) is not separable, by adding almost separability of the range. Moreover, the topology in \(L^0\) should be slightly modified, i.e., convergence in measure on every set of finite measure.

Even when the (complete) metric space \(E\) and the \(\sigma\)-algebra \(\mathcal{F}\) are separable, the separability of the (complete) metric space \(L^0\) is an issue, because some property of the measure \(\mu\) are also involved.

If \(E\) is a Banach space (i.e., complete normed space) with norm \(|\cdot|_E\) then the function

\[
d_\mu(f,0) = \inf \{r > 0 : \mu(\{\omega \in \Omega : |f(\omega)|_E > r\}) \leq r\}
\]

is not necessarily homogeneous, for instance if \(f = 1_F\) with \(F \in \mathcal{F}\) then \(d(cf,0) = c \wedge \mu(F)\), for every \(c \geq 0\). Nevertheless, \(d_\mu(cf,0) \leq (1 \vee |c|)d_\mu(f,0)\) and therefore \(cf \rightarrow 0\) if \(f \rightarrow 0\). Moreover, if \(\mu(\{\omega \in \Omega : f(\omega) \neq 0\}) < \infty\) then for every \(\varepsilon > 0\) there exist \(\delta > 0\) such that \(\mu(\{\omega \in \Omega : d_\mu(f(\omega), 0) > 1/\delta\}) < \varepsilon\)
and therefore \( d_\mu(cf, 0) \leq \varepsilon \) whenever \( |c| < \varepsilon \delta \). Thus, besides \( L^0(\Omega, \mu; E) \) being a complete metric space, it is not quite a topological vector space, i.e., the vector addition is continuous but the scalar multiplication is continuous only on functions vanishing outside of a set of finite measure.

If \( E = \mathbb{R} \) then \( L^1(\Omega, \mathcal{F}, \mu) = L^1(\Omega, \mathcal{F}, \mu; \mathbb{R}) \) is the vector space of real-valued integrable functions, where the expression

\[
\|f\|_1 = \int_\Omega |f| \, d\mu
\]

defines a semi-norm, i.e., we need to consider equivalence class of functions and consider the quotient space \( L^1(\Omega, \mu) \) as a subspace of \( L^0(\Omega, \mu) \), and \( \| \cdot \|_1 \) becomes a norm on \( L^1(\Omega, \mu) \). It is simple to verify that \( L^1(\Omega, \mu) \) is a closed subspace of the complete space \( L^0(\Omega, \mu) \), therefore \( L^1(\Omega, \mu) \) results a Banach space. Note that if \( \mathcal{R} = [-\infty, +\infty] \) then \( L^0(\Omega, \mu; \mathcal{R}) \) is not necessarily equal to \( L^0(\Omega, \mu; \mathbb{R}) \), but, since any integrable function is finite almost everywhere, we do have \( L^1(\Omega, \mu; \mathcal{R}) = L^1(\Omega, \mu; \mathbb{R}) \).

B.5 Integration Theory

Recall that a simple function \( \varphi : \Omega \to \mathbb{R} \) is a measurable functions assuming a finite number of values, i.e., a linear (finite with real coefficients) combination of characteristic functions. Any simple function has a standard represented as \( \varphi(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{E_i}(x) \), with \( a_i \neq a_j \) for \( i \neq j \) and \( \{E_i\} \) a finite sequence of disjoint measurable sets. Denote by \( S = S(\Omega, \mathcal{F}) \) the set of all simple functions on a measurable space \((\Omega, \mathcal{F})\). Clearly \( S \) is stable under the addition, multiplication, \( \max (\lor) \) and \( \min (\land) \), i.e., if \( \varphi, \psi \in S \) and \( a, b \in \mathbb{R} \) then \( a \varphi + b \psi, \varphi \lor \psi, \varphi \land \psi \in S \). Also, we have seen in Theorem B.9, that simple functions can be used to approximate pointwise any measurable function.

Once a measure space \((\Omega, \mathcal{F}, \mu)\) has been given, it is clear that for any measurable set \( F \) we should assign the value \( \mu(F) \) as the integral of the characteristic function \( \mathbb{1}_F \). Then, by imposing linearity, for a simple function \( \varphi(x) \) we should have

\[
\int_\Omega \varphi(x) \mu(dx) = \sum_{i=1}^{n} a_i \mu(\varphi^{-1}(\{a_i\})) \text{,}
\]

under the convention that the sum is possible, i.e., we set \( a \times (+\infty) = 0 \) if \( a = 0 \), \( a \times (+\infty) = +\infty \) if \( a > 0 \), \( a \times (-\infty) = -\infty \) if \( a < 0 \), and the case \( +\infty \land -\infty \) is forbidden. Hence, we can approximate any nonnegative measurable function \( f \) by an increasing sequence of simple functions to have

\[
\lim_n \int_\Omega f \mathbb{1}_{f < 2^n} \, d\mu = \lim_n \sum_{k=0}^{2^n-1} k2^{-n} \mu(f^{-1}([k2^{-n}, (k + 1)2^{-n}])) \text{,}
\]

which is always meaningful, and then writing \( f = f^+ - f^- \) we treat the general case. Details of these arguments follow.
B.5.1 Definition and Properties

Let \((\Omega, \mathcal{F}, \mu)\) be a measurable space. If \(\varphi : \Omega \to [0, \infty)\) is a simple function with standard represented as \(\varphi(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{E_i}(x)\), with \(a_i \neq a_j\) for \(i \neq j\) and \(\{E_i\}\) a finite sequence of disjoint measurable sets, then we define the integral of \(\varphi\) over \(\Omega\) with respect to the measure \(\mu\) as

\[
\int_{\Omega} \varphi \, d\mu = \int_{\Omega} \varphi(\omega) \mu(d\omega) = \int_{\Omega} \varphi(\omega) \mu(\omega) = \sum_{i=1}^{n} a_i \mu(F_i),
\]

under the only convention \(0 \times (+\infty) = 0\), since \(\varphi \geq 0\).

**Proposition B.42.** If \(\varphi\) and \(\psi\) are nonnegative simple functions then

(a) \(\int_{\Omega} c \varphi \, d\mu = c \int_{\Omega} \varphi \, d\mu, \quad \forall c \geq 0\),

(b) \(\int_{\Omega} (\varphi + \psi) \, d\mu = \int_{\Omega} \varphi \, d\mu + \int_{\Omega} \psi \, d\mu\),

(c) if \(\varphi \leq \psi\), then \(\int_{\Omega} \varphi \, d\mu \leq \int_{\Omega} \psi \, d\mu\) (monotony),

(d) the function \(A \to \int_{A} \varphi \, d\mu\) is a measure on \(\mathcal{F}\).

**Proof.** The property (a) follows directly from the definition of the integral.

To check the identity (b) take standard representations \(\varphi = \sum_{i=1}^{n} a_i \mathbb{1}_{F_i}\) and \(\psi = \sum_{j=1}^{m} b_j \mathbb{1}_{G_j}\). Since \(F_i = \bigcup_{j=1}^{m} F_i \cap G_j\) and \(G_j = \bigcup_{i=1}^{n} F_i \cap G_j\), both disjoint unions, the finite additivity of \(\mu\) implies

\[
\int_{\Omega} (\varphi + \psi) \, d\mu = \int_{\Omega} \sum_{j=1}^{m} \sum_{i=1}^{n} (a_i + b_j) \mathbb{1}_{F_i \cap G_j} \, d\mu =
\]

\[
= \sum_{j=1}^{m} \sum_{i=1}^{n} (a_i + b_j) \mu(F_i \cap G_j) =
\]

\[
= \sum_{j=1}^{m} \sum_{i=1}^{n} a_i \mu(F_i \cap G_j) + \sum_{j=1}^{m} \sum_{i=1}^{n} b_j \mu(F_i \cap G_j) =
\]

\[
= \sum_{i=1}^{n} a_i \mu(F_i) + \sum_{j=1}^{m} b_j \mu(G_j) = \int_{\Omega} \varphi \, d\mu + \int_{\Omega} \psi \, d\mu.
\]

as desired.

To show (c), if \(\varphi \leq \psi\) then \(a_i \leq b_j\) each time that \(F_i \cap F_j \neq \emptyset\), hence

\[
\int_{\Omega} \varphi \, d\mu = \sum_{j=1}^{m} \sum_{i=1}^{n} a_i \mu(F_i \cap G_j) \leq \sum_{j=1}^{m} \sum_{i=1}^{n} b_j \mu(F_i \cap G_j) = \int_{\Omega} \psi \, d\mu.
\]
For (d), we have to prove only the countable additivity. If \( \{A_j\} \) are disjoint and \( A = \bigcup_{j=1}^{\infty} A_j \) then

\[
\int_A \varphi \, d\mu = \sum_{i=1}^{n} a_i \mu(A \cap F_i) = \sum_{i=1}^{\infty} \sum_{j=1}^{n} a_i \mu(A_j \cap F_i) = \sum_{j=1}^{\infty} \sum_{i=1}^{n} a_i \mu(A_j \cap F_i) = \sum_{j=1}^{\infty} \int_{A_j} \varphi \, d\mu,
\]

and we conclude.

\[\square\]

**Definition B.43.** If \( f \) is a nonnegative measurable function then we define the integral of \( f \) of \( \Omega \) with respect to \( \mu \) as

\[
\int_{\Omega} f \, d\mu = \sup \left\{ \int_{\Omega} \varphi \, d\mu : \varphi \text{ simple}, \, 0 \leq \varphi \leq f \right\},
\]

which is nonnegative and perhaps \(+\infty\). If \( f \) is a measurable function with valued in \([-\infty, +\infty] \), writing \( f = f^+ - f^- \), then

\[
\int_{\Omega} f \, d\mu = \int_{\Omega} f^+ \, d\mu - \int_{\Omega} f^- \, d\mu < \infty,
\]

whenever the above expression is defined (i.e., \( \pm \infty \mp \infty \) is not allowed), and in this case \( f \) is called quasi-integrable. If both integrals are finite then we say that \( f \) is integrable.

By means of the previous proposition, part (c), implies that both definitions agree on simple functions, and parts (a) and (c) remain valid if \( \varphi = f \) and \( \psi = g \) for any integrable functions. To check the linearity, we use the following result. Since \( f^+, f^- \leq |f| = f^+ + f^- \), given a measurable functions \( f \), we deduce that \( f \) is integrable if and only if \( |f| \) is integrable.

Sometimes, an integrable function (as above, with finite integral) is called summable, while a quasi-integrable function (as above, with possible infinite integral) is called integrable.

We keep the notation

\[
\int_{A} f \, d\mu = \int_{\Omega} f 1_{A} \, d\mu, \quad \forall A \in \mathcal{F}
\]

and the inequality

\[
c \mu(\{|f| \geq c\}) \leq \int_{\Omega} |f| 1_{\{|f| \geq c\}} \, d\mu \leq \int_{\Omega} |f| \, d\mu, \quad \forall c \geq 0,
\]

shows that if \( f \) is integrable then the set \( \{|f| \geq c\} \) has finite \( \mu \)-measure, for every \( c > 0 \), and so the set \( \{f \neq 0\} \) is \( \sigma \)-finite. On the other hand, a measurable function \( f \) is allowed to assume the values \( +\infty \) and \( -\infty \), but an integrable function is finite almost everywhere, i.e., \( \mu(\{|f| = \infty\}) = 0 \).
Remark B.44. Instead of initially defining the integral for nonnegative simple functions with the convention $\int_0^{+\infty} = 0$, we may consider only (nonnegative) integrable simple functions in Proposition B.42. In this case, only (nonnegative) measurable functions which vanish outside of a $\sigma$-finite set can be expressed as a (monotone) limit of integrable (nonnegative) integrable simple functions, see Proposition B.9.

A key point is the monotone convergence

**Theorem B.45** (Beppo Levi). If $\{f_n\}$ is a monotone increasing sequence of nonnegative measurable functions then

$$\int_{\Omega} \lim_{n} f_n \, d\mu = \lim_{n} \int_{\Omega} f_n \, d\mu \quad \text{or} \quad \int_{\Omega} \left( \sup_n f_n \right) \, d\mu = \sup_n \left\{ \int_{\Omega} f_n \, d\mu \right\}.$$ 

**Proof.** Since $f_n \leq f_{n+1}$ for every $n$, the limiting function $f$ is defined as taking values in $[0, +\infty]$ and the monotone limit of the integral exists (finite or infinite). Moreover

$$\int_{\Omega} f_n \, d\mu \leq \int_{\Omega} f \, d\mu \quad \text{and} \quad \lim_{n \to \infty} \int_{\Omega} f_n \, d\mu \leq \int_{\Omega} f \, d\mu.$$ 

To check inverse inequality, for every $\alpha \in (0, 1)$ and every simple function $\varphi$ such that $0 \leq \varphi \leq f$ define $F_n = \{ x : f_n(x) \geq \alpha \varphi(x) \}$. Thus $\{F_n\}$ is an increasing sequence of measurable sets with $\bigcup_{n} F_n = \Omega$ and

$$\int_{\Omega} f_n \, d\mu \geq \int_{F_n} f_n \, d\mu \geq \alpha \int_{F_n} \varphi \, d\mu.$$ 

By means of Proposition B.42, part (d), and the continuity from below of a measure, we have

$$\lim_{n} \int_{F_n} \varphi \, d\mu = \int_{\Omega} \varphi \, d\mu, \quad \text{and} \quad \lim_{n \to \infty} \int_{\Omega} f_n \, d\mu \geq \alpha \int_{\Omega} \varphi \, d\mu.$$ 

Since this holds for any $\alpha < 1$, we can take $\alpha = 1$. Taking the sup in $\varphi$ we deduce

$$\lim_{n} \int_{\Omega} f_n \, d\mu \geq \int_{\Omega} f \, d\mu,$$ 

as desired inequality. \qed

The additivity follows from Beppo Levi Theorem, i.e., if $\{f_n\}$ is a finite or infinite sequence of nonnegative measurable functions and $f = \sum_n f_n$ then

$$\int_{\Omega} f \, d\mu = \sum_n \int_{\Omega} f_n \, d\mu.$$ 

Indeed, first for any two functions $g$ and $h$, we can find two monotone increasing sequences $\{g_n\}$ and $\{h_n\}$ of nonnegative simple functions pointwise convergent to $g$ and $h$ respectively. By the monotone convergence theorem, we have

$$\int_{\Omega} g \, d\mu = \lim_{n} \int_{\Omega} g_n \, d\mu \quad \text{and} \quad \int_{\Omega} h \, d\mu = \lim_{n} \int_{\Omega} h_n \, d\mu.$$ 

Since $g_n \leq g$ and $h_n \leq h$, we have

$$\int_{\Omega} g_n \, d\mu \leq \int_{\Omega} g \, d\mu \quad \text{and} \quad \int_{\Omega} h_n \, d\mu \leq \int_{\Omega} h \, d\mu.$$ 

By the monotone convergence theorem again, we have

$$\lim_{n} \int_{\Omega} g_n \, d\mu = \int_{\Omega} g \, d\mu \quad \text{and} \quad \lim_{n} \int_{\Omega} h_n \, d\mu = \int_{\Omega} h \, d\mu.$$ 

Therefore, we have

$$\int_{\Omega} f \, d\mu = \int_{\Omega} g \, d\mu + \int_{\Omega} h \, d\mu = \sum_n \int_{\Omega} f_n \, d\mu.$$ 

The additivity follows from the fact that $f_n$ is a sequence of nonnegative measurable functions and $f = \sum_n f_n$.

\[\text{Preliminary}\] Menaldi November 11, 2016
to \( g \) and \( h \). Thus \( \{g_n + h_n\} \) is a monotone increasing sequence pointwise convergent to \( g + h \), and by means of Theorem B.45

\[
\int_\Omega (g + h) \, d\mu = \lim_{n} \int_\Omega (g_n + h_n) \, d\mu = \lim_{n} \int_\Omega g_n \, d\mu + \lim_{n} \int_\Omega h_n \, d\mu = \int_\Omega g \, d\mu + \int_\Omega h \, d\mu.
\]

Hence, by induction we deduce

\[
\int_\Omega \left( \sum_{n=1}^m f_n \right) \, d\mu = \sum_{n=1}^m \int_\Omega f_n \, d\mu,
\]

and applying again Theorem B.45 as \( m \to \infty \) follows the desired equality.

- **Remark** B.46. Because the integral is unchanged when the integrand is modified in a negligible set, the results of Beppo Levi Theorem B.45 remain valid for an almost monotone sequence \( \{f_n\} \), i.e., when \( f_{n+1} \geq f_n \) a.e., of measurable functions non necessarily nonnegative, but such that \( f_1^- \) is integrable. \( \square \)

Based on the monotone convergence, we deduce two results on the passage to the limit inside the integral. First, Fatou lemma or \( \lim \inf \) convergence

**Theorem B.47** (Fatou). If \( \{f_n\} \) is a sequence of nonnegative measurable functions then

\[
\int_\Omega \lim \inf_n f_n \, d\mu \leq \lim \inf_n \int_\Omega f_n \, d\mu.
\]

**Proof.** For each \( k \) we have that \( \inf_{n \geq k} f_n \leq f_j \) for every \( j \geq k \), which implies that

\[
\int_\Omega \inf_{n \geq k} f_n \, d\mu \leq \int_\Omega f_j \, d\mu \quad \text{and} \quad \int_\Omega \inf_{n \geq k} f_n \, d\mu \leq \inf_{j \geq k} \int_\Omega f_j \, d\mu.
\]

Hence, applying Theorem B.45 as \( k \to \infty \) we have

\[
\int_\Omega \lim \inf_n f_n \, d\mu = \int_\Omega \lim \sup_k \left( \lim \inf_{n \geq k} f_n \right) \, d\mu = \lim_k \int_\Omega \inf_{n \geq k} f_n \, d\mu \leq \lim \inf_n \int_\Omega f_n \, d\mu,
\]

i.e., the desired result. \( \square \)

Secondly, Lebesgue or dominate convergence

**Theorem B.48** (Lebesgue). Let \( \{f_n\} \) be a sequence of measurable functions such that there exists an integrable function \( g \) satisfying \( |f_n(x)| \leq g(x) \), for every \( x \) in \( \Omega \) and any \( n \). Then the functions \( \bar{f} = \lim \sup_n f_n \) and \( \underline{f} = \lim \inf_n f_n \) are integrable and

\[
\int_\Omega f \, d\mu \leq \lim \inf_n \int_\Omega f_n \, d\mu \leq \lim \sup_n \int_\Omega f_n \, d\mu \leq \int_\Omega \bar{f} \, d\mu. \tag{B.11}
\]
In particular,

$$\lim_{n}\int_{\Omega} f_{n} \, d\mu = \int_{\Omega} f \, d\mu.$$ 

provided \( f = \overline{f} \), i.e., \( f_{n} \) converges pointwise to \( f \).

**Proof.** First, note that the condition \( |f_{n}(x)| \leq g(x) \) (valid also for the limit \( \overline{f} \) or \( \underline{f} \)) implies that \( f_{n} \) (and the limit \( \overline{f} \) or \( \underline{f} \)) is integrable. Next, apply Fatou lemma to \( g + f_{n} \) and \( g - f_{n} \) to obtain

$$\int_{\Omega} (g + \underline{f}) \, d\mu \leq \liminf_{n} \int_{\Omega} (g + f_{n}) \, d\mu$$

and

$$\limsup_{n} \int_{\Omega} (g + f_{n}) \, d\mu \leq \int_{\Omega} (g + \overline{f}) \, d\mu,$$

Finally, using the fact that \( g \) is integrable, we deduce (B.11), which implies the desired equalities. \( \square \)

**Remark B.49.** We could re-phase the previous Theorem B.48 as follows: If \( \{f_{n}\} \) and \( \{g_{n}\} \) are sequences of measurable functions satisfying \( |f_{n}| \leq g_{n} \), a.e. for any \( n \), and

$$g_{n} \rightarrow g \text{ a.e. and } \int_{\Omega} g_{n} \, d\mu \rightarrow \int_{\Omega} g \, d\mu < \infty,$$

then the inequality (B.11) holds true. Indeed, applying Fatou lemma to \( g_{n} + f_{n} \) we obtain

$$\int_{\Omega} (g + \underline{f}) \, d\mu = \int_{\Omega} \liminf_{n} (g_{n} + f_{n}) \, d\mu \leq \liminf_{n} \int_{\Omega} (g_{n} + f_{n}) \, d\mu = \int_{\Omega} g \, d\mu + \liminf_{n} \int_{\Omega} f_{n} \, d\mu,$$

which yields the first part of the inequality (B.11), after simplifying the (finite) integral of \( g \). Similarly, by using \( g_{n} - f_{n} \), we conclude. \( \square \)

In the above presentation, we deduced Fatou and Lebesgue Theorems B.47 and B.48 from Beppo Levi Theorem B.45, actually, from any one of them, we can obtain the other two.

**Remark B.50.** A basic consequence of the previous definition of integral on a measure space \((\Omega, \mathcal{F}, \mu)\) is the following list of properties:

1. If \( f \) is an integrable function and \( N \) is a set of measure zero then

$$\int_{N} f \, d\mu = 0.$$
(2) If \( f \) is a strictly positive integrable function and \( E \) is a measurable set such that
\[
\int_E f \, d\mu = 0
\]
then \( E \) is a set of measure zero.

(3) If an integrable function \( f \) satisfies
\[
\int_E f \, d\mu = 0,
\]
for every measurable set \( E \), then \( f = 0 \) a.e.

(4) If \( f \) is a measurable function and \( g \) is an integrable function such that \( |f| \leq g \) a.e., then \( f \) is also an integrable function.

(5) If \( h \) is a nonnegative measurable function then the integral expression
\[
\lambda(A) = \int_A h \, d\mu, \quad \forall A \in \mathcal{F}.
\]
defines a measure on \((\Omega, \mathcal{F})\), which satisfies
\[
\int_\Omega f \, d\lambda = \int_\Omega f \, h \, d\mu,
\]
for every nonnegative measurable function \( f \).

- **Remark B.51.** As mentioned early, the use of the concept “almost everywhere” for a pointwise property in a measure space \((\Omega, \mathcal{F}, \mu)\) is very important, essentially, insisting in a pointwise property could be unwise. For instance, the statement \( f = 0 \) a.e. means strictly speaking that the set \(\{x : f(x) \neq 0\}\) belongs to \(\mathcal{F}\) and \(\mu(\{x : f(x) \neq 0\}) = 0\), but it also could be understood in a large sense as requiring that there exists a set \(N\) in \(\mathcal{F}\) such that \(\mu(N)\) and \(f(x) = 0\) for every \(x\) in \(\Omega \setminus N\). Thus, the large sense refers to the strict sense when \((\mathcal{F}, \mu)\) is complete and certainly, both concepts are the same if the measure space \((\Omega, \mathcal{F}, \mu)\) is complete. Sometimes, we may build-in this concept inside the definition of the integral by adding the condition almost everywhere, i.e., using
\[
\int_\Omega f \, d\mu = \sup \left\{ \int_\Omega \varphi \, d\mu : \varphi \text{ simple, } 0 \leq \varphi \leq f \text{ a.e.} \right\}
\]
as the definition of integral (where the a.e. inequality is understood in the large sense) for any nonnegative “almost” measurable function, i.e., any function \(f\) such that there exist a negligible set \(N\) and a nonnegative measurable function \(g\) such that \(f(x) = g(x)\), for every \(x\) in the complement \(N^c\). Therefore, parts (1) and (2) of Remark B.50 are necessary to prove that the above definition of integral (with the a.e. inequality) is indeed meaningful and non-ambiguous.

- **Remark B.52.** Certainly, it can be proved that every Riemann integrable function is Lebesgue measurable and both integrals coincide. Moreover, a bounded function is Riemann integrable if and only if it is continuous almost everywhere.
B.5.2 Cartesian Products

As we have seen that we can change the values of an integrable function in a set of measure zero without any changes in its integral, however, we need to know that the resulting function is measurable, e.g., we should avoid the situation of measure zero without any changes in its integral, however, we need to know

As we have seen that we can change the values of an integrable function in a set

B.5.2 Cartesian Products

Appendix B. Measure and Integration

Let \((X, \mathcal{X}, \mu)\) be a \(\sigma\)-finite measure space and \((Y, \mathcal{Y})\) be a measurable space.

A function \(\nu : X \times Y \to [0, +\infty]\) is called a \(\sigma\)-finite regular transition measure if

(a) the mapping \(x \mapsto \nu(x, B)\) is \(\mathcal{X}\)-measurable for every \(B \in \mathcal{Y}\),

(b) the mapping \(B \mapsto \nu(x, B)\) is a measure on \(\mathcal{Y}\) for every \(x \in X\),

(c) there exists increasing sequences \(\{X_n\} \subset \mathcal{X}\) and \(\{Y_n\} \subset \mathcal{Y}\) such that \(\bigcup_{n=1}^{\infty} X_n = X, \bigcup_{n=1}^{\infty} Y_n = Y\),

\[\nu(x, Y_n) < \infty, \quad \forall x \in X, \quad \int_{X_n} \mu(dx)\nu(x, Y_n) < \infty, \quad \forall n. \quad (B.12)\]

If \(\mu(X) = 1\) and \(\nu(x, Y) = 1\) for every \(x \in X\) then \(\nu\) is called a transition probability measure. The qualification regular is attached to the condition (b), a non regular transition measure would satisfy almost everywhere the \(\sigma\)-additivity property, i.e., besides the condition \(\nu(x, \emptyset) = 0\), for every sequence of disjoint set \(\{B_k\} \subset \mathcal{Y}\) there exists a set \(A \in \mathcal{X}\) with \(\mu(A) = 0\) such that \(\nu(x, \bigcup_{k} B_k) = \sum_{k} \nu(x, B_k)\), for every \(x \in X \setminus A\).

Note the following two particular cases: (1) \(\nu(x, B) = \nu(B)\) independent of \(x\), for a given \(\sigma\)-finite measure \(\nu\) on \(\mathcal{Y}\), and (2) \(\nu(x, B) = \sum_{k=1}^{\infty} a_k(x) 1_{f_k(x) \in B}\), for sequences \(\{a_k\}\) and \(\{f_k\}\) of measurable functions \(a_k : X \to [0, \infty)\) and \(f_k : X \to Y\), i.e., a sum of Dirac measures \(\nu = \sum_{k} a_k(x) \delta_{f_k(x)}\). Remark that, for the case (2), the assumptions on transition measure \(\nu\) are equivalent to the measurability of the functions \(a_k\) and \(f_k\), for every \(k\).

For any \(E \subset X \times Y\) and any \(x \in X\), we define the sections as the sets \(E_x = \{y \in Y : (x, y) \in E\} \subset Y\) (similarly \(E^y\), by exchanging \(X\) with \(Y\)). Note that \((E \cup F)_x = E_x \cup F_x\) and \((E \setminus F)_x = E_x \setminus F_x\), but we may have \(E \cap F = \emptyset\) with \(E_x \cap F_x \neq \emptyset\). Recall that the product \(\sigma\)-algebra \(\mathcal{X} \times \mathcal{Y}\) is generated by the semi-algebra of rectangle \(A \times B\) with \(A \in \mathcal{X}\) and \(B \in \mathcal{Y}\).

Proposition B.53. Let \(\nu(\cdot, \cdot)\) be a \(\sigma\)-finite transition measure from \(\sigma\)-finite measure \((X, \mathcal{X}, \mu)\) into \((Y, \mathcal{Y})\) as above, and let \(E\) be a set in product \(\sigma\)-algebra \(\mathcal{X} \times \mathcal{Y}\). Then (a) all sections are measurable, i.e., \(E_x \in \mathcal{Y}\), for every \(x \in X\); (b) the mapping \(x \mapsto \nu(x, E_x)\) is \(\mathcal{X}\)-measurable and (c) the mapping

\[E \rightarrow (\mu \times \nu)(E) = \int_X \mu(dx)\nu(x, E_x), \quad \forall E \in \mathcal{X} \times \mathcal{Y}\]

is a \(\sigma\)-finite measure, in particular the expression

\[(\mu \times \nu)(A \times B) = \int_A \nu(x, B) \mu(dx), \quad \forall A \in \mathcal{X}, B \in \mathcal{Y},\]

uniquely determines the values of the product measure \(\mu \times \nu\).
Proof. First remark that for any $E = A \times B$ the sections satisfy $E_x = B$ if $x \in A$ and $E_x = \emptyset$ if $x \notin A$. Hence $\nu(x, E_x) = 1_A \nu(x, B)$, for any rectangle $E$ and with the convention that $0 \infty = 0$.

Take increasing sequences $\{X_n\} \subset \mathcal{X}$ and $\{Y_n\} \subset \mathcal{Y}$ as in (B.12). It is clear that if the conditions (a) and (b) hold for $E \cap (X_n \times Y_n)$ instead of $E$, for every $n$, then they should be valid for $E$. Thus we may assume

$$\nu(x, Y) < \infty, \quad \forall x \in X,$$
$$\int_X \mu(dx) \nu(x, Y) < \infty,$$

without any loss of generality.

Let $\mathcal{D}$ be the class of sets $E$ in $\mathcal{X} \times \mathcal{Y}$ for which the conditions (a) and (b) are satisfied. Because $(F \cup E)_x = F_x \cup E_x$ and $(F \setminus E)_x = F_x \setminus E_x$, the family $\mathcal{D}$ is a $\lambda$-class, which contains the $\pi$-class of all rectangle. Hence, a monotone argument (see Proposition B.5) shows that $\mathcal{D} = \mathcal{X} \times \mathcal{Y}$.

To check (c), we need to verify that the product $\mu \times \nu$ is $\sigma$-additive on the semi-algebra of measurable rectangle. To this purpose, note that if $E = \sum_{k=1}^{\infty} E_k$, $E = A \times B$ and $E_k = A_k \times B_k$ then

$$\mathbb{1}_A(x) \mathbb{1}_B(y) = \sum_{k=1}^{\infty} \mathbb{1}_{A_k}(x) \mathbb{1}_{B_k}(y), \quad \forall x, y$$

Thus, the $\sigma$-additivity of the measure $\nu(x, \cdot)$ implies

$$\mathbb{1}_A(x) \nu(x, B) = \sum_{k=1}^{\infty} \mathbb{1}_{A_k}(x) \nu(x, B_k), \quad \forall x \in X,$$

and the monotone convergence (Theorem B.45) yields

$$\int_A \mu(dx) \nu(x, B) = \sum_{k=1}^{\infty} \int_{A_k} \mu(dx) \nu(x, B_k), \quad \forall A \in \mathcal{X}, B \in \mathcal{Y}.$$ 

At this point, either by Proposition B.18 or repeating the above argument with any $E \in \mathcal{X} \times \mathcal{Y}$ and remarking that $\mathbb{1}_{E}(x, y) = \mathbb{1}_{E_x}(y)$, we deduce

$$E \to (\mu \times \nu)(E) = \int_X \mu(dx) \nu(x, E_x) = \int_X \mu(dx) \int_Y \mathbb{1}_E(x, y) \nu(x, dy),$$

for every $E \in \mathcal{X} \times \mathcal{Y}$, is a $\sigma$-finite measure.

\[ \Box \]

\textbf{Remark B.54.} It is clear that in Proposition B.53 we also proved that the function $y \mapsto \mu(E_y)$ is $\mathcal{Y}$-measurable, and if the transition measure $\nu$ is actually a measure on $(Y, \mathcal{Y})$ then we deduce the equality

$$\int_X \nu(E_x) \mu(dx) = \int_Y \mu(E_y) \nu(dy), \quad \forall E \in \mathcal{X} \times \mathcal{Y}$$

as expected.

\[ \Box \]
By means of Proposition B.9, we can approximate a measurable functions by a pointwise convergence sequence of simple functions to deduce from Proposition B.53 that if \( f : X \times X \to \bar{\mathbb{R}} \) is a \( \mathcal{X} \times \mathcal{Y} \)-measurable function then for every \( y \in Y \), the section function \( x \mapsto f(x, y) \) is \( \mathcal{X} \)-measurable. Certainly, we may replace the extended real \( \bar{\mathbb{R}} \) by any separable metric space to deduce that the sections of a product-measurable functions are indeed measurable. Note that the converse is not valid in general, i.e., although if a contra-example is not easy to get, we may have a non measurable subset \( E \) of \( X \times Y \) such that the sections \( E_x \) and \( E_y \) are measurable, for every fixed \( x \) and \( y \).

Moreover, for any \( N \in \mathcal{X} \times \mathcal{Y} \) we have \( (\mu \times \nu)(N) = 0 \) if and only if its sections \( N_x \) have \( \nu(x, \cdot) \)-measure zero, for \( \mu \)-almost every \( x \), i.e., there exists a set \( A_N \in \mathcal{X} \) such that \( \mu(A_N) = 0 \) and \( \nu(x, N_x) = 0 \), for every \( x \in X \setminus A_N \). Hence, if \( (\lambda, \mathcal{F}) \) is the completion of the product measure \( \mu \times \nu \), and if \( f : X \times Y \to \bar{\mathbb{R}} \) is \( \mathcal{F} \)-measurable then there exists a \( \mathcal{X} \times \mathcal{Y} \)-measurable function \( \bar{f} \) such that \( \lambda(\{(x, y) : f(x, y) \neq \bar{f}(x, y)\}) = 0 \). Moreover, there exists a set \( N \in \mathcal{X} \times \mathcal{Y} \) with \( \lambda(N) = 0 \) such that \( f(x, y) = \bar{f}(x, y) \) for every \( (x, y) \notin N \). Thus we have

**Corollary B.55.** Let \( (\lambda, \mathcal{F}) \) be the completion of the product measure \( \mu \times \nu \), as given by Proposition B.53. If \( f : X \times Y \to \bar{\mathbb{R}} \) is \( \mathcal{F} \)-measurable then there exists a set \( A_f \) in \( \mathcal{X} \) with \( \mu(A_f) = 0 \) such that the function \( y \mapsto f(x, y) \) is \( \mathcal{Y} \)-measurable, for every \( x \in X \setminus A_f \).

**Proof.** In view of the approximation by simple functions (see Proposition B.9), we need to show the result only for \( f = 1_E \) with \( E \in \mathcal{F} \).

Now, for a \( \lambda \)-measurable set \( E \) there exists sets \( E', N \in \mathcal{X} \times \mathcal{Y} \) such that \( (E \setminus E') \cup (E' \setminus E) \subset N \), i.e., \( |1_E - 1_{E'}| \leq 1_N \). Because \( \nu(x, \cdot) \) is \( \sigma \)-finite regular transition measure, there is an increasing sequence \( \{Y_n\} \subset \mathcal{Y} \) such that \( \nu(x, Y_n) < \infty \) for every \( x \in X \), for every \( n \). Thus, \( \nu(x, Y_n \cap E_x) < \infty \) and \( |\nu(x, Y_n \cap E_x) - \nu(x, Y_n \cap E'_x)| \leq \nu(x, N_x) \), for every \( n \) and every \( x \in X \). Since

\[
0 = \lambda(N) = \int_X \mu(dx) \nu(x, N_x) = \int_X \mu(dx) \int_Y 1_{N}(x, y)\nu(x, dy),
\]

there exists a set \( A_E \in \mathcal{X} \) with \( \mu(A_E) = 0 \) such that \( \nu(x, N_x) = 0 \) for every \( x \in X \setminus A_E \). Hence \( \nu(x, E_x) = \nu(x, E'_x) \), for every \( x \notin A_E \). \qed

**Remark B.56.** Recall that the approximation of measurable functions by integrable simple functions (as in Proposition B.9) can only be used a in \( \sigma \)-finite space, i.e., if the space is not \( \sigma \)-finite then there are nonnegative measurable functions which are nonzero on a non \( \sigma \)-finite set, and therefore, they can not be a pointwise limit of integrable simple functions. On the other hand, there are ways of dealing with product of non \( \sigma \)-finite measures, essentially, only \( \sigma \)-finite measurable sets (i.e., covered by a sequence \( \{A_k \times B_k : k \geq 1\} \) of rectangles where each \( A_k \) and \( B_k \) is measurable and the product measure of \( \bigcup_k A_k \times B_k \) is finite) are considered and the product measure is defined on a \( \sigma \)-ring, instead of a \( \sigma \)-algebra. The difficulty is the measurability of the mapping \( x \mapsto \nu(x, E_x) \), for an arbitrary measurable set \( E \) on the product space, for instance see Pollard [102, Section 4.5, pp. 93-95]. \qed
Theorem B.57 (Fubini-Tonelli). Let $\lambda$ be the completion of the product measure $\mu \times \nu$ defined in Proposition B.53 and let $f: X \times Y \to [0, \infty]$ be a $\mathcal{X} \times \mathcal{Y}$-measurable (respect., $\lambda$-measurable) function. Then (a) the function $f(x, \cdot)$ is $\mathcal{Y}$-measurable for every $x$ in $X$ (respect., for $\mu$-almost everywhere $x$ in $X$); (b) the function

$$x \mapsto \int_Y f(x, y) \nu(x, dy)$$

(respect., measurable with respect to the completion of $\mu$); (c) we have

$$\int_{X \times Y} f(x, y) \lambda(dx, dy) = \int_X \mu(dx) \int_Y f(x, y) \nu(x, dy).$$

Proof. Let $E \subset X \times Y$ be a $\lambda$-measurable set, i.e., there exists $E', N \in \mathcal{X} \times \mathcal{Y}$ such that $(E \setminus E') \cup (E' \setminus E) \subset N$ and $(\mu \times \nu)(N) = 0$. If $f = 1_E$ then Proposition B.53 and Corollary B.55 proves the validity of the assertions for this particular case, and so for any simple function. Next, we conclude by approximating $f$ by a monotone sequence of nonnegative simple functions.

If $f: X \times Y \to \bar{\mathbb{R}}$ is $\lambda$-integrable then $f$ takes finite valued outside of a set $N \in \mathcal{X} \times \mathcal{Y}$ with $(\mu \times \nu)(N) = 0$. Applying (a), (b) and (c) for $f^+$ and $f^-$ we deduce that (1) $f(x, \cdot)$ is $\nu(x, \cdot)$-integrable for $\mu$-almost everywhere $x$ in $X$; (2) the integral of $f(x, y)$ with respect to $\nu(x, dy)$ is $\mu$-integrable; (3) the iterate integral reproduces the double integral, i.e., (B.13) holds.

In the particular case of a constant transition measure $\nu(x, \cdot) = \nu(\cdot)$, we may consider also $\nu \times \mu$ and we deduce from (B.13) the exchange of the integration order, i.e.,

$$\int_{X \times Y} f(x, y) \lambda(dx, dy) = \int_X \mu(dx) \int_Y f(x, y) \nu(dy) =$$

$$= \int_Y \nu(dy) \int_X f(x, y) \mu(dx),$$

for every $f$ either nonnegative and measurable or integrable in the product space. This is the traditional Fubini-Tonelli Theorem.

It is clear that these arguments extend to a finite product, with suitable transition measures. The reader may take a look at Ambrosio et al. [8, Chapter 6, pp. 83–118] and Taylor [122, Chapter 7, 324–347].

B.5.3 Some Inequalities

Now, let $L^0 = L^0(\Omega, \mathcal{F}, \mu; E)$ be the space of all almost measurable $E$-valued functions, where $(E, |\cdot|_E)$ is a Banach space. For $1 \leq p \leq \infty$ and any $f \in L^0$
we consider
\[ \|f\|_p = \left( \int_{\Omega} |f|^p d\mu \right)^{1/p} < \infty, \quad \forall 1 \leq p < \infty, \]  
\[ \|f\|_{\infty} = \inf \{ C \geq 0 : |f|_E \leq C, \text{ a.e.} \}, \]
where \( \|f\|_{\infty} = \infty \) if \( \mu(\{ x : |f(x)|_E \geq C \}) > 0 \), for every \( C > 0 \). We define \( L^p = L^p(\Omega, F, \mu; E) \) as the subspace of \( L^0(\Omega, F, \mu; E) \) such that \( \|f\|_p < \infty \) and for \( p = \infty \) we add the condition (which is already included if \( p < \infty \)) that \( \{ f \neq 0 \} \) is a \( \sigma \)-finite (i.e., a countable union of sets with finite measure). Recall that elements \( f \) in \( L^0 \) are equivalence classes (i.e., functions defined almost everywhere), and that \( f \) takes valued in some separable subspace of \( E \), when \( E \) is not separable.

Most of what follows is valid for a (separable) Banach space \( E \), but to simplify, we consider only the case \( E = \mathbb{R} \) or \( E = \mathbb{R}^d \), with the Euclidean norm is denoted by \( | \cdot | \).

We have already shown that \( (L^1, \| \cdot \|_1) \) and \( (L^\infty, \| \cdot \|_{\infty}) \) are Banach spaces. The general case \( 1 < p < \infty \) requires some estimates to prove that \( \| \cdot \|_p \) is indeed a norm.

First, recalling that the \( - \ln \) function is a strictly convex function,
\[ \ln(ax + by) \geq a \ln x + b \ln y, \quad \forall a, b, x, y > 0, \ a + b = 1, \]
we check that the arithmetic mean is larger that the geometric mean, i.e.,
\[ x^a y^b \leq ax + by, \quad \forall a, b, x, y > 0, \ a + b = 1, \]
where the equality holds only if \( x = y \).

(a) Hölder inequality: for any \( p, q \geq 1 \) with \( 1/p + 1/q = 1 \) (where the limit case \( 1/\infty = 0 \) is used) we have
\[ \|fg\|_1 \leq \|f\|_p \|g\|_q, \quad \forall f \in L^p, \ g \in L^q, \]  
\[ \|fg\|_1 \leq \|f\|_p \|g\|_q, \quad \forall f \in L^p, \ g \in L^q, \]
where the equality holds only if for some constant \( c \) we have \( |f|^p = c |g|^q \), almost everywhere. Indeed, if \( \|fg\|_1 > 0 \) then \( \|f\|_p > 0 \) and \( \|g\|_q > 0 \). Taking \( a = 1/p, \ b = 1/q, \ x = |f|^p/\|f\|_p^p \) and \( y = |g|^q/\|g\|_q^q \) in (B.15) and integrating in \( \mu \), on deduce (B.16).

If \( 1 \leq p < r < q \leq \infty \) and \( f \) belongs to \( L^p \cap L^q \) then \( f \) belongs to \( L^r \) and
\[ (1/p - 1/q) \ln \|f\|_r \leq (1/r - 1/q) \ln \|f\|_p + (1/p - 1/r) \ln \|f\|_q. \]
Indeed, for some \( \theta \) in \( (0, 1) \) we have \( 1/r = \theta/p + (1 - \theta)/q \) and Hölder inequality yields
\[ \|f\|_r = \|f^\theta f^{1-\theta}\|_r = \|f^{r \theta} f^{r(1-\theta)}\|_r^r \leq \left\{ \|f^{r \theta}\|_{p/r \theta} \|f^{r(1-\theta)}\|_{q/r(1-\theta)} \right\}^{1/r} \]
\[ = \left\{ \|f\|_{p \theta} \|f\|_{q(1-\theta)} \right\}^{1/r} = \|f\|_{p \theta} \|f\|_{q}^{1-\theta}, \]
and the desired estimate follows.

(b) **Minkowski inequality:** if \( 1 \leq p \leq \infty \) then
\[
\| f + g \|_p \leq \| f \|_p + \| g \|_p, \quad \forall f, g \in L^p.
\] (B.17)

Indeed, only the case \( 1 < p < \infty \) need to be considered. Thus the inequality
\[
|f + g|^p \leq (|f| + |g|)^p \leq 2^p(|f|^p + |g|^p)
\]
shows that \( f + g \) belongs to \( L^p \). With \( q = p/(p-1) \) we have
\[
\| f + g \|_p^{-1} \leq (\| f \|_p + \| g \|_p)^{p-1}.
\]
Next, applying (B.16) to
\[
|f + g|^p = |f + g| |f + g|^{p-1} \leq |f| |f + g|^{p-1} + |g| |f + g|^{p-1}
\]
we obtain (B.17).

Therefore \( (L^p, \| \cdot \|_p) \) is a normed space, and the inequality
\[
\varepsilon^p \mu(\{|f| \geq \varepsilon\}) \leq \| f \|_p^p,
\]
shows that if \( \{f_n\} \) is a Cauchy sequence in \( L^p \) then it is also a Cauchy sequence in \( L^0 \). Hence \( L^p \) is complete, i.e., it is a Banach space.

- **Remark** B.58. If \( 0 < p < 1 \) and \( f, g \) belongs to \( L^p \) then \( f + g \) belongs to \( L^p \) and
\[
\| f + g \|^p_p \leq \| f \|^p_p + \| g \|^p_p.
\]

This follows from the elementary inequality \((a + b)^p \leq a^p + b^p\), for every \( a, b \) in \([0, \infty)\) and \( 0 < p < 1 \), which is deduced from \([a/(a + b)]^p + [b/(a + b)]^p \geq a/(a + b) + b/(a + b) = 1\). Hence \( L^p \) with the distance \( d_p(f, g) = \| f - g \|_p \), \( 0 < p < 1 \), is a (complete metric) topological vector space. Also we have
\[
\| f + g \|_p \geq \| f \|_p + \| g \|_p, \quad \forall f, g \in L^p, \ 0 < p < 1,
\]
\[
\| fg \|_1 \leq \| f \|_p \| g \|_q, \quad \forall f \in L^p, \ g \in L^q,
\]
again \( 1/p + 1/q = 1 \), but in this case \( q < 0 \). It is possible to show that
\[
\lim_{p \to 0} \| f \|^p_p = \mu(\{\omega \in \Omega : f(\omega) \neq 0\}),
\]
\[
\lim_{p \to 0} \| f \|_p = \exp \left( \int_{\Omega} \ln |f| \, d\mu \right), \quad \text{if } \mu(\Omega) = 1 \text{ and } f \neq 0 \ a.e.,
\]
provided \( f \) belongs to some \( L^p(\Omega, \mathcal{F}, \mu) \) with \( p > 0 \). Indeed, the first inequality follows after splitting the integral over the regions \( 0 < |f(x)| \leq 1 \) and \( |f(x)| > 1 \). To check the second inequality, we assume \( |f| > 0 \ a.e. \) to show (with the help of the mean value theorem) that
\[
\ln \| f \|_p = \| f \|_{-q} \int_{\Omega} |f|^q \ln |f| \, d\mu,
\]
for some \( q \) in \((0, p)\). Hence, as in the argument to prove first inequality, we have
\[
\| f \|_{-q} \int_{\Omega} |f|^q \ln |f| \, d\mu \to \int_{\Omega} \ln |f| \, d\mu.
\]
Notice that if \( |f| > 0 \) on a set \( \Omega_0 \) with \( 0 < \mu(\Omega_0) < 1 \) then we use the previous argument on the space \( \Omega_0 \) with the measure \( A \mapsto \mu(A)/\mu(\Omega_0) \). \qed
Remark B.59. First, if $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ are two $\sigma$-finite measure spaces then Minkowski’s integral inequality, states that

$$\left[ \int_X \left( \int_Y f(x, y) \nu(dy) \right)^p \mu(dx) \right]^{1/p} \leq \int_Y \left( \int_X |f(x, y)|^p \mu(dx) \right)^{1/p} \nu(dy)$$

for any real-valued $(\mu \times \nu)$-measurable function $f$. Moreover, this inequality can be generalized in the following way. If $f(x, y)$ is a nonnegative measurable function on the product space $X \times Y$ and $1 \leq p \leq \infty$ then

$$\left\| \int_Y f(\cdot, y) \nu(dy) \right\|_{L^p(X)} \leq \int_Y \left\| f(\cdot, y) \right\|_{L^p(X)} \nu(dy),$$

where the integral in $Y$ is regarded as a limit of sums, i.e., approximating $f$ by an increasing sequence of simple measurable functions and taking limit. This is usually referred to as Minkowski inequality for integrals.

Based on Hölder inequality, we can define the duality paring

$$\langle f, g \rangle = \int_{\Omega} f g \, d\mu, \quad \forall f \in L^p, \ g \in L^q, \ \frac{1}{p} + \frac{1}{q} = 1,$$

(B.18)

which has the property $|\langle f, g \rangle| \leq \|f\|_p \|g\|_q$.

Proposition B.60 (dual norm). For any function $f$ in $L^0(\Omega, \mathcal{F}, \mu)$ with $\sigma$-finite support $\{f \neq 0\}$ we have

$$\|f\|_p = \sup \{ \langle f, g \rangle : g \in L^q, \ \text{with} \ \|g\|_q = 1 \}, \quad 1 \leq p \leq \infty,$$

(B.19)

where $\langle \cdot, \cdot \rangle$ is the duality paring (B.18), and the supremum is attained with $g = \text{sign}(f) |f|^{p-1} \|f\|_p^{1-p}$, if $p < \infty$ and $0 < \|f\|_p < \infty$.

Proof. Temporarily denote by $\|f\|_p$ the right-hand term of (B.19). Thus Hölder inequality yields $\|f\|_p \leq \|f\|_p$.

For $p < \infty$ and $0 < \|f\|_p < \infty$ define $g = \text{sign}(f) |f|^{p-1} \|f\|_p^{1-p}$ to get $\|g\|_q = 1$, $1/p + 1/q = 1$, and $\langle f, g \rangle = \|f\|_p$. On the other hand, if $0 < a < \|f\|_\infty$ then define the function $g = \text{sign}(f) 1_A/\mu(A)$ with $A = \{x : |f(x)| > a\}$ to get $\|g\|_1 = 1$ and $\langle f, g \rangle \geq a$. Hence we have the reverse inequality $\|f\|_p \leq \|f\|_p$, provided $p = \infty$ or $\|f\|_p < \infty$.

If $\|f\|_p = \infty$ then $f$ is a pointwise limit of a bounded $\mu$-measurable bounded functions $f_n$ such that $\mu(\{f_n \neq 0\}) < \infty$ and $|f_n| \leq |f_{n+1}| \leq |f|$. Then $\|f_n\|_p$ increases to $\|f\|_p = \infty$ and $\|f_n\|_p = \|f_n\|_p \leq \|f\|_p$, i.e., $\|f\|_p = \infty$.

Remark B.61. The above proof shows that we may replace the condition $\|g\|_q = 1$ by $\|g\|_q \leq 1$ and the equality (B.19) remain true. Moreover, we may take the supremum only over simple functions $g$ in $L^q$ satisfying $\|g\|_q = 1$, i.e.,

$$\|f\|_p = \sup \{ \langle f, \varphi \rangle : \varphi \in S^1, \ \text{with} \ \|\varphi\|_q = 1 \},$$

where $S^1 = S^1(\Omega, \mathcal{F}, \mu)$ is the space of simple functions, $\varphi = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$, with $\{A_i\}$ measurable and $\mu(A_i) < \infty$, for every $i$.

For instance, the reader may consult Folland [44, Section 193–197], Jones [70, Chapter 12, pp. 277–291] for more details.
B.5.4 Orthogonal Projection

Some of the properties valid in the Euclidean spaces \( \mathbb{R}^n \) or \( \mathbb{C}^n \) can be extended to some infinite dimensional spaces, such as \( L^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^n) \) or \( L^2(\Omega, \mathcal{F}, \mu; \mathbb{C}^n) \). Perhaps, at this level, the reader should take a look at the beginning of the book Halmos [65] for a short introduction to Hilbert spaces.

Our interest is on the orthogonal projection and the representation of linear continuous functionals for the \( L^2 \) space over the complex number, we have

\[
(f,g) = \int_{\Omega} f(x) \overline{g(x)} \mu(dx), \quad \forall f, g \in L^2(\Omega, \mathcal{F}, \mu; \mathbb{C}),
\]

and \( \|f\|^2 = (f,f) \), where the notation \( (\cdot,\cdot) \) is reserved for the duality, even when discussing real-valued functions \( f \) and the complex-conjugate operator \( f \mapsto \overline{f} \) is not used. This special form of the norm yields the so-called parallelogram equality \( \|f+g\|^2 + \|f-g\|^2 = 2\|f\|^2 + 2\|g\|^2 \), for every \( f,g \in H \), and the identity \( \|f+g\|^2 - \|f-g\|^2 = 4(f,g) \) allows the re-definition of the scalar product in term of the norm.

Actually recall that a Hilbert space is a vector space (on \( \mathbb{R} \) or \( \mathbb{C} \)) with a scalar (or inner) product satisfying:

a. \( (f,f) \geq 0, \quad \forall f \in H, \quad \text{and} \quad (f,f) = 0 \) only if \( f = 0 \);

b. \( (af + bg,h) = a(f,h) + b(g,h), \quad \forall f, g, h \in H \) and \( a, b \in \mathbb{R} \) (or \( \mathbb{C} \));

c. \( (f,g) = \overline{(g,f)}, \quad \forall f, g \in H \);

plus the completeness axiom: every Cauchy sequence \( \{f_n\} \subset H \), i.e., \( (f_n - f_m, f_n - f_m) \to 0 \) as \( n, m \to \infty \), is convergent, i.e., there exists \( f \in H \) such that \( (f_n - f, f_n - f) \to 0 \) as \( n, m \to \infty \). Hence, by considering the nonnegative quadratic \( r \mapsto \|f+rg\|^2 \) and using the linearity we deduce the Cauchy inequality,

\[
|(f,g)| \leq \|f\| \|g\|, \quad \forall f, g \in H,
\]

where the equality holds if and only if \( f \) and \( g \) are co-linear, i.e., \( f = cg \) or \( cf = g \) for some constant \( c \).

Two elements \( f, g \) in a Hilbert space \( H \) are called orthogonal if \( (f,g) = 0 \), and we may define the orthogonal complement of any nonempty subset \( V \subset H \) as \( V^\perp = \{ h \in H : (h,v) = 0, \forall v \in V \} \). From the continuity and the linearity of the scalar product we deduce that \( V^\perp \) is a closed subspace of \( H \).

**Proposition B.62** (Orthogonal Projection). Let \( K \) be a closed convex set of \( H \). Then there exists a unique operator \( P : H \to K \) such that \( f \mapsto Pf \) satisfies

\[
(Pf - f, k - Pf) \geq 0, \quad \forall k \in K. \tag{B.20}
\]

Moreover, we have the estimate \( \|Pf - Pg\| \leq \|f - g\| \) for every \( f \) and \( g \) in \( H \); and if \( K \) is a closed subspace then \( P \) is linear and (B.20) becomes \( (Pf - f, k) = 0 \) for every \( k \) in \( K \).
Proof. First check the uniqueness. For any \( g \) in \( H \), \( Pf \) satisfies

\[
(Pg - g, k - Pg) \geq 0, \quad \forall k \in K.
\]

Take \( k = Pf \) and add (B.20) with \( k = Pg \) to deduce \( (f - g, Pf - Pg) \geq \|Pf - Pg\|^2 \), which yields the estimate and the uniqueness. If \( K \) is a closed subspace then \( k - Pf \in K \) if and only if \( k \in K \), i.e., (B.20) is equivalent to \((Pf - f, k) = 0\) for every \( k \in K \) and the linearity of \( P \) follows.

Next, for every fixed \( f \) in \( H \), consider the nonlinear functional \( h \mapsto I(h) = (h - 2f, h) \) on \( H \) and set \( a = \inf \{I(h) : h \in K\} \). Since \( I(h) \geq \|h\|^2 - 2\|f\| \|h\|, \) we obtain \( a \geq -\|f\|^2 > -\infty \), and so we can find a minimizing sequence \( \{h_n\} \subset K \) such that \( a \leq I(h_n) \leq a + n^{-1} \), for every \( n \geq 1 \). Because \( K \) is convex, \( h_{n,m} = (h_n + h_m)/2 \) belongs to \( K \) and we obtain

\[
\|h_n\|^2 + \|h_m\|^2 - 2\|h_{n,m}\|^2 = I(h_n) + I(h_m) - 2I(h_{n,m}) \leq 1/n + 1/m,
\]

after canceling the linear part of \( I \). Hence, applying the parallelogram equality we have

\[
\|h_n - h_m\|^2 = 2\|h_n\|^2 + 2\|h_m\|^2 - \|h_n - h_m\|^2 \leq 2/n + 2/m,
\]

which proves that \( \{h_n\} \) is a Cauchy sequence in \( K \). The whole space \( H \) is complete and \( K \) is closed, therefore, there exists \( h \) in \( K \) such that \( \|h_n - h\| \to 0 \). Now, for every \( k \) in \( K \) we have \( h + \varepsilon(k - h) \) in \( K \), for any \( \varepsilon \) in \([0,1]\), and so \( I(h + \varepsilon(k - h)) \geq I(h) \), i.e.,

\[
2\varepsilon(h - f, k - h) + \varepsilon^2\|k - h\|^2 \geq 0.
\]

Thus, dividing by \( \varepsilon \) and then vanishing \( \varepsilon \), we get (B.20) with \( Pf = h \). \( \square \)

Sometimes, we write \( P = P_K \) to emphasize the dependency on \( K \). Also, \( P_K \) is called the orthogonal projection over \( K \). It is clear that \( P_Kf = f \) for every \( f \) in \( K \), i.e., \( P_K \) is idempotent. If \( K \) is a closed subspace then \( Pf - f \) belongs to \( K_\perp \), i.e., \( f = Pf + (f - Pf) \), which means \( H = K \oplus K_\perp \). For any nonempty subset \( V \) of \( H \), we have defined its orthogonal complement \( V_\perp = \{h \in H : (h, v) = 0, \forall v \in V\} \), but only when \( V = K \) is a closed subspace we obtain \( V = (V_\perp)_\perp \). Also, by writing \( f = Pf + (f - Pf) \) we deduce \( (Pf, g) = (Pf, Pg) = (f, Pg) \), for every \( f, g \in H \), i.e., the projection is a symmetric operator.

If \((H, \|\cdot\|)\) is a Hilbert space then we denote by \( H' \) its dual space, i.e., the space of all continuous linear functionals \( T : H \to \mathbb{K} \), with \( \mathbb{K} = \mathbb{R} \) or \( \mathbb{K} = \mathbb{C} \). We can check that \( H' \) endowed with the dual norm

\[
\|Tf\|_{H'} = \|Tf\|' = \sup \{|Tf| : \|f\| \leq 1\}
\]

is a Banach space, and more detail is needed to see that \( \|\cdot\|' \) satisfies the parallelogram equality, and so, \( H' \) is a Hilbert space.

Thus, if \( f \) belongs to \( H \) we can define \( \Phi f : H \to \mathbb{R}, \Phi f(h) = (h, f) \), which results an element in \( H' \). It is clear that the map \( f \mapsto \Phi f \) is (sesqui-)linear from \( H \) into \( H' \), and Cauchy inequality shows that \( \|\Phi f\|' = \|f\| \) for every \( f \) in \( H \).
Theorem B.63 (Riesz Representation). Let $H$ a Hilbert space. If $T: H \rightarrow \mathbb{K}$, with $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, is a continuous linear functional then there exists $f$ in $H$ such that $T(h) = (h, f)$, for every $h$ in $H$. Moreover, the application $\Phi$ defined above is an isometry from $H$ onto its dual $H'$.

Proof. It is clear that only the fact that $\Phi$ is onto should be shown, i.e., given $T$ we can find $f$. To this purpose, denote by $\text{Ker}(T)$ the kernel or null space of $T$, i.e., all elements in $h \in H$ such that $T(h) = 0$. If $\text{Ker}(T) = H$ then $f = 0$ satisfies $\Phi(f) = T$, otherwise, there exits $g \neq 0$ in the orthogonal complement $\text{Ker}(T) \perp$, and after diving by $T(g)$ if necessary, we may suppose $T(g) = 1$. Now, for any $h$ in $H$ we have $T(h - T(h)g) = 0$ and so $h - T(h)g$ belongs to $\text{Ker}(T)$, i.e., $(h - T(h)g, g) = 0$. This can written as $T(g, g) = (h, g)$, for every $h$ in $H$. Hence, $f = g/(g, g)$ satisfies the desired condition. \hfill \Box

Among other things this proves the

Corollary B.64. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $T : L^2 \rightarrow \mathbb{R}$ be a linear functional, which is continuous, i.e., for some constant $C > 0$,

$$|T(f)| \leq C \|f\|_2, \quad \forall f \in L^2.$$  

Then there exists a unique function $g = g_T$ in $L^2$ such that

$$T(f) = \int_{\Omega} fg \, d\mu, \quad \forall f \in L^2,$$

and $\|T\|' = \|g\|_2$. \hfill \Box

B.5.5 Lebesgue Spaces

First, to allow explicit calculation, recall that it can be proved that every Riemann integrable function is Lebesgue measurable and both integrals coincide. Moreover, a bounded function is Riemann integrable if and only if it is continuous almost everywhere.

Now, perhaps the most typical measures are the Lebesgue measure in $\mathbb{R}^d$ and the counting measure in $\mathbb{N}$, with the corresponding $L^p = L^p(\mathbb{R}^d)$ space of Lebesgue almost everywhere measurable real-valued functions with norm

$$\|f\|_p = \int_{\mathbb{R}^d} |f(x)|^p \, dx < \infty$$

and $\ell^p = \ell^p(\mathbb{R})$ space of real-valued (or complex-valued or $\mathbb{R}^d$-valued) sequences $x = \{x_n\}$ such that

$$\|a\|_p = \|\{a_n\}\|_p = \sum_{n=1}^{\infty} |a_n|^p < \infty,$$
with $1 \leq p < \infty$. Also $L^\infty = L^p(\mathbb{R}^d)$ is the space of all Lebesgue essentially bounded (i.e., almost everywhere measurable and bounded outside of a negligible set) real-valued functions, namely
\[
\|f\|_\infty = \inf \{C > 0 : |f(x)| \leq C, \text{ a.e.} \},
\]
where the infimum is $\infty$ if the function is not bounded outside a negligible set. Similarly, $\ell^\infty = \ell^\infty(\mathbb{R}^d)$ is the space of all bounded real-valued sequences with the norm
\[
\|a\|_\infty = \|\{a_n\}\|_\infty = \sup \{|a_n| : n \geq 1\}.
\]
Certainly, we have the spaces $L^p(A)$ for any measurable non-negligible subset $A \subset \mathbb{R}^d$ (of particular interest is the case when $A = \Omega$ an open set), $L^p(\mathbb{R}^d; \mathbb{C})$ or $L^p(\mathbb{R}^d; \mathbb{R}^n)$ (functions with complex values or with values in $\mathbb{R}^n$), $\ell^p(\mathbb{R})$ or $\ell^p(\mathbb{C})$ (sequences with complex values or with values in $\mathbb{R}^n$, $n \geq 1$). Moreover, we may use $\ell^p(\mathbb{Z}; \mathbb{R})$ the space of all double-sided sequence $\{a_n : n \in \mathbb{Z}\}$, with $\mathbb{Z} = \{0, \pm 1, \ldots\}$ the integers numbers. Actually, we may replace $\mathbb{Z}$ by any countable set, or even any set of indexes $I$, where real-valued “sequences” means functions $a : I \to \mathbb{R}$ with countable support, i.e., such that $\{i \in I : a_i \neq 0\}$ is finite or countable.

As we have seen, these are Banach spaces, which are separable if $1 \leq p < \infty$. If $A$ have a finite measure then $L^p(A) \subset L^q(A)$ for any $1 \leq p < q \leq \infty$, and on the contrary, $\ell^q \subset \ell^p$. In general, $L^p \cap L^q$ is a subspace of $L^r$ for any $1 \leq p \leq r \leq q \leq \infty$.

Recall the convolution defined on $\mathbb{R}^d$ by the expression
\[
(f \ast g)(x) = \int_{\mathbb{R}^d} f(x-y) g(y) \, dy = \int_{\mathbb{R}^d} f(y) g(x-y) \, dy,
\] (B.21)
which is defined almost everywhere for any $f$ (or $g$) in $L^1$ and $g$ (or $f$) in $L^\infty$. To define the convolution we use the topological group structure $(\mathbb{R}^d, +)$. In general, if $(\Omega, +)$ is a locally compact (abelian) group then a translation-invariant Radon measure on $\Omega$ is called a Haar measure, and there is one and only one (up to a multiplicative constant) Haar measure, e.g. see Folland [44, Section 11.1, pp. 339–348] or Cohn [28, Chapter 9, pp. 297-327]. For instance, the Lebesgue measure is a Haar measure on $(\mathbb{R}^d, +)$ with the Euclidean topology and the counting measure is a Haar measure on $(\mathbb{Z}, +)$ or $(\mathbb{R}^d, +)$ with the discrete topology. Thus,
\[
(a \ast b)_n = \sum_k a_{n-k} b_k = \sum_k a_k b_{n-k}
\] (B.22)
is a discrete version of (B.21). We are more interested in the continuous case.

If $f$ and $g$ have support in $\mathbb{R}^d_+ = [0, \infty)^d$, then we have
\[
(f \ast g)(x) = \int_{(0,x)} f(x-y) g(y) \, dy = \int_{(0,x)} f(y) g(x-y) \, dy,
\]
where $(0,x) = (0,x_1) \times \cdots \times (0,x_d)$ is a bounded $d$-dimensional interval, and so, with finite measure, i.e., the convolution can be considered in $L^1_{\text{loc}}$. 

Proposition B.65 (Young Inequality). If $f$ belongs to $L^p(\mathbb{R}^d)$ and $g$ belongs to $L^q(\mathbb{R}^d)$ then $f \ast g$ belongs to $L^r(\mathbb{R}^d)$ and
\[
\|f \ast g\|_r \leq \|f\|_p \|g\|_q,
\]
provided $1 \leq p, q, r \leq \infty$ and $1/p + 1/q - 1/r = 1$.

Proof. We integrate in $y$ the expression
\[
|f(x-y)g(y)| = (|f(x-y)|^{p/r} |g(y)|^{q/r}) \times (|f(x-y)|^{p(1-p-1/r)} \times (|g(y)|^{q(1/q-1/r)})
\]
and we apply Hölder inequality with the exponents $r, p_1$ and $q_1$ satisfying $1/p_1 = 1/p - 1/r$ and $1/q_1 = 1/q - 1/r$, to obtain
\[
|(f \ast g)(x)|^r \leq \left( (|f|^p \ast |g|^q)(x) \right) \left( \|f\|_{p^{-p}} \|g\|_{q^{-q}} \right).
\]
Hence, integrating in $x$ we deduce
\[
\|f \ast g\|_r \leq \|f\|_p \|g\|_q \|f\|_{p^{-p}} \|g\|_{q^{-q}} = \|f\|_p \|g\|_q,
\]
i.e., the desired estimate, for $p, q, r$ finite.

Analogously, we treat the limiting cases when some of the exponents are infinite. \qed

The following properties proved for $L^1$ can be extended to $L^p$, with $1 \leq p < \infty$

(a) The translation is continuous in $L^p(\mathbb{R}^d)$, i.e., if $\tau_a f(\cdot) = f(\cdot + a)$ then $\|\tau_a f - f\|_p \to 0$ as $a \to 0$, for every $f$ in $L^p$. \qed

(b) The space $C_0^0$ of all continuous functions on $\mathbb{R}^d$ with compact support is dense in $L^p$, i.e., for every $\epsilon > 0$ and $f$ in $L^p(\mathbb{R}^d)$ there exists $g_\epsilon$ in $C_0^0(\mathbb{R}^d)$ such that $\|f - g_\epsilon\|_p < \epsilon$. \qed

(c) The kernel convolution converges in $L^p$, i.e., $\|f \ast k_\epsilon - f\|_p \to 0$ as $\epsilon \to 0$, for every $f$ in $L^p$. Indeed, we apply Hölder inequality to the right-hand term of
\[
|(f \ast k_\epsilon)(x) - f(x)| \leq \int_{\mathbb{R}^d} (|f(x-y) - f(x)| |k_\epsilon(y)|^{1/p}) (|k_\epsilon(y)|^{1/q}) dy,
\]
with $1/p + 1/q = 1$, and we integrate in $dx$ to obtain
\[
\|f \ast k_\epsilon - f\|_p^p \leq \|k_\epsilon\|_1^{p/q} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-y) - f(x)|^p |k_\epsilon(y)| dy.
\]
Exchanging the order of integration, and splitting the integral in $dy$ into the regions $\{|y| < \delta\}$ and $\{|y| \geq \delta\}$ we have
\[
\|f \ast k_\epsilon - f\|_p^p \leq \|k_\epsilon\|_1^{p/q} \left[ \int_{\{|y| < \delta\}} \phi(y) |k_\epsilon(y)| dy + \int_{\{|y| \geq \delta\}} \phi(y) |k_\epsilon(y)| dy \right],
\]
where
\[
\phi(y) = \int_{\mathbb{R}^d} \left| f(x - y) - f(x) \right|^p \, dx.
\]

The continuity of the translation \((a)\) shows that \(\phi(y) \to 0\) as \(|y| \to 0\), and so, for every \(\varepsilon > 0\) there exists \(\delta > 0\) such that
\[
\int_{\{|y| < \delta\}} \phi(y) |k_\varepsilon(y)| \, dy \leq \varepsilon_1 \int_{\{|y| < \delta\}} |k_\varepsilon(y)| \, dy \leq \varepsilon_1 \|k\|_1 \leq \varepsilon_1.
\]

Since \(\phi\) is bounded, i.e., \(\|\phi(y)\|_\infty \leq (2\|f\|_p)^p\), we obtain
\[
\int_{\{|y| \geq \delta\}} \phi(y) |k_\varepsilon(y)| \, dy = \int_{\{|y| \geq \delta/\varepsilon\}} \phi(\varepsilon y) |k(y)| \, dy \leq (2\|f\|_p)^p \int_{\{|y| \geq \delta/\varepsilon\}} |k(y)| \, dy,
\]

where the right-hand side tends to 0 as \(\varepsilon \to 0\). This proves that \(f \ast k_\varepsilon \to f\) in \(L^p\), as \(\varepsilon \to 0\).

Based on these properties we have

**Proposition B.66.** Let \(\Omega\) be an open subset of \(\mathbb{R}^d\) and \(C_0^\infty(\Omega)\) be the space of all real-valued functions having derivatives of any order and compact supports. Then \(C_0^\infty(\Omega)\) is dense in \(L^p(\Omega)\), for any \(1 \leq p < \infty\).

**Proof.** It is clear that we can find a sequence \(\{\Omega_n : n \geq 1\}\) of open sets with compact closure satisfying
\[
\overline{\Omega}_n \subset \Omega_{n+1} \subset \overline{\Omega}_{n+1} \subset \Omega, \quad \forall n \quad \text{and} \quad \bigcup_n \Omega_n = \bigcup_n \overline{\Omega}_n = \Omega.
\]

By means of the dominate convergence we check that
\[
\int_{\Omega} \left| 1_{\Omega_n}(x) f(x) - f(x) \right|^p \, dx \to 0,
\]
i.e., \(\|1_{\Omega_n} f - f\|_p \to 0\) as \(n \to \infty\). Hence, we are reduced to approximate functions with compact supports.

Therefore, let \(f\) be a function in \(L^p(\Omega)\) which vanishes outside of some compact set \(K = K_f \subset \Omega\). It is then clear that there exists a continuous function \(k\) with compact support inside \(\Omega\) such that \(k = 1\) on \(K\) and \(0 \leq k \leq 1\) on \(\Omega\). Now, for every \(\varepsilon > 0\) there exists a continuous function \(g_\varepsilon\) with compact support such that
\[
\int_{\mathbb{R}^d} \left| 1_K(x) f(x) - g_\varepsilon(x) \right|^p \, dx < \varepsilon,
\]
which implies that \(\|f - kg_\varepsilon\|_p < \varepsilon\). Actually, by means of a convolution with a smooth kernel, we can choose \(kg_\varepsilon\) in \(C_0^\infty(\Omega)\). \(\square\)
Proposition B.67. If $1 \leq p < \infty$ and $A$ is a measurable subset of $\mathbb{R}^d$ then $L^p(A)$ is separable Banach space.

Proof. It is clear that only the case $A = \mathbb{R}^d$ needs consideration. Indeed, for any function in $L^p(A)$ can be extended by zero to be obtain an element in $L^p(\mathbb{R}^d)$ and backward, any function $f$ in $L^p(\mathbb{R}^d)$ becomes a function in $L^p(A)$ by setting $g = 1_A f$, which is a continuous linear transformation.

There are several ways to check that $L^p = L^p(\mathbb{R}^d)$ is separable. For instance, we may consider functions of the form $p(x)1_B$ where $p$ are polynomials with rational coefficients and $B$ are closed balls centered at the origin of radius $1/n$, for $n = 1, 2, \ldots$.

Alternatively, we may consider simple functions of the form $\sum_{j=1}^n a_j 1_{A_j}$, where $a_j$ are rational numbers and $\{A_j\}$ are disjoint $d$-intervals with rational extremes, i.e., of the form $\prod_{i=1}^d [\alpha_i, \beta_i]$, with $\alpha_i$ and $\beta_i$ rational numbers. It is clear that any simple function can be approximate in the $L^p$-norm with simple functions of the above form.

• Remark B.68. It is clear some of the arguments used in the above Proposition B.67 can be applied to any Radon measure $(\mathcal{F}, \mu)$ in $\mathbb{R}^d$, so that $L^p(\mathbb{R}^d, \mathcal{F}, \mu)$ is a separable Banach space.

The particular case $L^2(A)$ or $L^2(A; \mathbb{C})$ is a real or complex separable Hilbert space with the scalar or inner product

$$(f, g) = \int_A f(x)g(x) \, dx \quad \text{or} \quad (f, g) = \int_A f(x)\bar{g}(x) \, dx,$$

where $\bar{g}$ means the complex-conjugate. We denote by $\| \cdot \| = \| \cdot \|_2$ the corresponding norm.

The following definitions apply to any Hilbert space, but we focus in $L^2$. A family of functions $\{\varphi_i : i \in I\}$ is orthogonal if $(\varphi_i, \varphi_j) = 0$ for every $i \neq j$; it is orthonormal if also $\|\varphi_i\| = 1$, for every $i$; and it is called complete if the only function orthogonal to any $\varphi_i$ is the zero, i.e., if $(f, \varphi_i) = 0$ for every $i$ implies $f = 0$. The (finite) linear combinations of elements in the family is called the span, and a family of functions $\{\varphi_i : i \in I\}$ is called a basis if its span is dense in $L^2$.

Proposition B.69. There exists a complete orthonormal basis for $L^2$. Moreover, any orthonormal basis is countable and complete.

Proof. If $\{\varphi_i : i \in I\}$ is an orthonormal basis then

$$\|\varphi_i - \varphi_j\|^2 = (\varphi_i - \varphi_j, \varphi_i - \varphi_j) = \|\varphi_i\|^2 + \|\varphi_j\|^2 = 2,$$

for any $i \neq j$. Because $L^2$ is separable, the set of indices $I$ can be at most countable.

If $\{\varphi_i : i \geq 1\}$ is a orthogonal basis and $(f, \varphi_i) = 0$ for every $i$ then $(f, \varphi) = 0$ for any $\varphi$ linear combination of elements in the basis, and so

$$\|f\|^2 = (f, \bar{f} - \bar{\varphi}) \leq \|f\| \|f - \varphi\|.$$
Since linear combinations are dense in \( L^2 \), the quantity \( \|f - \varphi\| \) can be made arbitrary small, which implies that \( f = 0 \), i.e., \( \{\varphi_i : i \geq 1\} \) is complete.

Finally, we apply the Gram-Schmidt procedure to a countable dense set \( \{\phi_i : i \geq 1\} \) to obtain an orthonormal family \( \{\varphi_i : i \geq 1\} \), which is a basis by construction. Thus, we get a complete orthonormal family or system or basis.

It is clear that we have proved that any separable Hilbert space has a (countable) complete orthonormal basis \( \{\varphi_i : i \geq 1\} \).

Recall that \( \ell^2(\mathbb{R}) \) or \( \ell^2(\mathbb{C}) \) is the space of all real-valued or complex-valued sequences \( a = \{a_i : i \geq 1\} \) such that \( \|a\|_2 = \left( \sum_{i \geq 1} |a_i|^2 \right)^{1/2} \) is finite, which is a separable Hilbert space with the scalar or inner product \( (a, b) = \sum_{i \geq 1} a_i \bar{b}_i \).

**Proposition B.70.** Let \( \{e_i : i \geq 1\} \) be a complete orthonormal basis in a separable Hilbert space \( H \), e.g., \( H = L^2 \), with norm \( \|\cdot\| \) and inner product \( (\cdot, \cdot) \). Then for any given element \( h \) in \( H \) the series \( h_n = \sum_{i=1}^{n} (h, \bar{e}_i) e_i, n \geq 1 \), converges to \( h \) and Parseval’s formula

\[
\|h\|^2 = \sum_{i=1}^{\infty} |(h, e_i)|^2, \quad \forall h \in H,
\]

holds. Moreover, the mapping \( T : H \to \ell^2 \) defined by \( T(h) = \{(h, \bar{e}_i) : i \geq 1\} \) is a linear isometry.

**Proof.** By means of the linearity of the inner product we have

\[
\|h_n - h_m\|^2 = \sum_{i=m+1}^{n} |(h, e_i)|^2 \quad \text{and} \quad \|h - h_n\|^2 = \|f\|^2 - \sum_{i=1}^{n} |(h, e_i)|^2,
\]

which proves that the sequence of partial sum \( \{h_n : n \geq 1\} \) is convergent to some function \( g \) in \( L^2 \). Since \( h - g \) is orthogonal to any \( e_i \), we deduce that \( h = g \), \( \|h_n - h\| \to 0 \) and Parseval’s formula holds.

It is clear that \( T \) is linear and that \( T^{-1}(a) = \sum_{i \geq 1} a_i e_i \). Also, the parallelogram identity \( \|h + g\|^2 + \|h - g\|^2 = 2|(h, g) + (g, h)| \) shows that

\[
(T(h), T(g)) = \sum_{i \geq 1} (h, e_i)(g, e_i), \quad \forall h, g \in H,
\]
i.e., \( T \) preserves the inner product. \( \square \)

Perhaps the reader may want to take a look at the book Lieb and Loss [80, Chapters 1 and 2, pp. 1–77] for a concrete review on the previous material. Also, plenty of exercises can be found in the book by Gelbaum [51].

[ Preliminary ]

Menaldi November 11, 2016
B.5.6 Radon-Nikodym Derivative

Because measures can take infinite values, subtraction two measures is only allowed when at least one of them is finite. Thus a signed measure $\nu$ is a $\sigma$-additive set function on a measurable space $(\Omega, \mathcal{F})$ such that $\nu(\emptyset) = 0$. The $\sigma$-additivity implies that $\nu$ takes values in either $[-\infty, +\infty]$ or $(-\infty, +\infty]$; moreover, if $A = \sum_{i=1}^{\infty} A_i$ with $A_i \in \mathcal{F}$ and $|\nu(A)| < \infty$ then, by separating the positive and the negative terms, we deduce that the series $\sum_{i=1}^{\infty} \nu(A_i)$ is absolutely convergence. Also, for any $E \subset F$ measurable sets, the relation $\nu(F) = \nu(E) + \nu(F \setminus E)$ shows that if $|\nu(F)| < \infty$ then $|\nu(E)| < \infty$ (i.e., finite values can only be obtained by adding or subtraction real numbers. Hence it makes sense to say that a signed measure $\nu$ is finite if $|\nu(\Omega)| < \infty$, and similarly we define $\sigma$-finite signed measures.

The $\sigma$-additivity property applied to finite measures can be considered in a larger context, e.g., we may discuss measures with complex values (in $\mathbb{C}$) or with vector values in $\mathbb{R}^d$ or even more general with values in a topological vector space (usually a Banach space or a locally convex space). Hahn-Jordan decomposition affirms that if $\nu$ is a signed measure on $(\Omega, \mathcal{F})$ then there exists a measurable set $A \in \mathcal{F}$ such that $\nu^+(F) = \nu(F \cap A)$ and $\nu^-(F) = -\nu(F \cap A')$ are measures satisfying $\nu(F) = \nu^+(F) - \nu^-(F)$, for every $F \in \mathcal{F}$. Certainly, the set $A$ is not necessarily unique, but the positive and negative variations measures $\nu^+$ and $\nu^-$ are uniquely defined.

Also we define the measure $|\nu|(A) = \nu^+(A) + \nu^-(A)$, which is called the variation of $\nu$. Note that a signed measure $\nu$ is finite (i.e., $|\nu(\Omega)| < \infty$) if and only if $|\nu|$ is so, (i.e., $|\nu|(\Omega) < \infty$), and similarly for the concept of $\sigma$-finite.

**Definition B.71.** Let $\mu$ and $\nu$ be two signed measures on a measurable space $(\Omega, \mathcal{F})$. The signed measure $\nu$ is said to be absolutely continuous with respect to $\mu$ and written $\nu \ll \mu$ if for every $F \in \mathcal{F}$ with $|\mu|(F) = 0$ we also have $\nu(F) = 0$. On the contrary, these two measures $\mu$ and $\nu$ are called (mutually) singular and written $\mu \perp \nu$ (or $\nu \perp \mu$) if there exits $A \in \mathcal{F}$ such that $|\mu|(A) = 0$ and $|\nu|(\Omega \setminus A) = 0$.

It is clear that being singular is a symmetric property, while being absolutely continuous is not. Moreover, $\mu \perp \nu$ if and only if there exits $A \in \mathcal{F}$ such that for every $F \in \mathcal{F}$ we have

$$F \cap A = \emptyset \Rightarrow \mu(F) = 0 \quad \text{and} \quad F \subset A \Rightarrow \nu(F) = 0,$$

i.e., $\nu = 0$ on $A$ and $\mu = 0$ on $\Omega \setminus A$. Similarly, $\nu \ll \mu$ if an only if for every $F \in \mathcal{F}$ such that $\mu(E \cap F) = 0$, for every $E \in \mathcal{F}$, we have $\nu(E \cap F) = 0$, for every $E \in \mathcal{F}$.

If $f$ is a quasi-integrable function in $(\Omega, \mathcal{F}, \mu)$, i.e., $f = f^+ - f^-$ is measurable and either $f^+$ or $f^-$ is integrable, then the expression

$$F \mapsto \int_{F} f \, d\mu, \quad \forall F \in \mathcal{F}$$
defines a signed measure which is absolutely continuous with respect to \( \mu \). The converse is precisely the Radon-Nikodym Theorem, and the Lebesgue decomposition completes the argument, namely, any \( \sigma \)-finite signed measure \( \nu \), on a \( \sigma \)-finite measure space \(( \Omega, \mathcal{F}, \mu)\), can be written as \( \nu = \nu_a + \nu_s \), where \( \nu_a \ll \mu \) and \( \nu_s \perp \mu \).

**Theorem B.72.** Let \(( \Omega, \mathcal{F}, \mu)\) be a \( \sigma \)-finite measure space. Suppose that \( \nu \) is a \( \sigma \)-finite signed measure on \(( \Omega, \mathcal{F})\), which is absolutely continuous with respect to \( \mu \). Then there exists a quasi-integrable function \( f \) such that

\[
\nu(F) = \int_F f \, d\mu, \quad \forall F \in \mathcal{F},
\]

where the function \( f \) is uniquely defined except in a set of \( \mu \)-measure zero.

**Proof.** First note that by means of the Hahn-Jordan decomposition, we can write \( \nu = \nu^+ - \nu^- \), which effectively reduces the problem to the case of a \( \sigma \)-finite measure \( \nu \). Now, we proceed in several steps:

1. **(Step 1)** Since \( \nu \) is \( \sigma \)-finite, the whole space \( \Omega \) can be written as a disjoint countable union \( \bigcup_n \Omega_n^\nu \) with \(|\nu(\Omega_n^\nu)| < \infty \). Next, because \( \mu \) is also \( \sigma \)-finite, each \( \Omega_n^\nu \) can be written as a disjoint countable union \( \bigcup_k \Omega_n^{\mu,k} \) with \( |\mu(\Omega_n^{\mu,k})| < \infty \). Hence, relabeling the double sequence, we have \( \Omega = \bigcup_n \Omega_n \), with \( \Omega_n \cap \Omega_m = \emptyset \) if \( n \neq m \) and \(|\nu_n(\Omega_n)| + |\mu(\Omega_n)| < \infty \), for every \( n \). Therefore, it suffices to show the results for the case where \( \nu \) and \( \mu \) are finite measures.

2. **(Step 2)** If \( G \) is the class of nonnegative \( \mu \)-integrable functions \( g \) such that

\[
\nu(F) \geq \int_F g \, d\mu, \quad \forall F \in \mathcal{F},
\]

then there exits a function \( f \) in \( G \) such that

\[
\int_\Omega f \, d\mu = \sup_{g \in G} \int_\Omega g \, d\mu.
\]

Indeed, first note that if \( g_1 \) and \( g_2 \) belongs to \( G \) then \( g_1 \lor g_2 \) also belongs to \( G \). Thus, if \( \{g_n\} \) is a maximizing sequence then \( f_n = \max\{g_1, \ldots, g_n\} \) defines an increasing sequence in \( G \) such that

\[
\lim_n \int_\Omega f_n = \sup_{g \in G} \int_\Omega g \, d\mu.
\]

The monotone convergence theorem ensures that \( f = \lim_n f_n \) belongs to \( G \) and provides a maximizer.

3. **(Step 3)** If \( \lambda \neq 0 \) is a measure absolutely continuous with respect to \( \mu \) then there exists \( \varepsilon > 0 \) and \( A \in \mathcal{F} \) with \( \nu(A) > 0 \) such that \( \lambda(F \cap A) \geq \varepsilon \mu(F \cap A) \), for every \( F \in \mathcal{F} \). Indeed, let \( A_k \) the Hahn decomposition of the signed measure \( \lambda_k = \lambda - (1/k)\mu \), i.e., \( \lambda_k(F \cap A_k) \geq 0 \geq \lambda_k(F \setminus A_k) \), for every \( F \in \mathcal{F} \). Set \( A_0 = \bigcup_k A_k \) and \( B_0 = \bigcap_k B_k \), with \( B_k = \Omega \setminus A_k \). Since \( 0 \leq \lambda(B_0) \leq (1/k)\mu(B_0) \)
we have $\lambda(B_0) = 0$, and because $\lambda$ is nonzero and $A_0 = \Omega \setminus B_0$, we deduce $\lambda(A_0) > 0$, i.e., there exists $k$ such that $\lambda(A_k) > 0$. Hence, we choose $A = A_k$ and $\varepsilon = 1/k$, for this particular $k$.

(Step 4) To complete the proof we show that the measure

$$
\lambda(F) = \nu(F) - \int_F f \, d\mu, \quad \forall F \in \mathcal{F}
$$

vanishes. To this purpose, assume $\lambda \neq 0$ and get a contradiction. Because $\nu \ll \mu$ implies $\lambda \ll \mu$, we can use (Step 3) to get a measurable set $A$ and a $\varepsilon > 0$ such that $\nu(A) > 0$ such that $\lambda(F \cap A) \geq \varepsilon \mu(F \cap A)$, for every $F \in \mathcal{F}$. Choose $h = f + \varepsilon 1_A$ to get

$$
\int_F h \, d\mu = \int_F f \, d\mu + \varepsilon \mu(F \cap A) \leq \int_F f \, d\mu + \lambda(F \cap A) =
$$

$$
= \int_{F \setminus A} f \, d\mu + \nu(F \cap A) \leq \nu(F \setminus A) + \nu(F \cap A) = \nu(F),
$$

which shows that $h$ belongs to the class $\mathcal{G}$ and

$$
\int_\Omega h \, d\mu = \int_\Omega f \, d\mu + \varepsilon \mu(A) > \int_\Omega f \, d\mu,
$$

i.e., a contradiction. \hfill \Box

Sometimes, the function $f$ satisfying the conditions of Theorem B.72 is denoted by $\frac{d\nu}{d\mu}$ and called the Radon-Nikodym derivative.

- Remark B.73. It is simple to show that for any $\mu$ and $\nu$ are two finite measures on $(\Omega, \mathcal{F})$ we have $\nu \ll \mu$ if and only if for every $\varepsilon > 0$ there exist $\delta > 0$ such that $F \in \mathcal{F}$ and $\mu(F) < \delta$ imply $\nu(F) < \varepsilon$. Indeed, by contradiction, suppose that for some $\varepsilon > 0$ there is a sequence $\{F_n\}$ of measurable sets such that $\mu(F_n) < 2^{-n}$ and $\nu(F_n) \geq \varepsilon$. If $F_0 = \bigcap_{n} \bigcup_{k \geq n} F_k$ then we deduce $\mu(F_0) \leq \sum_{k \geq n} 2^k = 2^{n-1}$ and $\nu(F_0) \geq \varepsilon$, i.e., $\mu(F_0) = 0$ and we obtain a contradiction. \hfill \Box

A generalization of Radon-Nikodym arguments yields the so-called Lebesgue Decomposition: If $\mu$ and $\nu$ are a two $\sigma$-finite signed measures on a measurable space $(\Omega, \mathcal{F})$ then there exist two $\sigma$-finite signed measures $\nu_a$ and $\nu_s$ such that $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$ and $\nu_s \perp \mu$. Clearly, the pair $\nu_a, \nu_s$ is uniquely determinate.

## B.6 Essential Complements

It may be convenient to include a short discussion on change of variables for the Lebesgue integral as well as a quick presentation on the Lebesgue measure on manifolds.
B.6.1 Change of Variables

Spherical coordinates can be used in $\mathbb{R}^d$, i.e., every $x$ in $\mathbb{R}^d \setminus \{0\}$ can be written uniquely as $x = r \, x'$, where $0 < r < \infty$ and $x'$ belongs to $S^{d-1} = \{ x \in \mathbb{R}^d : |x| = 1 \}$.

**Theorem B.74.** The Lebesgue measure $dx$ in $\mathbb{R}^d$ can be expressed as a product measure $dr \times dx'$, where $dr$ is the Lebesgue measure on $(0, \infty)$ and $dx'$ is a (surface) measure on $S^{d-1}$. Moreover, for every nonnegative measurable function in $\mathbb{R}^d$ we have

$$\int_{\mathbb{R}^d} f(x) \, dx = \int_{(0, \infty) \times S^{d-1}} f(rx') \, r^{d-1} \, dr \times dx'. \tag{B.23}$$

In particular, if $f$ is homogeneous, i.e., $f(x) = g(|x|)$, then

$$\int_{\mathbb{R}^d} f(x) \, dx = \omega_{d-1} \int_0^\infty g(r) r^{d-1} \, dr,$$

where the value $\omega_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of the unit ball, i.e., $dx'(S^{d-1})$.

**Proof.** It is clear that for $d = 2$ (or $d = 3$) this is call polar (or spherical) coordinates. Moreover, the crucial point is to define the surface measure $dx'$ on $S^{d-1}$, which will agree with the $(d-1)$-dimensional superficial measure in $\mathbb{R}^d$ (i.e., Hausdorff measure, except for a multiplicative constant).

It is clear that

$$\Upsilon : \mathbb{R}^d \setminus \{0\} \to (0, \infty) \times S^{d-1}, \quad \Upsilon(x) = (r, x'), \quad r = |x|, x' = x/|x|$$

is a continuous bijection mapping with $\Upsilon^{-1}(r, x') = rx'$. Then, given a Borel set $B$ in $S^{d-1}$ we define $B_a = \{rx' : x' \in B, r \in (0, a]\}$, i.e., $B_a = \Upsilon^{-1}([0, a] \times E)$. Thus, for the desired surface measure $dx'$ we must satisfies (B.23) for $f = 1_{E_1}$, i.e.,

$$\ell(E_1) = \int_{\mathbb{R}^d} 1_{E_1}(x) \, dx = dx'(E) \int_0^1 r^{d-1} \, dr,$$

and therefore we can define $dx'(E) = dx(E_1)$, which results a measure on $S^{d-1}$. On the other hand, Theorem B.30 shows that $dx(E_a) = a^d dx(E_1)$ and thus

$$dx([a, b] \times E) = dx(E_b \setminus E_a) = \frac{b^d - a^d}{d} dx'(E) = dx'(E) \int_a^b r^{d-1} \, dr,$$

i.e., with $dx'$ defined as above, we have the validity of equality (B.23) for any function $f = 1_{[a, b] \times E}$. We conclude approximating any nonnegative measurable function by a sequence of simple functions. \[\square\]

\[^{1}\text{Recall the Gamma function } \Gamma(x) \text{ satisfying } \Gamma(n+1) = n(n - 1)\ldots1, \text{ for any integer } n, \text{ and } \Gamma(1/2) = \sqrt{\pi} \]
For instance, the interested reader may consult the book by Folland [44, Section 2.7, pp. 77–81] for more details. Note that

$$\int_{\{x \in \mathbb{R}^d : |x| \leq r\}} \text{d}x = \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} r^d$$

and

$$\int_{\{x \in \mathbb{R}^d : |x| = r\}} \text{d}x' = 2 \frac{\pi^{d/2}}{\Gamma(d/2)} r^{d-1}$$

are the volume and the surface area of a ball radius $r$.

**Remark B.75.** This change-of-variables yields that the function $x \mapsto |x|^{-\alpha}$ is Lebesgue integrable (a) on the unit ball $B = \{x \in \mathbb{R}^d : |x| < 1\}$ if an only if $\alpha < d$ and (b) outside the unit ball $\mathbb{R}^d \setminus B$ if and only if $\alpha > d$.

More general, we have

**Theorem B.76 (Change of variable).** Let $X$ and $Y$ be open subsets of $\mathbb{R}^d$ and $T : X \to Y$ be a homeomorphism of class $C^1$. A function $y \mapsto f(y)$ is Lebesgue measurable on $(Y, \mathcal{L}_y, dy)$ is and only if $x \mapsto f(T(x))$ is Lebesgue measurable on $(X, \mathcal{L}_x, dx)$. In this case, we have

$$\int_Y f(y) \text{d}y = \int_X f(T(x)) J_T(x) \text{d}x,$$

where $J_T(x) = |\det(\partial_x T(x))|$ denotes the Jacobian of $T$.

Based on Theorem B.30, we can easily prove the change of variable formula for an affine transformation $T$. Indeed, it suffices to approximate $f$ by a sequence of simple functions. Some more preparation is required for a nonlinear homeomorphism of class $C^1$, e.g., see Ambrosio et al. [8, Chapter 8, pp. 129–136] or Jones [70, Chapter 15, pp. 494–510] or Knapp [74, Section VI.5, pp. 320–326] or Schilling [111, Chapter 15, pp. 142–162]. Actually, essentially with the same arguments, we can prove the following estimate: For any function $T : X \to \mathbb{R}^d$ with $X$ an open subset of $\mathbb{R}^d$, and for any set $E \subset X$ where $T$ is differentiable at every point of $E$, we have

$$\ell^*(T(E)) \leq (\sup_{E} J_T) \ell^*(E),$$

where $\ell^*$ denotes the Lebesgue outer measure on $\mathbb{R}^d$. This implies Sard’s Theorem, i.e., the set of point $x$, where the function $T(x)$ is differentiable and the Jacobian $J_T(x) = 0$, is indeed negligible. Moreover, if $T$ is a measurable function from an open set $X \subset \mathbb{R}^d$ into $\mathbb{R}^d$, i.e., $T : X \to \mathbb{R}^d$, which is differentiable at every point of a measurable set $E \subset \mathbb{R}^d$ then

$$\ell^*(T(E)) \leq \int_E J_T(x) \text{d}x,$$

which implies a one-side inequality $\leq$ in the Theorem B.76, under the sole assumption that $T$ is only differentiable and $f$ is a nonnegative Borel function. The reader may take a look at Cohn [28, Chapter 6, pp. 167–195] for a carefully discussion, and to Duistermaat and Kolk [37, Chapter 6, pp. 423–486] for a number of details in the change of variable formula for the Riemann integral.
By means of the change of variables formula, we can define the surface measure of a $n$-dimensional $C^1$-manifold $M$ with local coordinates chart $T : \mathcal{O} \to \mathbb{R}^n$ and metric tensor given locally by a positive definite matrix $a = (a_{ij})$, $a_{ij} = (\partial_k T_i)(\partial_k T_j)$. Indeed, the expression

$$\mu(\mathcal{O}) = \int_{T(\mathcal{O})} \sqrt{\det(a(x))} \, dx,$$

$\forall \mathcal{O}$ open subset of $M$

is well defined and invariant within the manifold. For instance, if $M$ is the graph of a real-valued continuously differentiable function $y = u(x)$ with $x \in \Omega \subset \mathbb{R}^n$, then $M$ is an $n$-dimensional manifold in $\mathbb{R}^{n+1}$ and the map $T(x) = (u(x), x)$ provides a natural (local) coordinates with metric tensor given locally by the matrix $a_{ij} = \delta_{ij} + \partial_i u \partial_j u$. Thus $\sqrt{\det(a(x))} = \sqrt{1 + |\nabla u(x)|^2}$, and

$$\mu(M) = \int_{\Omega} \sqrt{1 + |\nabla u(x)|^2} \, dx$$

is the surface measure of $M$, in particular this is valid for the unit sphere $S^{n-1} = \{ x \in \mathbb{R}^n : |x| = 1 \}$.

For instance, the interested reader may consult Taylor [124, Chapter 7, pp. 83–106] for more details. In general, the reader may take a look at the textbooks by Apostol [9, Chapters 14 and 15, pp. 388–433] and Duistermaat and Kolk [38], for a detail account of the multidimensional Riemann integral.

In a more delicate setting, the Lebesgue measure represents the volume in $\mathbb{R}^d$ while the length and surface area are given by the Hausdorff measure, except for a factor. Recall that on the Borel space $\mathbb{R}^d$ we denote by $\ell_n = c_n \mathbb{R}_n$, where $c_n = 2^{-n} \pi^{n/2} / \Gamma(n/2 + 1)$, i.e., the $n$-dimensional surface measure, and $\ell_d$ is the Lebesgue measure.

Let us recall the polar decomposition of a linear mapping $T : \mathbb{R}^d \to \mathbb{R}^n$ into an symmetric linear map $\mathbb{R}^{d \wedge n} \to \mathbb{R}^{d \wedge n}$ and an orthogonal linear map $H : \mathbb{R}^{d \wedge n} \to \mathbb{R}^{d \wedge n}$ such that $T = HS$ if $d \leq n$ and $T = SH^*$ if $d \geq n$. Thus, the Jacobian $J(T)$ is defined as $J(T) = |\det(S)|$, the determinant (with positive sign) of the symmetric (square) part of $T$. Next, based on Rademacher’s Theorem, the differential $DF$ of a given Lipschitz mapping $f : \mathbb{R}^d \to \mathbb{R}^n$ exits as a linear map $DF(x) : \mathbb{R}^d \to \mathbb{R}^n$, $\ell_d$-almost every $x$. Hence, the Jacobian of $f$ is defined as the Jacobian of its differential $DF$ (as a linear map), i.e., $J(f, x) = J(DF(x))$.

**Theorem B.77.** Let $f : \mathbb{R}^d \to \mathbb{R}^n$ be a Lipschitz function. Then for every $\ell_d$-measurable set $E \subset \mathbb{R}^d$ the mapping $y \mapsto \ell_{d-n}(E \cap f^{-1}(y))$ is $\ell_n$-measurable, we have the co-area formula

$$\int_E J(f, x) \, dx = \int_{\mathbb{R}^n} \ell_{d-n}(E \cap f^{-1}(y)) \, dy,$$

when $d \geq n$,

and the area formula

$$\int_E J(f, x) \, dx = \int_{\mathbb{R}^n} \ell_0(E \cap f^{-1}(y)) \, dy,$$

when $d \leq n$,

where $dx = d\ell_d(x)$ and $dy = d\ell_n(y)$, for $x$ in $\mathbb{R}^d$ and $y$ in $\mathbb{R}^n$. 

[Preiliminary]

Menaldi

November 11, 2016
It is clear that the area formula is used for the length of a curve ($d = 1$, $n \geq 1$), surface area of a graph or surface area of a parametric hypersurface ($d \geq 1$, $n = d + 1$), and in general for submanifolds.

The both formulae generalize to change of variables, i.e., if $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is an $\ell_d$-integrable function then

$$\int_{\mathbb{R}^d} g(x) J(f, x) \, dx = \int_{\mathbb{R}^n} \left[ \sum_{x \in f^{-1}\{y\}} g(x) \right] \, dy, \quad \text{when } d \leq n$$

and, the restriction of $g$ to $f^{-1}\{y\}$, denoted by $g|_{f^{-1}\{y\}}$, is $\ell_{d-n}$-integrable for $\ell_n$-almost every $y$ and

$$\int_{\mathbb{R}^d} g(x) J(f, x) \, dx = \int_{\mathbb{R}^n} dy \int_{f^{-1}\{y\}} g(x) \ell_{d-n}(dx), \quad \text{when } d \geq n.$$

Note that $f^{-1}\{y\}$ is a closed set in $\mathbb{R}^d$ for every $y \in \mathbb{R}^n$.

The co-area formula can be used to compute level sets and polar (or spherical) coordinates, e.g., if $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is integrable then

$$\int_{\mathbb{R}^d} g(x) \, dx = \int_0^\infty dr \int_{\partial B(0, r)} g(x) \ell_{d-1}(dx),$$

where $\partial B(0, r)$ is the boundary (sphere) of the ball $B(0, r)$ with radius $r$ and center at the origin 0, and again, we have

$$\int_{\partial B(0, r)} g(x) \ell_{d-1}(dx) = \int_{S^{d-1}} g(rx') r^{d-1} \, dx',$$

where $S^{d-1} = \partial B(0, 1)$ is the unit sphere in $\mathbb{R}^d$ and $dx' = \ell_{d-1}(dx')$. In spherical coordinates this means

$$\int_{\Omega} f(x) \ell_d(dx) = \int_0^\infty dr \int_{\{x \in \Omega : |x| = r \}} f(x) \ell_{d-1}(dx) =$$

$$= \int_0^\infty r^{d-1} dr \int_{\{x' \in \mathbb{R}^d : |x'| = 1 \}} f(rx') \mathbb{1}_{\{rx' \in \Omega\}} \ell_{d-1}(dx'). \quad \text{(B.24)}$$

Moreover, the center of the spherical coordinates may be different from the origin 0. For instance, a prove of what was mentioned in this subsection can be found Evans and Gariepy [43] or Lin and Yang [82].

The interested reader may also check the Appendix C in Leoni [79, pp 543-579] for a quick refresh on Lebesgue and Hausdorff measure (and integration), in particular, if $A$ and $B$ are two measurable sets in $\mathbb{R}^d$ such that $A + B = \{a + b : a \in A, \ b \in B\}$ is also measurable then Brunn-Minkowski’s inequality reads as

$$(\ell_d(A))^{1/d} + (\ell_d(B))^{1/d} \leq (\ell_d(A + B))^{1/d}. \quad \text{(B.25)}$$
This estimate in turn can be used to deduce the isodiametric inequalities. Moreover, if we denote by $\ell^*_d$ the Lebesgue outer measure in $\mathbb{R}^d$ then the reader may find details (e.g., Stroock [118, Section 4.2, pp. 74-79]) on proving the so-called isodiametric inequality (see Remark B.32)

$$\ell^*_d(A) \leq \omega_d(r(A))^d, \quad \forall A \subset \mathbb{R}^d,$$

where $\omega_d = \pi^{d/2}/\Gamma(d/2+1)$ is the Lebesgue measure of the unit ball in $\mathbb{R}^d$, and $r(A)$ is the radius of $A$, i.e., $r(A) = \sup\{|x-y|/2 : x, y \in A\}$.

For a later use, the above co-area formulae can be summarized as

$$\int_{\Omega} f(x)|\nabla \varrho(x)| \, dx = \int_{\mathbb{R}} ds \int_{\varrho^{-1}(s)}^{} f(x) \ell_{d-1}(dx) \quad (B.26)$$

where $\ell_{d-1}$ is the $(d-1)$-dimensional Hausdorff (Lebesgue) measure in $\mathbb{R}^d$, $\Omega$ is an open subset of $\mathbb{R}^d$ and $\varrho$ is a real-valued Lipschitz function defined on $\Omega$. More general, if $\varrho = (\varrho_1, \ldots, \varrho_n)$ is a Lipschitz function defined on $\Omega$ with values in $\mathbb{R}^n$, for some $n = 1, \ldots, d-1$, then the formula (B.26) becomes

$$\int_{\Omega} f(x)\sqrt{\nabla \varrho^* \nabla \varrho} \, dx = \int_{\mathbb{R}^n} ds \int_{\varrho^{-1}(s)}^{} f(x) \ell_{d-n}(dx), \quad (B.27)$$

where the Jacobian $J(\rho, x) = \sqrt{\nabla \varrho^* \nabla \varrho}$ is written in term of the $n \times n$ square matrix $\nabla \varrho^* \nabla \varrho = (\sum_{k=1}^{d} \partial_k \varrho_i \partial_k \rho_j)$.

### B.6.2 Lebesgue Measure on Manifolds

First we recall the concept of manifold. If $U$ and $V$ are two open sets in $\mathbb{R}^d$ then a bijective mapping $\Phi : U \rightarrow V$ which is continuously differentiable up to the order $k$ together with its inverse $\Phi^{-1} : V \rightarrow U$ is called a homeomorphism of class $C^k$ (or a $C^k$ diffeomorphism). If $k = 0$ then $\Phi$ and its inverse are just continuous, and a (locally) Lipschitz homeomorphism (or a (local) bi-Lipschitz mapping) is when $\Phi$ and $\Phi^{-1}$ are both (local) Lipschitz continuous functions, i.e., for some constant $C \geq c > 0$,

$$c|x-y| \leq |\Phi(x) - \Phi(y)| \leq C|x-y|, \quad \forall x, y \in U$$

or if the ‘locally’ prefix is used, for any $x$ and $y$ in $K$, for any compact set $K \subset U$, where the constants $C$ and $c$ may depend on $K$. Also the case $k = \infty$ (i.e., continuously differentiable of any order) is included. In this context, a homeomorphism is also called a (local) change-of-variables or coordinates.

**Definition B.78.** A set $S \subset \mathbb{R}^d$ is called a $C^k$ submanifold of $\mathbb{R}^d$ at $x \in S$ of dimension $1 \leq m < d$ if there exists an open neighborhood $U$ of $x$ such that $S \cap U$ is the graph of a mapping $\psi$ of class $C^k$ from an open set $V \subset \mathbb{R}^m$ into $\mathbb{R}^{d-m}$, i.e., for some orthogonal change-of-variables $y = (y_1, \ldots, y_d)$, $y' = (y_1, \ldots, y_m)$,

$$S \cap U = \{(y', \psi(y')) \in \mathbb{R}^d : y' \in V \subset \mathbb{R}^m\},$$
and \( \psi \) is continuously differentiable up to the order \( k \). If this property holds for every \( x \) in \( S \), with the same constants \( k \) and \( m \), but possibly a different choice of the orthogonal coordinates (and \( \psi \)), then \( S \) is called a \( C^k \) submanifold (of \( \mathbb{R}^m \)) of dimension \( m \). The \( m \)-dimensional linear space of all tangent vectors, i.e., the graph of the \((d-m) \times m\) matrix gradient \( \nabla \psi \), namely,

\[
\text{graph}(\nabla \psi(x')) = \{(y', \nabla \psi(x')y') : y' \in \mathbb{R}^m\}, \quad \text{with } x = (x', \psi(x')),
\]
is called the tangent space at the point \( x \). The strictly positive function

\[
y = (y', \psi(y')) \mapsto J_\psi(y') = \sqrt{\det(\nabla \phi(y')^* \nabla \phi(y'))}, \quad \phi(y') = (y', \psi(y'))
\]
defined on \( S \cap U \) is called the Euclidean \( m \)-dimensional density function, where \((\cdot)^*\) means the transposed matrix, and \( \det(\cdot) \) is the determinant of a \( m \times m \) matrix. With obvious changes, continuous submanifolds \((k = 0)\), \( C^\infty \) submanifolds, and (locally) Lipschitz submanifolds \((\psi \text{ is locally Lipschitz})\) are also defined. For (locally) Lipschitz submanifolds, the tangent space and the Euclidean density may not be defined at every points.

Similarly, any open subset of \( \mathbb{R}^d \) and any point in \( \mathbb{R}^d \) can be regarded as submanifolds of dimension \( m = d \) and \( m = 0 \), respectively. Certainly, instead of calling \( S \) a \( C^k \) submanifold (of \( \mathbb{R}^m \)) of dimension \( m \), we may call \( S \) a manifold of dimension \( m \) (in \( \mathbb{R}^d \)). A typical example of a \( C^\infty \) manifold is the sphere \( S^{d-1} = \{x \in \mathbb{R}^d : |x| = 1\} \). Indeed, for any \( x^0 \) in \( S^{d-1} \) there is at least one coordinate nonzero, e.g., \( x^0 > 0 \), and so,

\[
x_1 = \psi(x_2, \ldots, x_d) = \sqrt{1 - x_2^2 - \cdots - x_d^2}
\]
yields a local description. It is clear from the definition that a homeomorphism \( \Phi \) of class \( C^k \) preserves submanifolds, i.e., if \( S \) is a manifold then \( \Phi(S) \) is also a manifold.

\textbf{Remark B.79.} The mapping \( \phi(y') = (y', \psi(y')) \) from \( V \) into \( \mathbb{R}^d \) is injective and of class \( C^k \), and its inverse \( (y', \psi(y')) \mapsto y' \) is necessarily continuous (actually, of class \( C^k \)) and the \( d \times m \) matrix gradient \( \nabla \phi = (I_m, \nabla \psi)^* \) is injective. Similarly, the mapping \( g(y) = g(y', r) = r - \psi(y') \) from \( U \) into \( \mathbb{R}^{d-m} \) satisfies \( S \cap U = \{y \in U : g(y) = 0\} \) and the \((d-m) \times d \) matrix gradient \( \nabla g = (-\nabla \psi|_{I_{d-m}}) \) is surjective, indeed, the number \( d-m \) of equation requires for a local description of \( S \) is called the co-dimension. Moreover, these \( d-m \) coordinates can be flattened, i.e., \( \Phi(S \cap U) = O \times 0_{d-m} \), for a suitable homeomorphism \( \Phi \) from \( U \) into \( \mathbb{R}^d \) and an open subset \( O \) of \( \mathbb{R}^m \). Actually, as long as we work within the class \( C^k \) with \( k \geq 1 \), these three functions \( \phi, g \) and \( \Phi \) are of class \( C^k \) and they provide an equivalent definition of submanifold, via the implicit and the inverse function theorems (which are not valid for Lipschitz functions). For instance, a set \( S \) is a \( C^k \) submanifold of \( \mathbb{R}^d \) at \( x \in S \) of dimension \( m \) if there exists an open set \( V \) of \( \mathbb{R}^m \) and an injective function \( \phi : V \to \mathbb{R}^d \) such that (a) \( x \) belongs to \( \phi(V) \), (b) \( \phi \) and its inverse \( \phi^{-1} \) are of class \( C^k \), \( k \geq 1 \), and (c) the matrix \( \nabla \phi(x) \) has rank

Indeed, from such a function $\phi$ the implicit function theorem applies, and after re-ordering the variables, the equation $\phi(y') = z$ can be solved, locally, as $z = (y', \psi(y'))$ to fit Definition B.78. A function $\phi$ satisfying (a), (b), (c) is called a local chart of $S$, and a family such functions is called an atlas of $S$. Essentially, any property of an object acting on a manifold is defined in terms of the local coordinates (or charts) on an open subset $\bar{V}$ of $\mathbb{R}^m$ then the tangent space at the point $x = \phi(x') = \tilde{\phi}(x')$ is given by

$$\{\nabla \phi(x')y' \in \mathbb{R}^d : y' \in \mathbb{R}^m\} = \{\nabla \tilde{\phi}(x')\tilde{y}' \in \mathbb{R}^d : \tilde{y}' \in \mathbb{R}^m\},$$

while, for the Euclidean density $J_\psi(y')$ the invariance is expressed by the relation

$$J_{\tilde{\phi}}(\tilde{y}') = J_\phi(\phi^{-1} \circ \tilde{\phi}(\tilde{y}')) | \det (\nabla (\phi^{-1} \circ \tilde{\phi}(\tilde{y}')))|,$$

for any $\tilde{y}'$ in $\tilde{\phi}^{-1}(\phi(V) \cap \tilde{\phi}(\bar{V}))$. Actually, any nonnegative function $\rho$ defined on $S$ by local coordinates $\rho_\phi(y') = \rho(\phi(y'))$ that follows the above invariance is called a density on $S$.

In particular, if $m = d - 1$ (i.e., the hyper-area) then $\psi$ is real-valued, $\phi(y') = (y', \psi(y'))$, $y'$ in $\mathbb{R}^{d-1}$, and

$$J_\psi(y') = \sqrt{\det (\nabla \phi(y')^* \nabla \phi(y'))} = \sqrt{1 + |\nabla \psi(y')|^2},$$

where $\nabla \psi$ is the gradient of $\psi$, i.e., the $(d - 1)$-dimensional vector of all partial derivatives. This means that if $y' = (y_1, \ldots, y_{d-1})$ and $y_d = \psi(y_1, \ldots, y_{d-1})$ then the vector

$$\mathbf{n}(y', \psi(y')) = \frac{(- \partial_1 \psi(y'), \ldots, - \partial_{d-1} \psi(y'), 1)}{[1 + (\partial_1 \psi(y'))^2 + \cdots + (\partial_{d-1} \psi(y'))^2]^{1/2}},$$

represents the unit normal vector (field) to the surface $S$. This yields $d - 1$ independent tangential unit vectors $\mathbf{t}_i$, for $i = 1, \ldots, d - 1$, i.e.,

$$\mathbf{t}_1 = \frac{(1, 0, \ldots, 0, \partial_1 \phi(y'))}{[1 + (\partial_1 \phi(y'))^2]^{1/2}}, \ldots, \mathbf{t}_{d-1} = \frac{(0, 0, \ldots, 1, \partial_1 \phi(y'))}{[1 + (\partial_{d-1} \phi(y'))^2]^{1/2}},$$

which are orthogonal to $\mathbf{n}$ as expected.

**Remark** B.80. The concept of manifold applied to an open set $\Omega \subset \mathbb{R}^d$ with boundary $\partial \Omega$ could read as follows: either (a) the boundary $\partial \Omega = S \subset \mathbb{R}^d$ is a $(d - 1)$-dimensional manifold satisfying Definition B.78 and

$$\Omega \cap U = \{(y', y_d) \in \mathbb{R}^d : y_d < \psi(y'), y' \in V \subset \mathbb{R}^{d-1}\},$$
or (b) the closure $\overline{\Omega}$ is a $d$-dimensional manifold with boundary $\partial\Omega = S$, i.e., as in Definition B.78 with $\phi: U \to \mathbb{R}^d$,

$$\Omega \cap U = \{ y = (y', y_d) \in \mathbb{R}^d : \phi_d(y) < 0 \} \quad \text{and} \quad S \cap U = \{ y = (y', y_d) \in \mathbb{R}^d : \phi_d(y) = 0 \}.$$ 

In this case, the normal direction $n$ is one-sided, i.e., the “graph” cannot traverses the tangent plane. As mentioned early, both viewpoints (a) and (b) are equivalent within the class $C^k$, $k \geq 1$, but for only continuous or Lipschitz manifolds, (a) implies (b), but (b) does not necessarily implies (a). For instance, the reader is referred to Grisvard [60, Section 1.2, pp. 4–14].

Similarly, if $m = 1$ (i.e., the arc-length) then $\psi$ takes values in $\mathbb{R}^{d-1}$, $\phi(y') = (y', \psi(y'))$, $y'$ in $\mathbb{R}^1$, and

$$J_\psi(y') = \sqrt{\det (\nabla \phi(y')^* \nabla \phi(y'))} = \sqrt{1 + |d\psi(y')|^2},$$

where $d\psi(y')$ is the $(d - 1)$-dimensional vector of the first derivative of $\psi$. This means that if $y' = y_1$, $\psi = (\psi_2, \ldots, \psi_d)$ and $\psi'$ denotes the derivative, then the vector

$$\tau(y_1, \psi(y_1)) = \pm \frac{(1, \psi_2'(y_1), \ldots, \psi_d'(y_1))}{[1 + (\psi_2'(y_1))^2 + \cdots + (\psi_d'(y_1))^2]^{1/2}},$$

represents the unit tangent vector (field) to the curve $S$. This means that for $d = 3$, we have the arc-length with $m = 1$ and the area with $m = 2$, as expected.

To patch all the pieces of a submanifold we need a partition of the unity:

**Theorem B.81** (continuous PoU). Let $\{O_\alpha : \alpha\}$ be an open cover of $S \subset \mathbb{R}^d$, i.e., $O_\alpha$ are open sets and $\bigcup_\alpha O_\alpha \supset S$. Then there exists a continuous partition of the unity subordinate to $\{O_\alpha : \alpha\}$, i.e., there exists a sequence of continuous functions $\chi_i : \mathbb{R}^d \to [0, 1]$, $i = 1, 2, \ldots$, such that the support of each function $\chi_i$ is a compact set contained in some element $O_\alpha$ of the cover, and for any compact set $K$ of $\bigcup_\alpha O_\alpha$ there exists a finite number $k$ such that $\sum_{i=1}^k \chi_i = 1$ on $K$.

**Proof.** First, if $I$ and $J$ are $d$-intervals (or $d$-rectangles) in $\mathbb{R}^d$ such that $I$ is compact, $J$ is open with compact closure and $I \subset J$ then there exists a continuous function $\varpi : \mathbb{R}^d \to [0, 1]$ satisfying $\varpi(x) = 1$ for every $x$ in $I$ and $\varpi(x) = 0$ for every $x$ outside of $J$, actually, an explicit construction of the $\varpi$ is clearly available.

Second, consider a sequence of compact set $K_n$ such that $\bigcup_n K_n = \bigcup_\alpha O_\alpha$ to deduce that for every $x$ in $K_n$ must belong to some open set $O_\alpha$, and so, there are $d$-intervals $I$ compact and $J$ open with compact closure such that $x$ belongs to the interior of $I$ and $I \subset J \subset \overline{J} \subset O_\alpha$. By compactness, there exists a finite number of $I_i \subset J_i$ with the above property which form a finite cover of $K_n$, i.e., $\bigcup_{i=1}^k \tilde{I}_i \supset K_n$. Hence, there is a sequence of $d$-intervals $I_i \subset J_i$, $I_i$
is compact, $J_i$ is open with a compact closure contained in some $\Omega_\alpha$, and such that $\bigcup_{i=1}^k \tilde{I}_i = \bigcup_{\alpha} O_\alpha$.

Next, denote by $\varpi_i$ a continuous function such that $\varpi = 1$ on $I_i$ and $\varpi = 0$ outside $J_i$ to define $\chi_1 = \varpi_1$ and

$$
\chi_i = (1 - \varpi_1)(1 - \varpi_2) \cdots (1 - \varpi_{i-1})\varpi_i,
$$

for $i \geq 2$. The support of $\chi_i$ is certainly contained in some $\Omega_\alpha$ and since

$$
\sum_{i=1}^k \chi_i = 1 - \prod_{i=1}^k (1 - \varpi_i), \quad \forall k \geq 1,
$$

we deduce that $\sum_{i=1}^\infty \chi_i = 1$ on any compact $K$ of $\bigcup_{\alpha} O_\alpha$, where the series is locally finite, i.e., only a finite number of $\chi_i$ have support in $K$. \qed

It is not hard to modify the argument so that the functions $\chi_i$ are of class $C^k$, but to actually see that $\chi_i$ may be chosen of class $C^\infty$, we make use of the fact that

$$
g(x) = \begin{cases} 
0 & \text{if } x \leq 0 \\
 e^{-1/x} & \text{if } x > 0 
\end{cases}
$$

is a function of class $C^\infty$.

Therefore, apply Theorem B.81 to the open cover $\{U\}$ of the submanifold $S$ as in Definition B.78 to find a continuous partition of the unity $\{\chi_i\}$ with a compact support contained in the open set $U_i \subset \mathbb{R}^d$, and charts $\psi_i$ defined on an open set $V_i \subset \mathbb{R}^m$ such that

$$
S \cap U_i = \{\phi_i(y') = (y', \psi_i(y')) \in \mathbb{R}^d : y' \in \mathbb{R}^m\}.
$$

Now, the Lebesgue measure defined on $\mathbb{R}^m$ can be transported to $S$. Indeed, a nonnegative function $f$ defined on a submanifold $S$ in $\mathbb{R}^d$ of dimension $m$ is called integrable with respect to the surface Lebesgue measure $\sigma(dx)$ if the function

$$
y' \mapsto \sum_i \chi_i(\phi_i(y')) f(\phi_i(y')) \sqrt{\det(\nabla \phi_i(y')^* \nabla \phi_i(y'))}
$$

is Lebesgue integrable on $\mathbb{R}^m$ and

$$
\int_S f(x) \sigma(dx) = \sum_i \int_{\mathbb{R}^m} \chi_i(\phi_i(y')) f(\phi_i(y')) \sqrt{\det(\nabla \phi_i(y')^* \nabla \phi_i(y'))} \, dy'
$$

is the definition of the integral. These definition are independent of the particular partition of the unity and the charts chosen. Indeed, if $\phi$ and $\tilde{\phi}$ have a common image $S_0$ then the formula for the change-of-variables $y' \mapsto \tilde{\phi}^{-1}(\phi(y'))$
in the \(m\)-dimensional integral yields
\[
\int_{\phi^{-1}(S_0)} f(\phi(y')) \sqrt{\det \left( \nabla \phi_i(y')^* \nabla \phi_i(y') \right)} \, dy' = \\
= \int_{\phi^{-1}(S_0)} f(\tilde{\phi}_i(y')) \sqrt{\det \left( \nabla \tilde{\phi}_i(y')^* \nabla \tilde{\phi}_i(y') \right)} \, dy',
\]
as expected. In particular, any linear (affine) submanifold \(S\) in \(\mathbb{R}^d\) of dimension \(m\) can be represented as
\[
S = \{(y', \psi(y')) \in \mathbb{R}^d : y' \in \mathbb{R}^m\}, \quad \text{with } \psi(y') = a + y'A,
\]
where \(A\) is a \(m \times (d-m)\) matrix \(A\) of maximal rank and \(a\) is a row vector in \(\mathbb{R}^{d-m}\). Hence, \(\phi(y') = (y', a + y'A)\), \(\nabla \phi(y') = (I_m, A)^*\), and \(\det((\nabla \phi(y'))^* \nabla \phi(y')) = \det(I_m + AA^*)\), independent of \(y'\), and it represents the \(m\)-volume of the image \(\{y'A \in \mathbb{R}^d : y' \in Q' \subset \mathbb{R}^m\}\), where \(Q'\) is the unit cube in \(\mathbb{R}^m\). In general
\[
\sigma(\phi(Q)) = \int_Q \sqrt{\det \left( \nabla \phi(y')^* \nabla \phi(y') \right)} \, dy',
\]
for any cube \(Q \subset \mathbb{R}^m\) inside the open set \(D_\phi\) where the local chart \(\phi\) is defined. Actually, the above equality holds true for any Lebesgue measurable set \(A = Q \subset D_\phi\) in \(\mathbb{R}^m\) and \(\sigma\) becomes a Borel measure on \(S \subset \mathbb{R}^d\), and except for a multiplicative constant, this surface Lebesgue measure agrees with the \(m\)-dimensional Hausdorff measure as discussed in next section.

For the case of the hyper-area \((m = d-1)\), if for instance, the local charts are taken
\[
\phi(y_1, \ldots, y_{d-1}) = (y_1, \ldots, y_{d-1}, \psi(y_1, \ldots, y_{d-1}))
\]
then the surface Lebesgue measure has locally the form
\[
\int_{\mathbb{R}^{d-1}} f(y', \psi(y')) \sqrt{1 + |\nabla \psi(y')|^2} \, dy', \quad y' = (y_1, \ldots, y_{d-1}).
\]
If the submanifold \(S\) is only (locally) Lipschitz then the Euclidean density is defined almost everywhere in \(\mathbb{R}^m\), and the surface Lebesgue measure \(\sigma\) still makes sense as a Borel measure on \(S\).

In particular, if \(\Omega\) is an open subset of \(\mathbb{R}^d\) with a Lipschitz boundary \(\partial \Omega\) (see Remark B.80) then the surface Lebesgue measure \(d\sigma\) is can be used to define the space \(L^1(\partial \Omega)\), i.e., the space of functions \(f : \Omega \to \mathbb{R}\) such that the composition function
\[
y' \mapsto \sum_i \chi_i(\phi_i(y')) f(\phi_i(y')) \sqrt{\det \left( \nabla \phi_i(y')^* \nabla \phi_i(y') \right)}
\]
is integrable in \(\mathbb{R}^{d-1}\), for some local coordinates \(\psi_i : V_i \to \mathbb{R}^d, \phi_i(y') = (y' \psi(y'))\), and a subordinate partition of the unity \(\{\chi_i\}\). As mentioned early, all properties of function defined on the boundary \(\partial \Omega\) are studied by local coordinates.
Moreover, if $\Omega$ a bounded domain as above and $F$ is a continuously differentiable functions defined on the closure $\overline{\Omega}$ with values in $\mathbb{R}^d$ then the divergent theorem, i.e.,

$$\int_{\Omega} \nabla F \, d\ell = \int_{\partial \Omega} F \cdot n \, d\sigma$$

holds true, where $n$ is the outward unit normal vector defined almost everywhere with respect to the surface Lebesgue measure $\sigma$. Similarly, with the integration-by-parts or Green formula.

Recall that by definition complex-valued measures are finite measure, i.e., a complex measure $\mu$ has a real-part $\Re(\mu)$ and an imaginary-part $\Im(\mu)$ both of which are finite real-valued measures on a measurable space $(\Omega, \mathcal{F})$. Thus, following the complex numbers arithmetic, a real- or complex-valued measurable function $f$ is integrable with respect to a complex valued measure $\mu$ if and only if the real-valued function $|f|$ is integrable with respect to the real-valued measures $\Re(\mu)$ and $\Im(\mu)$.

In general, every integral with complex values is reduced to its real and imaginary parts, and then each one is studied separately and put back together when the result make sense, i.e., both parts are finite and $\mathbb{H}^3$ complex plane is identified with $\mathbb{R}^2$ for all practical use. Hence, of particular interest is the integral over a complex Lipschitz curve, which treated as a generalization of the Riemann-Stieltjes contour integral over a complex $C^1$-curve, i.e., the complex line integral

$$\int_C f(z) \, dz = \int_a^b f(x(t) + iy(t)) (a'(t) + iy'(t)) \, dt,$$

where the curve $C$ is parameterized as $z = x(t) + iy(t)$, with $t$ from $a$ to $b$ and Lipschitz functions $t \mapsto x(t)$ and $t \mapsto y(t)$.

For instance, the reader may check the textbook Amann and Escher [7, Sections VII.9 and VII.10, pp. 242–280 and Chapter XI–X, pp. 235–456] or Duistermaat and Kolk [38, Chapter 7, pp. 487–535] or Giaquinta and G. Modica [54, Chapter 4, pp. 213–282] or Haroske and Triebel [66, Appendix A, pp. 245–249]. Regarding manifolds, depending on the reader interest, the following textbooks, Auslander and MacKenzie [13], Berger and Gostiaux [17], Boothby [20], Gadea and Muñoz Masqué [48], and Tu [132] could be consulted.

### B.6.3 Smooth Approximations

By definition, for any integrable function $f$ there exists a sequence $\{f_k : k \geq 1\}$ of simple functions such that $\|f_k - f\|_1 \to 0$, which implies that the vector space of (integrable) simple functions is dense in $L^1$, and in particular we deduce that $L^1_0 \cap L^\infty$ is dense in $L^1$. Also we have
Proposition B.82. Given a Lebesgue integrable function $f$ and a real number $\varepsilon > 0$ there exists a continuous functions $g$ such that

$$\int_{\mathbb{R}^d} |f(x) - g(x)| \, dx = \|f - g\|_1 < \varepsilon,$$

and $g$ vanishes outside of some ball, i.e., the space of continuous functions with compact supports $C_0^0$ is dense in $L^1$.

Proof. Each real-valued measurable function $f$ can be written as $f = f^+ - f^-$, where $f^+$ and $f^-$ are nonnegative $m$-measurable functions. By Proposition B.9, for any nonnegative measurable function $f^\pm$ there exists an increasing sequence $\{f^\pm_k : k \geq 1\}$ of simple functions such that $f^\pm_k(x) \to f^\pm(x)$, for almost everywhere $x$ in $\mathbb{R}^d$, as $k \to \infty$. Hence, by the monotone convergence, we obtain

$$\lim_k \int_{\mathbb{R}^d} |f^\pm_k(x) - f^\pm(x)| \, dx = 0,$$

whenever $f$ is integrable in $\mathbb{R}^d$. Now, for a fixed $k$, the simple function $f^\pm_k$ is a finite combination of expression of the form $c \mathbb{1}_E$, with $E$ a (Borel) measurable set of finite measure and $c$ a real number. For each $E$ and $\varepsilon > 0$ there exists an open set $U \supset E$ such that $m(U \setminus E) < \varepsilon$. Since $U$ is an open set in $\mathbb{R}^d$, there exists an non-overlapping sequence $\{Q_i : i \geq 1\}$ of closed cubes such that $U = \bigcup_i Q_i$, which yields $m(U) = \sum_i m(Q_i)$, and

$$\lim_n \int_{\mathbb{R}^d} |\mathbb{1}_U(x) - \mathbb{1}_{F_n}(x)| \, dx = 0, \quad \text{with } F_n = \bigcup_{i=1}^n Q_i.$$

Given $\varepsilon > 0$ and the cubes $F_n$, we can easily find a continuous function $g_{\varepsilon,n}$ such that

$$\int_{\mathbb{R}^d} |g_{\varepsilon,n}(x) - \mathbb{1}_{F_n}(x)| \, dx < \varepsilon.$$

Combining all, the desired approximation follows.

Alternatively, given an integrable function $f$, the dominated convergence implies that, for every $\varepsilon > 0$ there exists $r > 0$ such that the function $f_r(x) = \mathbb{1}_{\{|x| \leq r\}} \mathbb{1}_{\{|f(x)| \leq r\}} f(x)$ satisfies

$$\int_{\mathbb{R}^d} |f(x) - f_r(x)| \, dx \leq \frac{\varepsilon}{2}.$$

Now, for this $r > 0$, we apply Lusin’s Theorem (e.g., see Part I) to obtain a closed set $C_r \subset B_r = \{x : |x| \leq r\}$ such that $f_r$ is continuous on $C_r$ and $m(B_r \setminus C_r) < \varepsilon/(5r)$. Next, essentially based on Tietze’s extension $f_r$ can be extended to a continuous function $g_r$ on $\mathbb{R}^d$ satisfying the conditions: (a) $|g_r| \leq r$ on $\mathbb{R}^d$ and (b) the set $N_r = \{x \in \mathbb{R}^d : g_r(x) \neq f_r(x)\}$ is contained in some ball
Appendix B. Measure and Integration

\[ B \text{ and } m(N_r) < \varepsilon/(4r). \text{ Hence} \]

\[
\int_{\mathbb{R}^d} |f(x) - g_r(x)| \, dx \leq \int_{\mathbb{R}^d} |f(x) - f_r(x)| \, dx + \]
\[
+ \int_B |f_r(x) - g_r(x)| \, dx \leq \frac{\varepsilon}{2} + 2r m(N_r) \leq \varepsilon,
\]
and \( g = g_r \) is the desired function. \( \square \)

The arguments used in proving Proposition B.82 can be extended to a more general setting, e.g., replacing the Lebesgue measure \( m \) on \( \mathbb{R}^d \) by a Borel measure \( \mu \) on a metric space \( \Omega \). There are other arguments for approximation typical in \( \mathbb{R}^d \), for instance, **mollification** and **truncation**.

Let us begin with the following results.

**Proposition B.83.** If \( f \) belong to \( L^1 \) then

\[
\lim_{a \to 0} \int_{\mathbb{R}^d} |f(x + a) - f(x)| \, dx = 0,
\]

i.e., the translation operator \( \tau_a f = f(\cdot - a) \) is continuous in \( L^1 \).

**Proof.** Indeed, let us denote by \( K \) the collection of all functions \( f \) in \( L^1 \) such that \( \|f(\cdot + a) - f\|_1 \to 0 \) as \( a \to 0 \). It is simple to check that \( K \) is a closed vector space, i.e.,

(a) if \( \alpha, \beta \in \mathbb{R} \) and \( f, g \in K \) then \( \alpha f + \beta g \in K \),

(b) if \( \{f_n : n \geq 1\} \subset K \) and \( \|f_n - f\|_1 \to 0 \) then \( f \in K \).

Now, we use the same argument of Proposition B.82 to successively approximate an integrable function by simple functions, next \( c \mathbb{1}_A \) with \( A \) measurable and \( m(A) < \infty \) by \( c \mathbb{1}_U \) with a bounded open set \( U \), and then for every \( \varepsilon > 0 \) and \( U \) we find a finite union of non-overlapping cubes \( Q = \bigcup_{i=1}^n Q_i \) with \( Q \subset U \) and \( m(U \setminus Q) < \varepsilon \), to establish that the family of (simple) functions of the form \( \sum_{i=1}^n a_i \mathbb{1}_{Q_i} \), where the cubes \( Q_i \) have edges parallel to the axis, can approximate and integrable function in the \( \| \cdot \|_1 \) norm.

Since the characteristic function of a \( d \)-interval (or a cube) belongs to \( K \), we deduce \( K = L^1 \).

Alternatively, we may claim that any integrable function can be approximated in the \( \| \cdot \|_1 \) norm by continuous functions with compact support (Proposition B.82), which also belong to \( K \). \( \square \)

For two integrable functions \( f \) and \( g \) we consider the **convolution** \( f \ast g \) given by the formula

\[
(f \ast g)(x) = \int_{\mathbb{R}^d} f(x - y) \, g(y) \, dy = \int_{\mathbb{R}^d} f(y) \, g(x - y) \, dy, \quad \forall x \in \mathbb{R}^d.
\]

(B.28)

It is clear that if either \( f \) or \( g \) is essentially bounded then \( x \mapsto (f \ast g)(x) \) is well defined and \( \|f \ast g\|_{\infty} \leq \|f\|_1 \|g\|_{\infty} \). Moreover, we can also check the inequality \( \|f \ast g\|_1 \leq \|f\|_1 \|g\|_1 \), which means that the convolution \( f \ast g \) is defined almost everywhere, i.e., \( L^1 \) is a commutative algebra with the convolution product.

**Definition B.84** (locally integrable). A function \( f : \mathbb{R}^d \to \mathbb{R} \) is called **locally integrable** if for every \( x \) in \( \mathbb{R}^d \) there exists an open neighborhood \( U_x \) such that \( f \) is integrable in \( U_x \), or equivalently, the restriction to any compact set in integrable. This class of functions is denoted by \( L^1_{\text{loc}} \), and we say that a sequence of locally integrable functions \( \{f_n : n \geq 1\} \) converges to \( f \) **locally in** \( L^1 \) or in \( L^1_{\text{loc}} \), if

\[
\lim_{n} \int_K |f_n(x) - f(x)| \, dx = 0, 
\]

for every compact set \( K \) of \( \mathbb{R}^d \). Similarly, \( L^\infty_{\text{loc}} \) is the space of **locally essentially bounded** functions, i.e., functions bounded almost everywhere on any compact set. We also have the spaces of equivalence classes \( L^1_{\text{loc}} \) and \( L^\infty_{\text{loc}} \).

Certainly, we mean \( f \) is Lebesgue measurable and \( f 1_K \) is in \( L^1 \). This concept of locally integrable can be used on locally compact spaces \( \Omega \) with a Borel measure \((\mu, \mathcal{B})\).

Also, recall that we say that a measurable function defined almost everywhere has compact support if it is equal to zero almost everywhere outside of a ball. The sub-vector space of \( L^1 \) (or \( L^1 \)) of all functions with compact support is denoted by \( L^1_0 \) (or \( L^1_0 \)), and similarly with \( L^\infty_0 \) (or \( L^\infty_0 \)). The convolution \( f \ast g \) is also defined if \( f \) and \( g \) are only locally integrable and one of then has compact support, i.e., \( f \in L^1_0 \) and \( g \in L^1_{\text{loc}} \) or \( g \in L^\infty_{\text{loc}} \) implies \( f \ast g \in L^1_{\text{loc}} \) or \( f \ast g \in L^\infty_{\text{loc}} \), respectively.

In general, given two Lebesgue measurable functions \( f \) and \( g \), we say that the convolution \( f \ast g \) is defined if the functions inside the integrals in the expression (B.28) are integrable for almost every \( x \). Remark that the convolution operation commutes with the translation operator, i.e., \( \tau_a(f \ast g) = (\tau_a f) \ast g = f \ast (\tau_a g) \).

**Proposition B.85.** Let \( f \) and \( g \) be two Lebesgue measurable functions in \( \mathbb{R}^d \).

(a) If \( f \) is integrable and \( g \) is essentially bounded then the convolution \( f \ast g \) is a bounded uniformly continuous function. Moreover, \( f \) is only locally integrable, \( g \) is only locally essentially bounded, and either \( f \) or \( g \) has a compact support then the convolution \( f \ast g \) is a continuous function.

(b) Denote by \( \partial_i f \) the partial derivative of \( f \) with respect to \( x_i \). If \( f \) is essentially bounded or integrable, \( g \) is integrable and the partial derivative \( \partial_i f \) is a bounded function then the \( i \)-partial derivative of the convolution \( f \ast g \) is a bounded uniformly continuous function and \( \partial_i(f \ast g) = (\partial_i f) \ast g \). Moreover, if \( f \) and \( g \) are only locally integrable, either \( f \) or \( g \) has a compact support, and the partial derivative \( \partial_i f \) is a bounded function then the \( i \)-partial derivative of the convolution \( f \ast g \) is continuous function and \( \partial_i(f \ast g) = (\partial_i f) \ast g \).
Proof. Consider the bound

$$\left| (f \ast g)(x + a) - (f \ast g)(x) \right| \leq \int_{\mathbb{R}^d} |f(x + a - y) - f(x - y)| |g(y)| \, dy,$$

where the integral is actually limited to the support of $g$. Thus, if $g$ is essentially bounded then Proposition B.83 proves most of the above claim (a). For the local version of this claim, we remark that if $f$ or $g$ has a compact support then the integral is only on a compact set $K$ (as long as $x$ remain in another compact region) instead of $\mathbb{R}^d$, and again, the continuity follows.

Next, by means of the Mean Value Theorem and the dominate convergence, we obtain $\partial_i (f \ast g) = (\partial_i f) \ast g$ and in view of (a), we deduce the claim (b).

Certainly, we can iterate the property (b) to deduce that $\partial^{\alpha} (f \ast g) = (\partial^{\alpha} f) \ast g$, for any multi-index $\alpha$ with $|\alpha| \leq n$, e.g., $f$ belongs to $C^n_b$ and $g$ is in $L^1$.

Regarding the claim (b), we assume that the partial derivative $\partial_i f$ exists a any point, so that the Mean Value Theorem can be applied, however, the expression $(\partial_i f) \ast g$ is a continuous function even if $\partial_i f$ is defined almost everywhere. Nevertheless, if we assume that $\partial_i f$ is defined only almost everywhere then we may have a non-constant function with $\partial_i f = 0$ a.e. (like the Cantor function).

To end this section let us state (without proof)

**Theorem B.86.** Let $f$ be a Lebesgue locally integrable function in $\mathbb{R}^d$. Then almost every point is a Lebesgue point for $f$, i.e., there exist a negligible $N = N_f$, $|N| = 0$, such that

$$\lim_{r \to 0} \frac{1}{|B(x, r)|} \int_{B(x, r)} |f(y) - f(x)| \, dy = 0, \quad \forall x \in \mathbb{R}^d \setminus N,$$

where $B(x, r)$ is the ball centered at $x$ with radius $r$. \hfill \Box

### B.6.4 Partition of the Unity

First recall that the function

$$k(x) = \begin{cases} c_r \exp \left[- \left( r^2 - |x|^2 \right)^{-1} \right] & \text{if } |x| < r, \\ 0 & \text{otherwise,} \end{cases} \quad (B.29)$$

for any given constant $r > 0$ and a suitable $c_r$ to meet the condition $\|k\|_1 = 1$, is an example of a smooth kernel (in $\mathbb{R}^d$) with compact support.

- **Remark B.87.** If $k$ is an integrable kernel, i.e., an integrable function such that

$$\int_{\mathbb{R}^d} k(x) \, dx = 1,$$
and \( \{ k_\varepsilon : \varepsilon > 0 \} \) its corresponding mollifiers, i.e., \( k_\varepsilon(x) = \varepsilon^{-d}k(x/\varepsilon) \), for every \( x \in \mathbb{R}^d \), then we have

\[
\lim_{\varepsilon \to 0} \| f \ast k_\varepsilon - f \|_1 = 0, \quad \forall f \in \mathcal{L}^1.
\]

Moreover, if either \( f \) is essentially bounded in \( \mathbb{R}^d \) or the kernel \( k \) satisfies

\[
k(x) = \alpha(x)|x|^{-d}, \quad \text{a.e. } x \in \mathbb{R}^d, \quad \text{with } \alpha(x) \to 0 \text{ as } |x| \to \infty,
\]

and \( f \) is uniformly continuous and bounded in a subset \( F \) of \( \mathbb{R}^d \), i.e., for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( |x-y| < \delta \) and \( x \in F \) imply \( |f(x) - f(y)| < \varepsilon \), and \( \sup_{x \in F} |f(x)| < \infty \), then \( (f \ast k_\varepsilon)(x) \to f(x) \) uniformly for \( x \in F \), as \( \varepsilon \to 0 \). This is usually referred to as approximation by convolution smooth functions. \( \square \)

Now, let \( K \) be a compact set in \( \mathbb{R}^d \) and \( U \) be an open set satisfying \( U \supset K \), with compact closure \( \overline{U} \). If \( B^1 = \{ x \in \mathbb{R}^d : |x| \leq 1 \} \) then for any \( \delta > 0 \) sufficiently small we have \( K \subset R = K + \delta B^1 \subset U \), and so we can select \( \varepsilon > 0 \) such that \( K + \varepsilon B^1 \subset R \) and \( R + \varepsilon B^1 \subset U \). Therefore, we may consider the convolution \( 1_R \ast k_\varepsilon \), where \( k \) is given by \( B.29 \) with \( r = 1 \). Since the support of \( k_\varepsilon \) satisfies \( \text{supp}(k_\varepsilon) \subset B^1 \) we deduce that there exists a function \( f = 1_R \ast k_\varepsilon \) with derivatives of any order such that \( f = 1 \) on \( K \) and \( f = 0 \) outside \( U \).

Another point is to use Remark \( B.87 \) with \( k \) given by \( (B.29) \) to show that for any \( f \) in \( \mathcal{L}^1 \) and \( \varepsilon > 0 \) there exists a function \( g \) with derivatives of any order such that \( \| f - g \|_1 < \varepsilon \), i.e., the space \( C_0^\infty(\mathbb{R}^d) = \bigcap_{n=1}^\infty C_0^n(\mathbb{R}^d) \) is dense in \( \mathcal{L}^1 \).

**Theorem B.88.** Let \( \{ \Omega_\alpha : \alpha \} \) be an open cover of an open subset \( \Omega \) of \( \mathbb{R}^d \), i.e., \( \Omega_\alpha \) is open and \( \Omega = \bigcup_\alpha \Omega_\alpha \). Then there exists a smooth partition of the unity \( \{ \chi_i : i \geq 1 \} \) subordinate to \( \{ \Omega_\alpha : \alpha \} \), i.e., (a) \( \chi_i \) belongs to \( C_0^\infty(\mathbb{R}^d) \), (b) for every \( i \) there exists \( \alpha = \alpha(i) \) such that \( \chi_i(x) = 0 \) for every \( x \in \Omega \setminus \Omega_\alpha \), namely, \( \text{supp}(\chi_i) \subset \Omega_\alpha \), (c) \( 0 \leq \chi_i(x) \leq 1 \) and \( \sum_i \chi_i(x) = 1 \), for every \( x \in \Omega \), where the series is locally finite, namely, for any compact set \( K \) of \( \Omega \) the set of indices \( i \) such that the support of \( \chi_i \) intersect \( K \), \( \text{supp}(\chi_i) \cap K \neq \emptyset \), is finite.

**Proof.** (1) First we show that there exists a locally finite subordinate open cover \( \{ U_i : i \geq 1 \} \) with compact closure \( \overline{U}_i \), i.e., for any compact set \( K \subset \Omega \) the set of indices \( \{ i \geq 1 : U_i \cap K \neq \emptyset \} \) is finite, and for every \( i \geq 1 \) there exists \( \alpha(i) \) such that \( \overline{U}_i \subset \Omega_{\alpha(i)} \).

Indeed, consider the compact sets

\[
K_n = \{ x \in \Omega : |x| \leq n \ \text{and} \ d(x, \mathbb{R}^d \setminus \Omega) \geq 1/n \},
\]

for \( n \geq 1 \), where \( d(x, A) = \inf \{|x - y| : y \in A\} \). We have \( \Omega = \bigcup_n K_n \), \( K_{n-1} \subset K_n^o \), where \( K_n^o \) is the interior of \( K_n \). For \( n \geq 3 \) define \( \Omega_{\alpha,n} = \Omega_{\alpha} \cap K_{n+1}^o \cap (\Omega \setminus K_{n-2}) \), and remark that \( \{ \Omega_{\alpha,n} : \alpha \} \) is a open cover of \( K_{n+1}^o \cap (\Omega \setminus K_{n-2}) \). On the other hand, for each \( x \in K_n \cap (\Omega \setminus K_{n-1}) \) there exists an open set \( U_n(x) \) with closure \( \overline{U}_n(x) \) included in \( \Omega_{\alpha,n} \) for some \( \alpha \). Hence, the family \( \{ U_n(x) : x \} \) forms an open cover of the compact set \( K_n \cap (\Omega \setminus K_{n-1}) \).
and so, there exists a finite subcover, i.e., \( x_1, \ldots, x_m, m = m(n) \) such that \( \{U_n(x_j) : j = 1, \ldots, m(n)\} \) cover \( K_n \cap (\Omega \setminus \mathcal{K}_n^{n-1}) \), for every \( n \geq 3 \). Thus, the family \( \{U_n(x_j) : j = 1, \ldots, m(n), n \geq 3\} \), now denoted by \( \{U_i : i \geq 1\} \), is countable and satisfies the required conditions.

(2) Next, we construct a continuous partition of the unity \( \{f_i : i \geq 1\} \) subordinate to \( \{U_i : i \geq 1\} \), and so, also subordinate to \( \{\Omega_\alpha : \alpha\} \). Indeed, we apply again the above argument (1), with \( \{U_i : i \geq 1\} \) instead of \( \{\Omega_\alpha : \alpha\} \), to obtain another locally finite subordinate cover \( \{V_i : i \geq 1\} \), which (after relabeling and deleting some \( U \)-open if necessarily) satisfies \( V_i \subset U_i \subset \Omega_\alpha, \ \alpha = \alpha(i) \), for every \( i \geq 1 \). Now, we use Urysohn’s Lemma to get a continuous function \( g_i \) satisfying \( g_i(x) = 1 \) for every \( x \in V_i \) and \( g_i(x) = 0 \) for any \( x \) in \( \mathbb{R}^d \setminus U_i \), i.e., \( \text{supp}(g_i) \subset \overline{U_i} \). Since the covers are locally finite, for any compact set \( K \) of \( \Omega \) there exists only finite many \( i \) such that \( U_i \cap K \neq \emptyset \) and so the finite sum \( g(x) = \sum_i g_i(x) \) defines a continuous function satisfying \( g(x) \geq 1 \), for every \( x \) in \( \Omega \). Hence, the family of continuous functions \( \{f_i : i \geq 1\} \), with \( f_i(x) = g_i(x)/g(x) \), is a partition of the unity subordinate to \( \{U_i : i \geq 1\} \), satisfying all the required conditions, except for the smoothness.

(3) To obtain a smooth partition we use the convolution with a smooth kernel \( k \) having compact support defined by (B.29) for \( r = 1 \), as in Remark B.87 with \( k_\varepsilon \). Indeed, again we apply (1) to get another locally finite subordinate cover \( \{W_i : i \geq 1\} \) which satisfies \( W_i \subset V_i \subset \overline{V_i} \subset U_i \subset \Omega_{\alpha(i)} \), \( \alpha = \alpha(i) \), for every \( i \geq 1 \). If \( 2\varepsilon_i = \min\{d(V_i, \Omega \setminus U_i), d(W_i, \Omega \setminus V_i)\} \) then the convolution \( \varphi_i = \mathbb{1}_{V_i} \ast k_{\varepsilon_i} \) is an infinitely differentiable (smooth) function and, since \( \text{supp}(k_{\varepsilon_i}) \) is included in the ball centered at the origin with radius \( \varepsilon_i \), we have

\[
0 \leq \varphi_i \leq 1 \text{ in } \mathbb{R}^d, \quad \text{supp}(\varphi_i) \subset \overline{U_i}, \quad \text{and } \varphi_i = 1 \text{ on } \overline{W_i}.
\]

Moreover, the finite sum \( \varphi(x) = \sum_i \varphi_i(x) \) defines a smooth function satisfying \( \varphi(x) \geq 1 \), for every \( x \) in \( \Omega \). Hence, the family of smooth functions \( \{\chi_i : i \geq 1\} \), with \( \chi_i(x) = \varphi_i(x)/\varphi(x) \), is a partition of the unity subordinate to \( \{U_i : i \geq 1\} \), satisfying all the required conditions. 

We note that in the above proof, we may go directly to (3) without using (2). However, (1) and (2) are valid for \( \sigma \)-compact locally compact Hausdorff topological spaces. Also, we may deduce (3) from (2) by using \( \varphi_i = g_i \ast k_{\varepsilon_i} \) with \( k \) as in (B.29) for \( r = 1 \) and \( 2\varepsilon_i = d(\overline{U_i}, \Omega \setminus \Omega_{\alpha(i)}) \). Indeed, we remark that \( g_i(x) > 0 \) implies \( \varphi_i(x) > 0 \) and then \( \varphi(x) = \sum_i \varphi_i(x) > 0 \), for every \( x \) in \( \Omega \). Alternatively, we may check that the functions \( g_i \) in (2) can be chosen infinitely differentiable, instead of just continuous. For instance, the reader may consult Folland [44, Section 4.5, pp. 132–136] and Malliavan [86, Section II.1, pp. 55–61].

### B.6.5 Representation Theorems

When discussing signed measures, it was clear that a signed measure cannot assume the values \( +\infty \) and \( -\infty \). However, a \( \sigma \)-finite signed measure \( \mu \) make sense, i.e., the measurable space \( (\Omega, \mathcal{F}) \) has a partition \( \Omega = \sum_k \Omega_k \) such that
the restriction of $\mu$ to $\Omega_k$, denoted by $\mu_k$, is a finite signed measure. This is essentially the situation of a linear functional on the space $L^1(\Omega, \mathcal{F}, \mu)$.

There are the various versions of the so-called Riesz representation theorems. For instance, recall the definition of the Lebesgue spaces $L^p = L^p(\Omega, \mathcal{F}, \mu)$, for $1 \leq p < \infty$ and its dual, denoted by $(L^p)'$, the Banach space of linear continuous (or bounded) functional on $L^p$, endowed with the dual norm

$$\|g\|_p = \sup \{ \langle g, \varphi \rangle : \|\varphi\|_p \leq 1 \}, \quad \forall g \in (L^p)'$$

where $\langle \cdot, \cdot \rangle$ denote the duality pairing, i.e., $g$ acting on (or applied to) $\varphi$, and for the supremum, the functions $\varphi$ can be taken in $L^p$ or just a simple function, actually, $\varphi$ belonging to some dense subspace of $L^p$ is sufficient.

**Theorem B.89.** For every $\sigma$-finite measure space $(\Omega, \mathcal{F}, \mu)$ and $p$ in $[1, \infty)$, the map $g \mapsto Tg$, defined by

$$\langle Tg, f \rangle = \int_{\Omega} gf \, d\mu,$$

gives a linear isometry from $(L^q, \| \cdot \|_q)$ onto the dual space of $(L^p, \| \cdot \|_p)$, with $1/p + 1/q = 1$.

**Proof.** First, Hölder inequality shows that $T$ maps $L^q$ into $(L^p)'$ with $\|Tg\|_p \leq \|g\|_q$. Moreover, by means of Proposition B.60 and Remark B.61 we have the equality, i.e., $\|Tg\|_p = \|g\|_q$, which proves that $T$ is an isometry.

To check that $T$ is onto, for any given element $g$ in the dual space $(L^p)'$ define

$$\nu_g(A) = \langle g, 1_A \rangle, \quad \forall A \in \mathcal{F}, \mu(A) < \infty.$$ 

Considering $\nu_g$ defined on measurable subsets $A \subset F$, for a fixed $F$ in $\mathcal{F}$ with $\mu(F) < \infty$, we have a signed measure $\nu_g$ on $F \subset \Omega$, which is absolutely continuous with respect to $\mu$. Thus Radon-Nikodym Theorem B.72 yields an almost everywhere measurable function, still denoted by $g_F$, such that

$$\nu_g(A) = \int_A g_F \, d\mu, \quad \forall A \in \mathcal{F}, A \subset F.$$ 

By linearity, we have

$$\langle g, 1_F \varphi \rangle = \int_F g \, \varphi \, d\mu,$$

for any simple functions $\varphi$. Again, Proposition B.60 and Remark B.61 imply this implies the equality $\|g 1_F\|_p = \|g_F\|_q$, where $g 1_F$ is the restriction of the functional $g$ to $F$, i.e., $\langle g 1_F, f \rangle = \langle g, 1_F f \rangle$.

Since for some sequence $\{f_n\}$ of functions in $L^p$ we have $\langle g, f_n \rangle \to \|g\|_p$, there exists a $\sigma$-finite measurable set $G$ (supporting all $f_n$) such that

$$\|g\|_p = \sup \{ \langle g, 1_G \varphi \rangle : \|\varphi\|_p \leq 1 \}, \quad (B.30)$$
and $G = \bigcup_n G_n$, for some monotone sequence $\{G_n\}$ of measurable sets with $\mu(G_n) < \infty$. Since $G_n \subset G_{n+1}$ implies $g_{\mathcal{G}_n} = g_{\mathcal{G}_{n+1}}$ on $G_n \setminus N_n$, with $\mu(N_n) = 0$, we can define a measurable function $g_{\mathcal{G}}$ such that $g_{\mathcal{G}} = 0$ outside of $G$ and $g_{\mathcal{G}} = g_{\mathcal{G}_n}$ on $G_n \setminus N_n$, for every $n \geq 1$, i.e., we have

$$\langle g, 1_G \varphi \rangle = \int g_{\mathcal{G}} \varphi \, d\mu, \quad \text{for any simple function } \varphi$$

Now, apply Proposition B.60 and Remark B.61 to deduce that $\|g 1_G\|_p = \|g_{\mathcal{G}}\|_q$. On the other hand, for any $\mu(F) < \infty$ with $F \cap G = \emptyset$ we must have $\nu_g(F) = 0$, i.e., $g_{\mathcal{G}} = 0$ almost everywhere. Indeed, if $\nu_g(F) > 0$ then

$$\langle g, 1_F + 1_G \varphi \rangle = \langle g, 1_F \rangle + \langle g, 1_G \varphi \rangle$$

yields $\|g\|_p = \langle g, 1_F \rangle + \|g 1_G\|_p$, which contradicts the equality (B.30). This proves that $g = g 1_G$ and $g = T(g_{\mathcal{G}})$.

Recalling that a Banach space is called reflexive if it is isomorphic to its double dual, we deduce that $L^p(\Omega, \mathcal{F}, \mu)$ is reflexive for $1 < p < \infty$. On the other hand, if $L^1$ is separable and $L^\infty$ is not separable then $L^1$ cannot be reflexive, since it can be proved that if the dual space is separable so is the initial space.

Given a Hausdorff topological space $X$, denote by $C(X)$ the linear space of all real-valued continuous functions on $X$. The minimal $\sigma$-algebra $\mathcal{B}_a$ for which all continuous (and bounded) real functions are measurable is called the Baire $\sigma$-algebra. If $X$ is a metric space then $\mathcal{B}_a$ coincides with the Borel $\sigma$-algebra $\mathcal{B}$, but in general $\mathcal{B}_a \subset \mathcal{B}$. If $X$ is compact then $C(X)$ with the sup-norm, namely, $\|f\|_\infty = \sup_x |f(x)|$ becomes a Banach space. The dual space $C(X)'$, i.e., the space of all continuous linear functional $T : C(X) \to \mathbb{R}$, with the dual norm

$$\|T\|'_\infty = \sup \{ |T(f)| : \|f\|_\infty \leq 1 \}$$

is also a Banach space.

If $X$ is a compact Hausdorff space then denote by $M(X)$ the linear space of all finite signed measures on $(X, \mathcal{B}_a)$, i.e., $\mu$ belongs to $M(X)$ if and only if $\mu$ is a linear combination (real coefficients) of finite measures, actually it suffices $\mu = \mu_1 - \mu_2$ with $\mu_i$ measures. We can check that

$$\|\mu\| = \inf \{ \mu_1(X) + \mu_2(X) : \mu = \mu_1 - \mu_2 \}$$

defines a norm, which makes $M(X)$ a Banach space. Moreover, we can write $\|\mu\| = |\mu|(X)$, where $|\mu|(X) = \mu^+(X) + \mu^-(X)$ and $\mu = \mu^+ - \mu^-$, with $\mu^+$ and $\mu^-$ measures such that for some measurable set $A$ we have $\mu^+(A) = 0$ and $\mu^-(X \setminus A) = 0$.

Theorem B.90. For every compact Hausdorff space $X$, the mapping $\mu \mapsto I_\mu$,

$$I_\mu(f) = \int_X f \, d\mu_1 - \int_X f \, d\mu_2, \quad \text{with } \mu = \mu_1 - \mu_2,$$

is a linear isometry from the space $(M(X), \|\cdot\|)$ onto $(C(X)', \|\cdot\|_\infty)$.
For instance, the reader may consult the book by Dudley [36, Theorems 6.4.1 and 7.4.1, p. 208 and p. 239] for a complete proof of the above theorems. For the extension to locally compact spaces, e.g., see Bauer [15, Sections 28–29, pp. 170–188], among others.

Remarks B.91. For locally compact space, the one-point (Alexandroff) compactification of $X$ yields the following version of Theorem B.90: If $X$ is a locally compact Hausdorff space then the dual of the space $(C_*(X), \| \cdot \|_\infty)$ of all continuous functions vanishing at infinity (i.e., $f$ such that for every $\varepsilon > 0$ there exists a compact $K_\varepsilon$ satisfying $|f(x)| \leq \varepsilon$ for every $x$ in $X \setminus K_\varepsilon$) is the space $(M(X), \| \cdot \|)$ of all finite Borel (or Radon) measures on $X$. For instance the reader may check Malliavin [86, Section II.6, pp. 94–100].

For a locally compact (Hausdorff) space $X$ denote by $C_0(X, \mathbb{R}^m)$ the linear space of all $\mathbb{R}^m$-valued continuous functions on $X$ with compact support, i.e., $f : X \to \mathbb{R}^m$ continuous and its $\text{supp}(f)$ (the closure of the set $\{x \in X : f(x) \neq 0\}$) is compact. Recall that a (outer) Radon measure on $X$ is a (signed) measure defined on the Borel $\sigma$-algebra which is finite for every compact subset of $X$.

**Theorem B.92.** Let $T : C_0(X, \mathbb{R}^m) \to \mathbb{R}$ be a linear functional satisfying
\[
\|T\|_K = \sup \{ T(f) : f \in C_0(X, \mathbb{R}^m), |f| \leq 1, \text{supp}(f) \subset K \} < \infty,
\]
for every compact subset $K$ of $X$. Then $\mu$ defined by
\[
\mu(U) = \sup \{ T(f) : f \in C_0(X, \mathbb{R}^m), |f| \leq 1, \text{supp}(f) \subset U \},
\]
for every open set $U$, is a Radon measure on $X$. Moreover, we have
\[
T(f) = \int_X f \sigma \, d\mu, \quad \forall f \in C_0(X, \mathbb{R}^m),
\]
where $\sigma : \mathbb{R}^m \to \mathbb{R}$ is a $\mu$-measurable function such that $|\sigma| = 1$. \(\square\)

For instance, a proof of this result (for $X = \mathbb{R}^n$) can be found in Evans and Gariepy [43, Section 1.8, pp. 49–54]. A simplified version (of this section and the previous one) is discussed in Stroock [118, Chapter 7, pp. 139–158]. In general, the reader may check Folland [44, Chapter 7, pp. 211–233] for a discussion on Radon measures and functional; and perhaps take a look at Kubrusly [75, Chapter 12, pp. 223–246] for some more details.
Notation

Some Common Uses:

\(\mathbb{N}, \mathbb{Q}, \mathbb{R}, \mathbb{C}\): natural, rational, real and complex numbers.

\(i, \Re(\cdot), I\): imaginary unit, the real part of complex number and the identity (or inclusion) mapping or operator.

\(\mathbb{1}_A\): usually denotes the characteristic function of a set \(A\), i.e., \(\mathbb{1}_A(x) = 1\) if \(x\) belongs to \(A\) and \(\mathbb{1}_A(x) = 0\) otherwise. Sometimes the set \(A\) is given as a condition on a function \(\tau\), e.g., \(\tau < t\), in this case \(\mathbb{1}_{\tau < t}(\omega) = 1\) if \(\tau(\omega) < t\) and \(\mathbb{1}_{\tau < t}(\omega) = 0\) otherwise.

\(\delta\): most of the times this is the \(\delta\) function or Dirac measure. Sometimes one write \(\delta_x(dy)\) to indicate the integration variable \(y\) and the mass concentrated at the point \(x\).

\(d\mu, \mu(dx), d\mu(x)\): together with the integration sign, usually these expressions denote integration with respect to the measure \(\mu\). Most of the times \(dx\) means integration respect to the Lebesgue measure in the variable \(x\), as understood from the context.

\(E^T, \mathcal{B}(E^T), \mathcal{B}^T(E)\): for \(E\) a Hausdorff topological (usually a separable complete metric, i.e., Polish) space and \(T\) a set of indexes, usually this denotes the product topology, i.e., \(E^T\) is the space of all function from \(T\) into \(E\) and if \(T\) is countable then \(E^T\) is the space of all sequences of elements in \(E\). As expected, \(\mathcal{B}(E^T)\) is the \(\sigma\)-algebra of \(E^T\) generated by the product topology in \(E^T\), but \(\mathcal{B}^T(E)\) is the product \(\sigma\)-algebra of \(\mathcal{B}(E)\) or generated by the so-called cylinder sets. In general \(\mathcal{B}^T(E) \subset \mathcal{B}(E^T)\) and the inclusion may be strict.

Most Commonly Used Function Spaces:

\(C(X)\): for \(X\) a Hausdorff topological (usually a separable complete metric, i.e., Polish) space, this is the space of real-valued (or complex-valued) continuous functions on \(X\). If \(X\) is a compact space then this space endowed with
sup-norm is a separable Banach (complete normed vector) space. Sometimes this space may be denoted by $C^0(X)$, $C(X, \mathbb{R})$ or $C(X, \mathbb{C})$ depending on what is to be emphasized.

$C_b(X)$: for $X$ a Hausdorff topological (usually a complete separable metric, i.e., Polish) space, this is the Banach space of real-valued (or complex-valued) continuous and bounded functions on $X$, with the sup-norm.

$C_0(X)$: for $X$ a \textit{locally compact} (but not compact) Hausdorff topological (usually a complete separable metric, i.e., Polish) space, this is the separable Banach space of real-valued (or complex-valued) continuous functions vanishing at infinity on $X$, i.e., a continuous function $f$ belongs to $C_0(X)$ if for every $\varepsilon > 0$ there exists a compact subset $K = K_\varepsilon$ of $X$ such that $|f(x)| \leq \varepsilon$ for every $x$ in $X \setminus K$. This is a proper subspace of $C_b(X)$ with the sup-norm.

$C_0(X)$: for $X$ a \textit{compact} subset of a locally compact Hausdorff topological (usually a Polish) space, this is the separable Banach space of real-valued (or complex-valued) continuous functions vanishing on the boundary of $X$, with the sup-norm. In particular, if $X = X_0 \cup \{\infty\}$ is the one-point compactification of $X_0$ then the boundary of $X$ is only $\{\infty\}$ and $C_0(X) = C_0(X_0)$ via the zero-extension identification.

$C_0(X)$, $C^0_0(X)$: for $X$ a proper \textit{open} subset of a locally compact Hausdorff topological (usually a Polish) space, this is the separable Fréchet (complete locally convex vector) space of real-valued (or complex-valued) continuous functions with a compact support $X$, with the inductive topology of uniformly convergence on compact subset of $X$. When necessary, this Fréchet space may be denoted by $C^0_0(X)$ to stress the difference with the Banach space $C_0(X)$, when $X$ is also regarded as a locally compact Hausdorff topological. Usually, the context determines whether the symbol represents the Fréchet or the Banach space.

$C^k_b(E)$, $C^k_0(E)$: for $E$ a \textit{domain} in the Euclidean space $\mathbb{R}^d$ (i.e, the closure of the interior of $E$ is equal to the closure of $E$) and $k$ a nonnegative integer, this is the subspace of either $C_b(E)$ or $C^0_0(E)$ of functions $f$ such that all derivatives up to the order $k$ belong to either $C_b(E)$ or $C^0_0(E)$, with the natural norm or semi-norms. For instance, if $E$ is open then $C^k_b(E)$ is a separable Fréchet space with the inductive topology of uniformly convergence (of the function and all derivatives up to the order $k$ included) on compact subset of $E$. If $E$ is closed then $C^k_b(E)$ is the separable Banach space with the sup-norm for the function and all derivatives up to the order $k$ included. Clearly, this is extended to the case $k = \infty$.

$B(X)$: for $X$ a Hausdorff topological (mainly a Polish) space, this is the Banach space of real-valued (or complex-valued) Borel measurable and bounded functions on $X$, with the sup-norm. Note that $\mathcal{B}(X)$ denotes the $\sigma$-algebra of Borel subsets of $X$, i.e., the smaller $\sigma$-algebra containing all open sets in
$X$, e.g., $B(\mathbb{R}^d)$, $\mathcal{B}(\mathbb{R}^d)$, or $B(E)$, $\mathcal{B}(E)$ for a Borel subset $E$ of $d$-dimensional Euclidean space $\mathbb{R}^d$.

$L^p(X,m)$: for $(X, \mathcal{X}, m)$ a complete $\sigma$-finite measure space and $1 \leq p < \infty$, this is the separable Banach space of real-valued (or complex-valued) $\mathcal{X}$-measurable (class) functions $f$ on $X$ such that $|f|^p$ is $m$-integrable, with the natural $p$-norm. If $p = 2$ this is also a Hilbert space. Usually, $X$ is also a locally compact Polish space and $m$ is a Radon measure, i.e., finite on compact sets. Moreover $L^\infty(X,m)$ is the space of all (class of) $m$-essentially bounded (i.e., bounded except in a set of zero $m$-measure) with essential-sup norm.

$L^p(\mathcal{O}), H^m_0(\mathcal{O}), H^m(\mathcal{O})$: for $\mathcal{O}$ an open subset of $\mathbb{R}^d$, $1 \leq p \leq \infty$ and $m = 1, 2, \ldots$, these are the classic Lebesgue and Sobolev spaces. Sometimes we may use vector-valued functions, e.g., $L^p(\mathcal{O}, \mathbb{R}^n)$.

$\mathcal{D}(\mathcal{O}), S(\mathbb{R}^d), \mathcal{D}'(\mathcal{O}), S'(\mathbb{R}^d)$: for $\mathcal{O}$ an open subset of $\mathbb{R}^d$, these are the classic test functions ($C^\infty$ functions with either compact support in $\mathcal{O}$ or rapidly decreasing in $\mathbb{R}^d$) and their dual spaces of distributions. These are separable Fréchet spaces with the inductive topology. Moreover, $S(\mathbb{R}^d) = \cap_m H^m(\mathbb{R}^d)$ is a countable Hilbertian nuclear space. Thus its dual space $S'(\mathbb{R}^d) = \cup_m H^{-m}(\mathbb{R}^d)$, where $H^{-m}(\mathbb{R}^d)$ is the dual space of $H^m(\mathbb{R}^d)$. Sometimes we may use vector-valued functions, e.g., $S(\mathbb{R}^d, \mathbb{R}^n)$.  

Bibliography


404 Bibliography


### Index

<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell$-class</td>
<td>320</td>
</tr>
<tr>
<td>$\mu$-equicontinuous</td>
<td>24</td>
</tr>
<tr>
<td>$\mu$-uniformly integrable</td>
<td>24</td>
</tr>
<tr>
<td>$\mu$-uniformly integrable of order $p$</td>
<td>35</td>
</tr>
<tr>
<td>$\mu^*$-measurable</td>
<td>329</td>
</tr>
<tr>
<td>$\mu^*$-measurable</td>
<td>335</td>
</tr>
<tr>
<td>$\pi$-class</td>
<td>320</td>
</tr>
<tr>
<td>$\sigma$-algebra</td>
<td>320</td>
</tr>
<tr>
<td>$\sigma$-finite</td>
<td>373</td>
</tr>
<tr>
<td>$\sigma$-finite transition measure</td>
<td>358</td>
</tr>
<tr>
<td>$\sigma$-ring</td>
<td>320</td>
</tr>
<tr>
<td>$p$-uniformly integrable</td>
<td>35</td>
</tr>
<tr>
<td>absolutely continuous</td>
<td>373</td>
</tr>
<tr>
<td>absorbing</td>
<td>98</td>
</tr>
<tr>
<td>algebra</td>
<td>320</td>
</tr>
<tr>
<td>almost everywhere</td>
<td>348</td>
</tr>
<tr>
<td>almost measurable</td>
<td>349</td>
</tr>
<tr>
<td>argument of monotone class</td>
<td>321</td>
</tr>
<tr>
<td>Arzela-Ascoli Theorem</td>
<td>63</td>
</tr>
<tr>
<td>atoms</td>
<td>328</td>
</tr>
<tr>
<td>Baire category Theorem</td>
<td>65</td>
</tr>
<tr>
<td>balanced</td>
<td>66, 98</td>
</tr>
<tr>
<td>Banach-Steinhaus Theorem</td>
<td>69</td>
</tr>
<tr>
<td>barrel space</td>
<td>102, 108, 111</td>
</tr>
<tr>
<td>basis</td>
<td>371</td>
</tr>
<tr>
<td>Beppo Levi Theorem</td>
<td>354</td>
</tr>
<tr>
<td>Bochner Theorem</td>
<td>188</td>
</tr>
<tr>
<td>Bochner’s integral</td>
<td>45</td>
</tr>
<tr>
<td>bounded</td>
<td>99</td>
</tr>
<tr>
<td>by local coordinates</td>
<td>156</td>
</tr>
<tr>
<td>canonical sample space</td>
<td>327</td>
</tr>
<tr>
<td>Caratheodory’s construction</td>
<td>329, 335</td>
</tr>
<tr>
<td>Cauchy sequence</td>
<td>67</td>
</tr>
<tr>
<td>Cauchy sequence in measure</td>
<td>344</td>
</tr>
<tr>
<td>Cauchy sequence in probability</td>
<td>344</td>
</tr>
<tr>
<td>change of variable</td>
<td>377</td>
</tr>
<tr>
<td>characteristic function</td>
<td>183</td>
</tr>
<tr>
<td>closed graph theorem</td>
<td>71</td>
</tr>
<tr>
<td>closed graphs</td>
<td>71</td>
</tr>
<tr>
<td>closed operators</td>
<td>71</td>
</tr>
<tr>
<td>compact</td>
<td>57</td>
</tr>
<tr>
<td>complete</td>
<td>67</td>
</tr>
<tr>
<td>completely regular</td>
<td>99</td>
</tr>
<tr>
<td>cone property</td>
<td>165</td>
</tr>
<tr>
<td>content</td>
<td>334</td>
</tr>
<tr>
<td>convergence in measure</td>
<td>345</td>
</tr>
<tr>
<td>converges weakly</td>
<td>77</td>
</tr>
<tr>
<td>convex</td>
<td>97</td>
</tr>
<tr>
<td>convolution</td>
<td>388</td>
</tr>
<tr>
<td>countable generated</td>
<td>321</td>
</tr>
<tr>
<td>cylindrical sets</td>
<td>323</td>
</tr>
<tr>
<td>density in Sobolev spaces</td>
<td>153</td>
</tr>
<tr>
<td>diagonal argument</td>
<td>78</td>
</tr>
<tr>
<td>Dini’s Theorem</td>
<td>3</td>
</tr>
<tr>
<td>distribution function</td>
<td>84</td>
</tr>
<tr>
<td>distribution sense</td>
<td>125</td>
</tr>
<tr>
<td>distributions</td>
<td>115</td>
</tr>
<tr>
<td>domain</td>
<td>58</td>
</tr>
<tr>
<td>domain of class $C^{m,\alpha}$</td>
<td>154</td>
</tr>
<tr>
<td>dual Banach space</td>
<td>47</td>
</tr>
<tr>
<td>dual norm</td>
<td>364</td>
</tr>
<tr>
<td>duality in $L^p$</td>
<td>393</td>
</tr>
<tr>
<td>Dunford-Pettis criterium</td>
<td>41, 79</td>
</tr>
<tr>
<td>Egorov Theorem</td>
<td>347</td>
</tr>
<tr>
<td>elementary functions</td>
<td>2</td>
</tr>
<tr>
<td>equi-bounded</td>
<td>63</td>
</tr>
<tr>
<td>equi-continuous</td>
<td>63, 68</td>
</tr>
<tr>
<td>equivalence class</td>
<td>349</td>
</tr>
<tr>
<td>essentially</td>
<td>43</td>
</tr>
<tr>
<td>Term</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>extension of measures</td>
<td>332</td>
</tr>
<tr>
<td>exterior measure</td>
<td>329</td>
</tr>
<tr>
<td>F-space</td>
<td>67</td>
</tr>
<tr>
<td>Fatou Theorem</td>
<td>355</td>
</tr>
<tr>
<td>finite</td>
<td>373</td>
</tr>
<tr>
<td>finite order</td>
<td>127</td>
</tr>
<tr>
<td>finite parts</td>
<td>120</td>
</tr>
<tr>
<td>Fréchet space</td>
<td>98</td>
</tr>
<tr>
<td>Frechet-Kolmogorov Theorem</td>
<td>82</td>
</tr>
<tr>
<td>Fubini-Tonelli Theorem</td>
<td>361</td>
</tr>
<tr>
<td>functionals on $C_0(X, \mathbb{R}^m)$</td>
<td>395</td>
</tr>
<tr>
<td>Hölder inequality</td>
<td>362</td>
</tr>
<tr>
<td>Haar measure</td>
<td>368</td>
</tr>
<tr>
<td>Hadamard finite part</td>
<td>120</td>
</tr>
<tr>
<td>Hahn-Banach Lemma</td>
<td>72</td>
</tr>
<tr>
<td>Hahn-Banach Theorem</td>
<td>73, 104</td>
</tr>
<tr>
<td>Hahn-Jordan decomposition</td>
<td>373</td>
</tr>
<tr>
<td>Hardy’s inequality</td>
<td>178</td>
</tr>
<tr>
<td>Hilbert space</td>
<td>365</td>
</tr>
<tr>
<td>inductive limits</td>
<td>60, 106</td>
</tr>
<tr>
<td>inner measure</td>
<td>335</td>
</tr>
<tr>
<td>integrable</td>
<td>8, 45, 353</td>
</tr>
<tr>
<td>integral</td>
<td>353</td>
</tr>
<tr>
<td>interior measure</td>
<td>335</td>
</tr>
<tr>
<td>Jacobian</td>
<td>378</td>
</tr>
<tr>
<td>jumps</td>
<td>120</td>
</tr>
<tr>
<td>lattice</td>
<td>2</td>
</tr>
<tr>
<td>lctvs</td>
<td>98</td>
</tr>
<tr>
<td>Lebesgue decomposition</td>
<td>375</td>
</tr>
<tr>
<td>Lebesgue invariance</td>
<td>343</td>
</tr>
<tr>
<td>Lebesgue points</td>
<td>390</td>
</tr>
<tr>
<td>Lebesgue Theorem</td>
<td>355</td>
</tr>
<tr>
<td>Lipschitz domain</td>
<td>155</td>
</tr>
<tr>
<td>Lipschitz function</td>
<td>378</td>
</tr>
<tr>
<td>local structure of distributions</td>
<td>144</td>
</tr>
<tr>
<td>locally</td>
<td>389</td>
</tr>
<tr>
<td>locally compact</td>
<td>57, 341</td>
</tr>
<tr>
<td>locally convex topological vector space</td>
<td>98</td>
</tr>
<tr>
<td>locally integrable</td>
<td>389</td>
</tr>
<tr>
<td>measurable</td>
<td>324</td>
</tr>
<tr>
<td>measurable space</td>
<td>322</td>
</tr>
<tr>
<td>measure semi-finite</td>
<td>336</td>
</tr>
<tr>
<td>Meyers-Serrin Theorem</td>
<td>152</td>
</tr>
<tr>
<td>Minkowski inequality</td>
<td>363, 364</td>
</tr>
<tr>
<td>mollification</td>
<td>388</td>
</tr>
<tr>
<td>monotone class</td>
<td>320, 321</td>
</tr>
<tr>
<td>Montel spaces</td>
<td>104, 146, 148</td>
</tr>
<tr>
<td>Morrey’s inequality</td>
<td>163</td>
</tr>
<tr>
<td>nowhere dense</td>
<td>65</td>
</tr>
<tr>
<td>open mappings</td>
<td>70</td>
</tr>
<tr>
<td>orthogonal</td>
<td>365, 371</td>
</tr>
<tr>
<td>orthogonal complement</td>
<td>365</td>
</tr>
<tr>
<td>orthogonal projection</td>
<td>365, 366</td>
</tr>
<tr>
<td>orthonormal</td>
<td>371</td>
</tr>
<tr>
<td>orthonormal basis</td>
<td>371</td>
</tr>
<tr>
<td>outer measure</td>
<td>329</td>
</tr>
<tr>
<td>partition of the unity</td>
<td>383, 390, 391</td>
</tr>
<tr>
<td>Poincaré’s inequality</td>
<td>162</td>
</tr>
<tr>
<td>polar decomposition</td>
<td>378</td>
</tr>
<tr>
<td>Polish space</td>
<td>67</td>
</tr>
<tr>
<td>pre-compact</td>
<td>63</td>
</tr>
<tr>
<td>pre-compact in $L^p$</td>
<td>80</td>
</tr>
<tr>
<td>pre-integral</td>
<td>3</td>
</tr>
<tr>
<td>pre-measure</td>
<td>334</td>
</tr>
<tr>
<td>product $\sigma$-algebra</td>
<td>323</td>
</tr>
<tr>
<td>product decomposition</td>
<td>376</td>
</tr>
<tr>
<td>product measure</td>
<td>358</td>
</tr>
<tr>
<td>quasi-integrable</td>
<td>353</td>
</tr>
<tr>
<td>quasi-normed space</td>
<td>98</td>
</tr>
<tr>
<td>Radon-Nikodym Theorem</td>
<td>374</td>
</tr>
<tr>
<td>reflexive</td>
<td>104, 394</td>
</tr>
<tr>
<td>relatively compact</td>
<td>63</td>
</tr>
<tr>
<td>Riemann integrable</td>
<td>357</td>
</tr>
<tr>
<td>Riemann-Lebesgue Theorem</td>
<td>190</td>
</tr>
<tr>
<td>Riesz Representation</td>
<td>367</td>
</tr>
<tr>
<td>ring</td>
<td>320</td>
</tr>
<tr>
<td>second category</td>
<td>65</td>
</tr>
<tr>
<td>segment property</td>
<td>153</td>
</tr>
<tr>
<td>seminorms</td>
<td>104</td>
</tr>
<tr>
<td>separability</td>
<td>371</td>
</tr>
<tr>
<td>separable</td>
<td>321–323</td>
</tr>
</tbody>
</table>
separation of convex sets, 104
sets of first category, 65
signed measure, 373
simple function, 327
singular, 373
singular integrals, 120
smooth domain, 154
smooth extension, 155
smooth functions, 129
Sobolev imbedding, 165
space of measures, 394
span, 371
strong convergence, 77
strong local Lipschitz property, 165
strong topology, 103
summable, 353
surface measure, 378
tempered distributions, 115
test functions, 111
topological vector space, 66
topology, 322
totally bounded, 63
transition probability measure, 358
translation, 123
truncation, 388
uniform absolutely continuous, 28
uniform integrability, 33
uniformly bounded, 63
uniformly boundedness principle, 69
vanishes near $x_0$, 127
Vitali Theorem, 29
Vitali-Hahn-Saks Theorem, 75
weak derivative, 125
weak topology, 103
weak* convergence, 77
weak* topology, 103
weakest topology, 323
weakly almost measurable, 46
Young inequality, 369