Chapter 12

Measure Theory and Function Spaces

In this chapter, we describe the basic ideas of measure theory and L^p spaces. We also define Sobolev spaces and summarize some of their main properties. Many results will be stated without proof, and we will not construct the most important example of a measure, namely, Lebesgue measure. Nevertheless, we hope that this discussion will allow the reader to use the concepts and results of measure theory as they are required in various applications.

12.1 Measures

The notion of measure generalizes the notion of volume. A *measure* μ on a set X associates to a subset A of X a nonnegative number $\mu(A)$, called the measure of A. It is convenient to allow for the possibility that the measure of a set may be infinite. It is too restrictive, in general, to require that the measure of every subset of X is well defined. Some sets may be too wild to define their measures in a consistent way. Sets that do have a well-defined measure are called *measurable* sets. Thus, a measure μ is a nonnegative, extended real-valued function defined on a collection of measurable subsets of X. We require that the measurable sets form a σ*-algebra*, meaning that complements, countable unions, and countable intersections of measurable sets are measurable. Moreover, as suggested by the properties of volumes, we require that the measure be *countably additive*, meaning that the measure of a countable union of disjoint sets is the sum of the measures of the individual sets. First, we give the formal definition of a σ-algebra.

Definition 12.1 A σ*-algebra* on a set X is a collection $\mathcal{A}$ of subsets of X such that:

(a) $\emptyset \in \mathcal{A}$;
(b) if $A \in \mathcal{A}$, then $A^c = X \setminus A \in \mathcal{A}$;
(c) if $\{A_i \mid i \in \mathbb{N}\}$ is a countable family of sets in $\mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

A *measurable space* $(X, \mathcal{A})$ is a set X and a σ-algebra $\mathcal{A}$ on X. The elements of $\mathcal{A}$

are called *measurable sets*.

It follows from the definition that $X \in \mathcal{A}$, and $\mathcal{A}$ is closed under countable intersections, since

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c.$$

Example 12.2 The smallest σ-algebra on an arbitrary set X is $\{\emptyset, X\}$. The largest σ-algebra is the power set $\mathcal{P}(X)$, that is, the collection of all subsets of X.

If $\mathcal{F}$ is an arbitrary collection of subsets of a set X, then the σ-algebra $\mathcal{A}(\mathcal{F})$ *generated* by $\mathcal{F}$ is the smallest σ-algebra on X that contains $\mathcal{F}$. This σ-algebra is the intersection of all σ-algebras on X that contain $\mathcal{F}$.

Example 12.3 Suppose that $(X, \mathcal{T})$ is a topological space, where $\mathcal{T}$ is the collection of open sets in X. The σ-algebra on X generated by $\mathcal{T}$ is called the *Borel σ-algebra* of X. We denote it by $\mathcal{R}(X)$. Since a σ-algebra is closed under complements, the Borel σ-algebra contains all closed sets, and is also generated by the collection of closed sets in X. Elements of the Borel σ-algebra are called *Borel sets*.

Example 12.4 The Borel σ-algebra of $\mathbb{R}$, with its usual topology, is generated by the collection of all open intervals in $\mathbb{R}$, since every open set is a countable union of open intervals. The collection of half-open intervals $\{(a, b] \mid a < b\}$ also generates $\mathcal{R}(\mathbb{R})$ (see Exercise 12.1). More generally, the Borel σ-algebra of $\mathbb{R}^n$ is generated by the collection of all cubes C of the form

$$C = (a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_n, b_n), \tag{12.1}$$

where $a_i < b_i$. It is tempting to try and construct the Borel sets by forming the collection of countable unions of closed sets, or the collection of countable intersections of open sets. The union of these collections, however, is not a σ-algebra. It can be shown that an uncountably infinite iteration of the formation of countable intersections and unions is required to obtain the Borel σ-algebra on $\mathbb{R}^n$, starting from the open sets. Thus, the structure of a general Borel set is complicated. This fact explains the nonexplicit definition of the σ-algebra generated by a collection of sets.

Next, we define measures and introduce some convenient terminology.

Definition 12.5 A *measure* μ on a set X is a map $\mu : \mathcal{A} \to [0, \infty]$ on a σ-algebra $\mathcal{A}$ of X, such that:

(a) $\mu(\emptyset) = 0$;

(b) if $\{A_i \mid i \in \mathbb{N}\}$ is a countable family of mutually disjoint sets in $\mathcal{A}$, meaning that $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{12.2}$$

The measure is *finite* if $\mu(X) < \infty$, and *σ-finite* if there is a countable family $\{A_i \in \mathcal{A} \mid i = 1, 2, \ldots\}$ of measurable subsets of X such that $\mu(A_i) < \infty$ and

$$X = \bigcup_{i=1}^{\infty} A_i.$$

A *measure space* is a triple $(X, \mathcal{A}, \mu)$ consisting of a set X, a σ-algebra $\mathcal{A}$ on X, and a measure $\mu : \mathcal{A} \to [0, \infty]$.

In the countable additivity condition (12.2), we make the natural convention that the sum of a divergent series of nonnegative terms is ∞. This countable additivity condition on μ makes sense because $\mathcal{A}$ is closed under countable unions. We will often write a measure space as (X, μ), or X, when the σ-algebra $\mathcal{A}$, or the measure μ, is clear from the context. It is also useful to consider *signed measures*, which take positive or negative values, *complex measures*, which take complex values, and *vector-valued measures*, which take values in a linear space, but we will not do so here.

Example 12.6 Let X be an arbitrary set and $\mathcal{A}$ the σ-algebra consisting of all subsets of X. The *counting measure* ν on X is defined by

$$\nu(A) = \text{the number of elements of } A,$$

with the convention that if A is an infinite set, then $\nu(A) = \infty$. The counting measure is finite if X is a finite set, and σ-finite if X is countable.

Example 12.7 We define the delta measure δ_{x_0} supported at $x_0 \in \mathbb{R}^n$ on the Borel σ-algebra $\mathcal{R}(\mathbb{R}^n)$ of $\mathbb{R}^n$ by

$$\delta_{x_0}(A) = \begin{cases} 1 & \text{if } x_0 \in A, \\ 0 & \text{if } x_0 \notin A. \end{cases}$$

This measure describes a "mass" distribution on $\mathbb{R}^n$ corresponding to a unit mass located at x_0. The formal density of this distribution is the delta function supported at x_0.

If a σ-algebra $\mathcal{A}$ is generated by a collection of sets $\mathcal{F}$, then we would like to define a measure on $\mathcal{A}$ by specifying its values on $\mathcal{F}$. The following theorem gives a useful sufficient condition to do this. A separate question, which we do not consider here, is when a function $\mu : \mathcal{F} \to [0, \infty]$ may be extended to a measure on the σ-algebra $\mathcal{A}(\mathcal{F})$ generated by $\mathcal{F}$.

Theorem 12.8 Suppose that $\mathcal{A}$ is the σ-algebra on X generated by the collection of sets $\mathcal{F}$. Let μ and ν be two measures on $\mathcal{A}$ such that

$$\mu(A) = \nu(A) \qquad \text{for every } A \in \mathcal{F}.$$

If there is a countable family of sets $\{A_i\} \subset \mathcal{F}$ such that $\bigcup_i A_i = X$ and $\mu(A_i) < \infty$, then $\mu = \nu$.

The following example of Lebesgue measure is fundamental.

Example 12.9 The Borel σ-algebra of $\mathbb{R}^n$, defined in Example 12.3, is generated by the collection of cubes C in (12.1). *Lebesgue measure* is the measure λ on the Borel σ-algebra $\mathcal{R}(\mathbb{R}^n)$ such that $\lambda(C) = \mathrm{Vol}(C)$, meaning that

$$\lambda\left((a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_n, b_n)\right) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n).$$

Lebesgue measure is σ-finite, since

$$\mathbb{R}^n = \bigcup_{i=1}^{\infty} (-i, i)^n .$$

As we discuss in Example 12.14 below, Lebesgue measure may be extended to the larger σ-algebra $\mathcal{L}(\mathbb{R}^n)$ of Lebesgue measurable sets, which is the completion of the Borel σ-algebra with respect to Lebesgue measure.

We have not explained why Lebesgue measure should exist at all, but Theorem 12.8 implies that it is unique if it exists. One can prove the existence of Lebesgue measure by construction, although the proof is not easy. The construction shows the following result.

Theorem 12.10 A subset A of $\mathbb{R}^n$ is Lebesgue measurable if and only if for every $\epsilon > 0$, there is a closed set F and an open set G such that $F \subset A \subset G$ and $\lambda(G \setminus F) < \epsilon$. Moreover,

$$\begin{aligned} \lambda(A) &= \inf\left\{\lambda(U) \mid U \text{ is open and } U \supset A\right\} \\ &= \sup\left\{\lambda(K) \mid K \text{ is compact and } K \subset A\right\}. \end{aligned}$$

Thus, a Lebesgue measurable set may be approximated from the outside by open sets, and from the inside by compact sets.

Lebesgue measure has several natural geometrical properties. It is translationally invariant, meaning that for every $A \in \mathcal{L}(\mathbb{R}^n)$ and $h \in \mathbb{R}^n$, we have $\lambda(\tau_h A) = \lambda(A)$, where

$$\tau_h A = \{y \in \mathbb{R}^n \mid y = x + h \text{ for some } x \in A\}.$$

The space $\mathbb{R}^n$ is a commutative group with respect to addition. An invariant measure on a locally compact group, such as Lebesgue measure, is called a *Haar measure.* If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map and

$$TA = \{y \in \mathbb{R}^n \mid y = Tx \text{ for some } x \in A\},$$

then $\lambda(TA) = |\det T|\lambda(A)$. Thus, Lebesgue measure is rotationally invariant, and it has the scaling property that $\lambda(tA) = t^n\lambda(A)$ for $t > 0$.

If $(X, \mathcal{A}, \mu)$ is a measure space, a subset A of X is said to have *measure zero* if it is measurable and $\mu(A) = 0$. Sets of measure zero play a particularly important role in measure theory and integration.

Example 12.11 A subset A of $\mathbb{R}^n$ is of measure zero with respect to the delta-measure δ_{x_0} defined in Example 12.7 if and only if A is a Borel set and $x_0 \notin A$.

Example 12.12 For each $x \in \mathbb{R}$, the Lebesgue measure of the set $\{x\}$ is equal to zero, since

$$\lambda(\{x\}) = \lim_{\epsilon\to 0^+} \lambda(\{y \mid |x - y| < \epsilon\}) = \lim_{\epsilon\to 0^+} 2\epsilon = 0.$$

Hence every countable subset $A = \{x_i \mid i \in \mathbb{N}\}$ of $\mathbb{R}$ has measure zero, since the countable additivity of Lebesgue measure implies that

$$\lambda(A) = \sum_{i=1}^{\infty} \lambda(\{x_i\}) = 0.$$

One can show that a subset A of $\mathbb{R}^n$ has Lebesgue measure zero if and only if for every $\epsilon > 0$, A is contained in a not necessarily disjoint union of open cubes, the sum of whose volumes is less than ϵ.

It follows from the additivity and nonegativity of a measure that any measurable subset of a set of measure zero has zero measure. We may extend a measure in a unique fashion to every subset of a set of measure zero by defining it to be zero.

Definition 12.13 A measure space is *complete* if every subset of a set of measure zero is measurable. If $(X, \mathcal{A}, \mu)$ is a measure space, the *completion* $\overline{\mathcal{A}}$ of the σ-algebra $\mathcal{A}$ with respect to a measure μ on $\mathcal{A}$ consists of all subsets A of X such that there exists sets E and F in $\mathcal{A}$ with

$$E \subset A \subset F, \qquad \text{and} \qquad \mu(F \setminus E) = 0.$$

The completion $\overline{\mu}$ of μ is defined on $\overline{\mathcal{A}}$ by

$$\overline{\mu}(A) = \mu(E) = \mu(F).$$

The complete measure space $\left(X, \overline{\mathcal{A}}, \overline{\mu}\right)$ is called the *completion* of $(X, \mathcal{A}, \mu)$.

Example 12.14 The Borel σ-algebra $\mathcal{R}(\mathbb{R}^n)$ is not complete. Its completion with respect to Lebesgue measure is the σ-algebra $\mathcal{L}(\mathbb{R}^n)$ of Lebesgue measurable sets. We will use the same notation λ to denote Lebesgue measure on the Borel sets and the Lebesgue measurable sets.

A property that holds except on a set of measure zero is said to hold *almost everywhere*, or *a.e.* for short. When we want to make explicit the measure μ with respect to which a set has measure zero, we write μ-a.e. We define the *essential supremum* of a set of real numbers $A \subset \mathbb{R}$ by

$$\text{ess sup}\, A = \inf\{C \mid x \leq C \text{ for all } x \in A \setminus N, \text{ where } \mu(N) = 0\}.$$

A *Borel measure* is a measure defined on the Borel σ-algebra of a topological space. Thus, the delta-measure and Lebesgue measure defined on $\mathcal{R}(\mathbb{R}^n)$ are examples of Borel measures. The following example gives a useful class of Borel measures on $\mathbb{R}$.

Example 12.15 Let $F : \mathbb{R} \to \mathbb{R}$ be an increasing, right-continuous function, meaning that $F(x) \leq F(y)$ for $x \leq y$, and

$$F(x) = \lim_{y \to x^+} F(y).$$

There is a unique measure μ_F on the Borel σ-algebra of $\mathbb{R}$ such that

$$\mu_F\left((a, b]\right) = F(b) - F(a).$$

From Exercise 12.3, if $b_n \to b^+$ is a decreasing sequence and $a < b$, then

$$\mu_F\left((a, b_n]\right) \to \mu_F\left((a, b]\right)^+,$$

which explains why F must be right-continuous. This measure is called a *Lebesgue-Stieltjes measure*, and F is called the *distribution function* of the measure. For example, if $F(x) = x$, then we obtain Lebesgue measure. If F is the right-continuous step function,

$$F(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

then we obtain the delta measure supported at the origin. If F is the Cantor function, defined in Exercise 1.19, then we obtain a measure such that the Cantor set C has measure one and $\mathbb{R} \setminus C$ has measure zero. Despite the fact that μ_F is supported on a set of Lebesgue measure zero, we have $\mu_F(\{x\}) = 0$ for every $x \in \mathbb{R}$.

Kolmogorov observed in the 1920s that measure theory provides the mathematical foundation of probability theory.

Definition 12.16 A *probability space* $(\Omega, \mathcal{A}, \mu)$ is a measure space such that $\mu(\Omega) = 1$. The measure μ is called a *probability measure.*

In modeling a random trial or experiment, we form a *sample space* Ω that consists of all possible outcomes of the trial, including outcomes that may occur with probability zero. An *event* is a measurable subset of Ω, and the collection of events forms a σ-algebra $\mathcal{A}$ on Ω. The *probability* $0 \le \mu(A) \le 1$ of an event $A \in \mathcal{A}$ is given by an appropriate probability measure defined on $\mathcal{A}$. The σ-algebra of a probability space has a natural interpretation as the collection of events about which information is available.

Example 12.17 Let $\Omega = \{n \in \mathbb{Z} \mid n \ge 0\}$ be the nonnegative integers and $\mathcal{A}$ the set of all subsets of Ω. Let μ be the measure on $\mathcal{A}$ such that

$$\mu(\{n\}) = \frac{\lambda^n}{n!}e^{-\lambda}, \qquad \mu(A) = \sum_{n \in A} \frac{\lambda^n}{n!}e^{-\lambda}.$$

This measure is called the *Poisson distribution*.

Example 12.18 Suppose that $\Omega = \mathbb{R}^n$ and $\mathcal{A}$ is the Lebesgue σ-algebra. The standard Gaussian probability measure on $\mathbb{R}^n$ is given by

$$\mu(A) = \frac{1}{(2\pi)^{n/2}} \int_A e^{-|x|^2/2}\, dx.$$

Here, the integral is the Lebesgue integral, defined below.

12.2 Measurable functions

Measurable functions are the natural mappings between measurable spaces. They play an analogous role to continuous functions between topological spaces.

Definition 12.19 Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A *measurable function* is a mapping $f : X \to Y$ such that

$$f^{-1}(B) \in \mathcal{A} \qquad \text{for every } B \in \mathcal{B}.$$

The measurability of $f : X \to Y$ depends only on the σ-algebras on X and Y, and not on what measure, if any, is defined on X or Y. When a measure μ is defined on X, we say that two measurable functions $f : X \to Y$ and $g : X \to Y$ are *equal a.e.* if

$$\mu\left(\{x \in X \mid f(x) \ne g(x)\}\right) = 0.$$

Two functions on a measure space that are equal a.e. will often be regarded as equivalent.

Example 12.20 A measurable map $T : X \to X$ on a measure space $(X, \mathcal{A}, \mu)$ is said to be *measure preserving* if

$$\mu\left(T^{-1}(A)\right) = \mu(A)$$

for all measurable sets A. Measure preserving maps arise naturally in physics and other applications. Ergodic theory studies the properties of various kinds of measure preserving maps (see Theorem 7.11 and Theorem 8.35, for example).

Example 12.21 A measurable map $X : \Omega \to \mathbb{R}$ on a probability space Ω is called a *random variable.*

If $\mathcal{B}$ is the σ-algebra generated by $\mathcal{F}$, then the condition that

$$f^{-1}(F) \in \mathcal{A} \qquad \text{for all } F \in \mathcal{F}$$

is sufficient to ensure that f is measurable. This follows from the fact that $\{f^{-1}(B) \mid B \in \mathcal{B}\}$ is the σ-algebra generated by $\{f^{-1}(F) \mid F \in \mathcal{F}\}$, and is therefore contained in $\mathcal{A}$.

Example 12.22 Every continuous function between topological spaces is Borel measurable. A continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is measurable with respect to the Lebesgue σ-algebra on the domain $\mathbb{R}^n$ and the Borel σ-algebra on the range $\mathbb{R}$.

From now on, we will restrict our attention to real-valued functions defined on a measure space $(X, \mathcal{A}, \mu)$. Complex-valued functions may be treated by splitting them into their real and imaginary parts. The fact that the real numbers are totally ordered makes it particularly easy to develop the integral in this case. The theory applies to real-valued functions defined on a general measure space X, but it is helpful to keep in mind the case when X is $\mathbb{R}^n$ equipped with Lebesgue measure.

It is often convenient to allow measures and functions to take on the values $-\infty$ or ∞. We therefore introduce the *extended real numbers* $\overline{\mathbb{R}} = [-\infty, \infty]$. We make the following definitions of algebraic operations involving $x \in \mathbb{R}$ and ∞:

$$\begin{aligned}
&x + \infty = \infty, \quad x - \infty = -\infty, \\
&x \cdot \infty = \infty, \quad x \cdot (-\infty) = -\infty \qquad \text{if } x > 0, \\
&x \cdot \infty = -\infty, \quad x \cdot (-\infty) = \infty \qquad \text{if } x < 0, \\
&|\infty| = |-\infty| = \infty.
\end{aligned}$$

We also define

$$0 \cdot \infty = 0 \cdot (-\infty) = 0.$$

For example, we will define the integral of a function that is infinite on a set of measure zero, or the integral of a function that is zero on a set of infinite measure, to be zero. We do not define $\infty - \infty$, and any expression of this form is meaningless.

We use the natural ordering and topology on $\overline{\mathbb{R}}$; as far as its ordering and topology are concerned, $\overline{\mathbb{R}}$ is isomorphic to the closed interval $[-1, 1]$. Any monotone sequence $\{x_n\}$ of points in $\overline{\mathbb{R}}$ has a limit. The limit of a monotone increasing sequence is $\sup\{x_n\}$ if the sequence is bounded, and ∞ if it is unbounded. The limit of a monotone decreasing sequence is $\inf\{x_n\}$ if the sequence is bounded, and $-\infty$ if it is unbounded. We equip $\overline{\mathbb{R}}$ with the Borel σ-algebra $\mathcal{R}(\overline{\mathbb{R}})$. This σ-algebra is generated by the semi-infinite intervals of the form $\{[-\infty, c) : c \in \mathbb{R}\}$, so we have the following criterion for the measurability of a function $f : X \to \overline{\mathbb{R}}$.

Proposition 12.23 Let $(X, \mathcal{A})$ be a measurable space. A function $f : X \to \overline{\mathbb{R}}$ is measurable if and only if the set $\{x \in X \mid f(x) < c\}$ belongs to $\mathcal{A}$ for every $c \in \mathbb{R}$.

In this proposition, the sets $\{f(x) \leq c\}$, $\{f(x) > c\}$, or $\{f(x) \geq c\}$ could be used equally well. A complex-valued function $f : X \to \mathbb{C}$ is measurable if and only if $f = g + ih$ where $g, h : X \to \mathbb{R}$ are measurable.

We say that a sequence of functions (f_n) from a measure space $(X, \mathcal{A}, \mu)$ to $\overline{\mathbb{R}}$ *converges pointwise* to a function $f : X \to \overline{\mathbb{R}}$ if

$$\lim_{n\to\infty} f_n(x) = f(x) \qquad \text{for every } x \in X.$$

The sequence *converges pointwise-a.e.* to f if it converges pointwise to f on $X \setminus N$, where N is a set of measure zero. The following result explains why all functions encountered in analysis are measurable.

Theorem 12.24 If (f_n) is a sequence of measurable functions that converges pointwise to a function f, then f is measurable. If $(X, \mathcal{A}, \mu)$ is a complete measure space and (f_n) converges pointwise-a.e. to f, then f is measurable.

A measurable function that takes on finitely many, finite values is called a *simple function*. Any measurable function may be approximated by simple functions.

Definition 12.25 A function $\varphi : X \to \mathbb{R}$ on a measurable space $(X, \mathcal{A})$ is a *simple function* if there are measurable sets $A_1, A_2, \ldots, A_n$ and real numbers $c_1, c_2, \ldots, c_n$ such that

$$\varphi = \sum_{i=1}^{n} c_i \chi_{A_i}. \tag{12.3}$$

Here, χ_A is the characteristic function of the set A, meaning that

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

The representation of a simple function as a sum of characteristic functions is not unique. A standard representation uses disjoint sets A_i and distinct values c_i. In that case, we have $\varphi(x) = c_i$ if and only if $x \in A_i$. Since the sets A_i are required to be measurable, a simple function is measurable.

Theorem 12.26 Let $f : X \to [0, \infty]$ be a nonnegative, measurable function. There is a monotone increasing sequence $\{\varphi_n\}$ of simple functions that converges pointwise to f.

Proof. For each $n \in \mathbb{N}$, we subdivide the range of f into $2^{2n}+1$ intervals

$$I_{n,k} = \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right) \quad \text{for } k = 1, 2, \dots, 2^{2n}, \quad I_{n,2^{2n}+1} = [2^n, \infty]$$

of length 2^{-n}. We define the measurable sets

$$A_{n,k} = f^{-1}\left(I_{n,k}\right) \qquad \text{for } k = 1, 2, \dots, 2^{2n}+1.$$

Then the increasing sequence of simple functions

$$\varphi_n = \sum_{k=1}^{2^{2n}+1} \left(\frac{k-1}{2^n}\right) \chi_{A_{n,k}}$$

converges pointwise to f as $n \to \infty$. □

An arbitrary measurable function $f : X \to \overline{\mathbb{R}}$ may be written in a canonical way as the difference of two nonnegative measurable functions,

$$f = f_+ - f_-, \qquad f_+ = \max\{f, 0\}, \quad f_- = \max\{-f, 0\}. \tag{12.4}$$

We call f_+ the *positive part* of f and f_- the *negative part.* We may then approximate each part by simple functions.

12.3 Integration

The Lebesgue integral provides an extension of the Riemann integral which applies to highly discontinuous and unbounded functions, and which behaves very well with respect to limiting operations. To construct the Lebesgue integral, we first define the integral of a simple function. We then define the integral of a general measurable function using approximations by simple functions.

Suppose that

$$\varphi = \sum_{i=1}^{n} c_i \chi_{A_i}$$

is a simple function on a measure space $(X, \mathcal{A}, \mu)$. We define the integral of φ with respect to the measure μ by

$$\int \varphi \, d\mu = \sum_{i=1}^{n} c_i \mu\left(A_i\right).$$

The value of the sum on the right-hand side is independent of how φ is represented as a sum of characteristic functions.

This definition is already well outside the scope of the Riemann integral. For instance, the characteristic function of the rationals is not Riemann integrable, but its Lebesgue integral is zero. The Riemann integral of a function $f : [a, b] \to \mathbb{R}$

is based upon the approximation of the function by simple step functions φ, in which the sets A_i are intervals. It may not be possible to approximate a highly discontinuous function by step functions, and then the Riemann definition of the integral fails. The Lebesgue integral uses approximations of the function by simple functions in which the sets A_i are general measurable sets. Because of the way the approximating simple functions are constructed in Theorem 12.26, the Lebesgue approach to integration is sometimes contrasted with the Riemann approach by saying that it subdivides the range of the function instead of the domain.

Definition 12.27 Let $f : X \to [0, \infty]$ be a nonnegative measurable function on a measure space $(X, \mathcal{A}, \mu)$. We define

$$\int f \, d\mu = \sup \left\{ \int \varphi \, d\mu \;\middle|\; \varphi \text{ is simple and } \varphi \leq f \right\}.$$

If $f : X \to \overline{\mathbb{R}}$ and $f = f_+ - f_-$, where f_+ and f_- are the positive and negative parts of f defined in (12.4), then we define

$$\int f \, d\mu = \int f_+ \, d\mu - \int f_- \, d\mu,$$

provided that at least one of the integrals on the right hand side is finite. If

$$\int |f| \, d\mu = \int f_+ \, d\mu + \int f_- \, d\mu < \infty,$$

then we say that f is *integrable* or *summable*. A complex-valued function $f : X \to \mathbb{C}$ is *integrable* if $f = g + ih$ where $g, h : X \to \mathbb{R}$ are integrable, and then

$$\int f \, d\mu = \int g \, d\mu + i \int h \, d\mu.$$

If A is a measurable subset of X, we define

$$\int_A f \, d\mu = \int f \chi_A \, d\mu.$$

The Lebesgue integral does not assign a value to the integral of a highly oscillatory function f for which both $\int f_+ \, d\mu$ and $\int f_- \, d\mu$ are infinite.

Example 12.28 The function

$$f(x) = \frac{d}{dx}\left[x^2 \sin\left(\frac{1}{x^2}\right)\right] = -\frac{2}{x}\cos\left(\frac{1}{x^2}\right) + 2x \sin\left(\frac{1}{x^2}\right),$$

is not Lebesgue integrable on $[0, 1]$. The function is not Riemann integrable on $[0, 1]$ either, since it is unbounded. Nevertheless, the improper Riemann integral

$$\lim_{\epsilon \to 0^+} \int_\epsilon^1 f(x) \, dx = \sin 1$$

exists, because of the cancellation between the large positive and negative oscillations in the integrand.

Depending on the context, we may write the integral in any of the following ways

$$\int f, \qquad \int f\, d\mu, \qquad \int f(x)\, d\mu(x), \qquad \int f(x)\, \mu(dx).$$

We will also write the integral $\int f\, d\lambda$ of a function defined on $\mathbb{R}^n$ with respect to Lebesgue measure λ as

$$\int f\, dx, \qquad \int f(x)\, dx.$$

Example 12.29 If δ_{x_0} is the delta-measure, and $f : \mathbb{R}^n \to \mathbb{R}$ is a Borel measurable function, then

$$\int f\, d\delta_{x_0} = f(x_0).$$

We have $f = g$ a.e. with respect to δ_{x_0} if and only if $f(x_0) = g(x_0)$.

Example 12.30 Let ν be the counting measure on the set $\mathbb{N}$ of natural numbers defined in Example 12.6. If $f : \mathbb{N} \to \mathbb{R}$, then

$$\int f\, d\nu = \sum_{n=1}^{\infty} f_n,$$

where $f_n = f(n)$. This integral is well defined if f is nonnegative, or if the sum on the right converges absolutely, in which case f is integrable with respect to ν. Thus, nonnegative and absolutely convergent series are a special case of the general Lebesgue integral.

Example 12.31 We denote the integral with respect to the Lebesgue-Stieltjes measure μ_F on $\mathbb{R}$ defined in Example 12.15 by

$$\int f d\mu_F = \int f dF.$$

If F is a piecewise smooth, monotone increasing function with a countable number of jump discontinuities, then the Lebesgue-Stieltjes integral includes a continuous integral from the smooth parts of F, and a discrete sum from the jumps.

12.4 Convergence theorems

Suppose that a sequence of functions (f_n) converges pointwise to a limiting function f. When can we assert that $\int f_n\, d\mu$ converges to $\int f\, d\mu$? The following example shows that some condition is required to ensure the convergence of the integrals.

Example 12.32 Define $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} n & \text{if } 0 < x < 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have $f_n(x) \to 0$ as $n \to \infty$ for every $x \in [0,1]$, but

$$\int_0^1 f_n(x)\, dx = 1$$

for every n, so $\int_0^1 f_n(x)\, dx$ does not tend to 0 as $n \to \infty$.

Two simple conditions that ensure the convergence of the integrals are the monotone convergence of the sequence, and a uniform bound on the sequence by an integrable function. The corresponding theorems, called the monotone convergence theorem and the Lebesgue dominated convergence theorem, are among the most important and frequently used results in integration theory.

A sequence of functions (f_n), where $f_n : X \to \overline{\mathbb{R}}$, is *monotone increasing* if

$$f_1(x) \leq \ldots \leq f_{n-1}(x) \leq f_n(x) \leq f_{n+1}(x) \leq \ldots \quad \text{for every } x \in X.$$

Theorem 12.33 (Monotone convergence) Suppose that (f_n) is a monotone increasing sequence of nonnegative, measurable functions $f_n : X \to [0,\infty]$ on a measure space $(X, \mathcal{A}, \mu)$. Let $f : X \to [0,\infty]$ be the pointwise limit,

$$f(x) = \lim_{n\to\infty} f_n(x).$$

Then

$$\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu.$$

The convergence of the sequence in Example 12.32 is not monotone. A generalization of this result, called Fatou's lemma, is the following.

Theorem 12.34 (Fatou) If (f_n) is any sequence of nonnegative measurable functions $f_n : X \to [0,\infty]$ on a measure space $(X, \mathcal{A}, \mu)$, then

$$\int \left(\liminf_{n\to\infty} f_n\right) d\mu \leq \liminf_{n\to\infty} \int f_n\, d\mu. \tag{12.5}$$

Example 12.32 shows that we may have strict inequality in (12.5). Intuitively, "mass" may "leak out to infinity" as $n \to \infty$, so the integral of the lim inf may be less than or equal to the lim inf of the integrals.

The crucial hypothesis in the next theorem is that every function in the sequence (f_n) is bounded independently of n by the same integrable function g. This theorem is the one of the most useful for applications.

Theorem 12.35 (Lebesgue dominated convergence) Suppose that (f_n) is a sequence of integrable functions, $f_n : X \to \overline{\mathbb{R}}$, on a measure space $(X, \mathcal{A}, \mu)$ that converges pointwise to a limiting function $f : X \to \overline{\mathbb{R}}$. If there is an integrable function $g : X \to [0, \infty]$ such that

$$|f_n(x)| \le g(x) \qquad \text{for all } x \in X \text{ and } n \in \mathbb{N},$$

then f is integrable and

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

The sequence in Example 12.32 is bounded uniformly in n by the function

$$g(x) = \begin{cases} 1/x & \text{if } 0 < x \le 1, \\ 0 & \text{if } x = 0, \end{cases}$$

but this function is not integrable on $[0, 1]$. The same result applies if X is a complete measure space, $f_n \to f$ pointwise-a.e., and $|f_n(x)| \le g(x)$ pointwise-a.e.

A corollary of the dominated convergence theorem is the following result for differentiation under an integral, which is proved by approximation of the derivative by difference quotients.

Corollary 12.36 (Differentiation under an integral) Suppose that $(X, \mathcal{A}, \mu)$ is a complete measure space, $I \subset \mathbb{R}$ is an open interval, and $f : X \times I \to \overline{\mathbb{R}}$ is a measurable function such that:

(a) $f(\cdot, t)$ is integrable on X for each $t \in I$;
(b) $f(x, \cdot)$ is differentiable in I for each $x \in X \setminus N$ where $\mu(N) = 0$;
(c) there is an integrable function $g : X \to [0, \infty]$ such that

$$\left| \frac{\partial f}{\partial t}(x, t) \right| \le g(x) \qquad \text{a.e. in } X \text{ for every } t \in I.$$

Then

$$\varphi(t) = \int_X f(x, t) \, d\mu(x)$$

is a differentiable function of t in I, and

$$\frac{d\varphi}{dt}(t) = \int_X \frac{\partial f}{\partial t}(x, t) \, d\mu(x).$$

12.5 Product measures and Fubini's theorem

One of the most elementary geometrical facts is that the area of a rectangle in the plane is the product of the lengths of its sides. From the point of view of measure theory, this means that Lebesgue measure on $\mathbb{R}^2$ is the product of Lebesgue measures on $\mathbb{R}$. We will describe a general construction of product measures here. A closely related result from elementary calculus is that the double integral of a continuous function over a smooth region in $\mathbb{R}^2$ can be computed as two iterated one-dimensional integrals. Fubini's theorem provides a generalization of this result, which states that an integral of a function on a product space can be computed as iterated integrals over the individual components of the product space. Fubini's theorem is another of the most useful results in the theory of Lebesgue integration.

The key hypothesis in Fubini's theorem is that the function is integrable on the product space. The following example shows that the equality of double and iterated integrals is not true, in general, without an integrability condition.

Example 12.37 Define $f : [0,1] \times [0,1] \to \mathbb{R}$ by

$$f(x,y) = \frac{x^2 - y^2}{(x^2 + y^2)^2}.$$

Then a straightforward computation shows that

$$\int_0^1 \left(\int_0^1 f(x,y)\, dy \right) dx = \int_0^1 \frac{1}{1+x^2}\, dx = \frac{\pi}{4},$$
$$\int_0^1 \left(\int_0^1 f(x,y)\, dx \right) dy = -\int_0^1 \frac{1}{1+y^2}\, dx = -\frac{\pi}{4}.$$

The function f in this example is not integrable, meaning that

$$\int_0^1 \int_0^1 |f(x,y)|\, dxdy = \infty.$$

First, we define the product of two σ-algebras.

Definition 12.38 Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. The *product σ-algebra* $\mathcal{A} \otimes \mathcal{B}$ is the σ-algebra on $X \times Y$ that is generated by the collection of sets

$$\{A \times B \mid A \in \mathcal{A},\ B \in \mathcal{B}\}. \tag{12.6}$$

The collection of sets in (12.6) does not form a σ-algebra, since the union of two such sets is not, in general, another such set. Next, we state a theorem which defines the product of two σ-finite measures.

Theorem 12.39 Suppose that $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are σ-finite measure spaces. There is a unique product measure $\mu \otimes \nu$, defined on $\mathcal{A} \otimes \mathcal{B}$, with the property that for every $A \in \mathcal{A}$ and $B \in \mathcal{B}$

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B).$$

Example 12.40 Suppose that $X = \mathbb{R}^m$ and $Y = \mathbb{R}^n$ are equipped with the Borel σ-algebras $\mathcal{A} = \mathcal{R}^m$ and $\mathcal{B} = \mathcal{R}^n$ and the Lebesgue measures $\mu = \lambda^m$ and $\nu = \lambda^n$. The product σ-algebra is $\mathcal{R}^m \otimes \mathcal{R}^n = \mathcal{R}^{m+n}$, and the product measure is $\lambda^m \otimes \lambda^n = \lambda^{m+n}$. Thus, Lebesgue measure on $\mathbb{R}^n$ is the n-fold product of Lebesgue measure on $\mathbb{R}$.

Let $f : X \times Y \to \overline{\mathbb{R}}$. We denote by $f^y : X \to \overline{\mathbb{R}}$ and $f_x : Y \to \overline{\mathbb{R}}$ the functions

$$f^y(x) = f(x, y), \qquad f_x(y) = f(x, y).$$

If $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are σ-finite measure spaces and $f : X \times Y \to \overline{\mathbb{R}}$ is an $\mathcal{A} \otimes \mathcal{B}$-measurable function, then one can prove that f^y is $\mathcal{A}$-measurable for every $y \in Y$ and f_x is $\mathcal{B}$-measurable for every $x \in X$. Furthermore, the function $I(x) = \int_Y f_x(y)\, d\nu(y)$ is μ-measurable, and the function $J(y) = \int_X f^y(x)\, d\mu(x)$ is ν-measurable. All the integrals appearing in the following statement of Fubini's theorem are therefore well defined.

Theorem 12.41 (Fubini) Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. Suppose that $f : X \times Y \to \overline{\mathbb{R}}$ is an $(\mathcal{A} \otimes \mathcal{B})$-measurable function.

(a) The function f is integrable, meaning that

$$\int_{X \times Y} |f|\, d\mu \otimes d\nu < \infty,$$

if and only if either of the following iterated integrals is finite:

$$\int_X \left(\int_Y |f_x(y)|\ d\nu(y) \right)\ d\mu(x),$$
$$\int_Y \left(\int_X |f^y(x)|\ d\mu(x) \right)\ d\nu(y).$$

(b) If f is integrable, then

$$\begin{aligned} \int_{X \times Y} f(x, y)\, d(\mu(x) \otimes \nu(y)) &= \int_X \left(\int_Y f_x(y)\, d\nu(y) \right)\ d\mu(x), \\ \int_{X \times Y} f(x, y)\, d(\mu(x) \otimes \nu(y)) &= \int_Y \left(\int_X f^y(x)\, d\mu(x) \right)\ d\nu(y). \end{aligned}$$

To apply this theorem, we usually check that one of the iterated integrals of $|f|$ is finite, and then compute the double integral of f by evaluation of an iterated integral.

Example 12.42 Suppose that x_{mn} is a doubly-indexed sequence of real or complex numbers, with $m, n \in \mathbb{N}$. The application of Fubini's theorem to counting measure on $\mathbb{N}$ implies that if

$$\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} |x_{mn}| \right) < \infty,$$

then

$$\sum_{m=1}^{\infty}\left(\sum_{n=1}^{\infty}x_{mn}\right)=\sum_{n=1}^{\infty}\left(\sum_{m=1}^{\infty}x_{mn}\right).$$

The product of two complete measures is not necessarily complete, and this leads to some technical complications in connection with Lebesgue measure.

Example 12.43 Suppose that $X=\mathbb{R}^m$ and $Y=\mathbb{R}^n$, with $m,n\geq 1$, are equipped with the Lebesgue σ-algebras $\mathcal{L}^m=\overline{\mathcal{R}^m}$ and $\mathcal{L}^n=\overline{\mathcal{R}^n}$, respectively. It is not true that $\mathcal{L}^m\otimes\mathcal{L}^n=\mathcal{L}^{m+n}$. Rather, we have

$$\mathcal{L}^{m+n}=\overline{\mathcal{L}^m\otimes\mathcal{L}^n}.$$

For example, if $E\subset\mathbb{R}^m$ is any non-Lebesgue measurable set (which cannot be a subset of a set of m-dimensional Lebesgue measure zero) and $y\in\mathbb{R}^n$, then the set $E\times\{y\}\subset\mathbb{R}^{m+n}$ does not belong to $\mathcal{L}^m\otimes\mathcal{L}^n$. It is, however, an $(\mathcal{L}^{m+n})$-measurable set, since it is a subset of $\mathbb{R}^m\times\{y\}$ which has $(m+n)$-dimensional Lebesgue measure zero.

The following version of Fubini's theorem applies in this context.

Theorem 12.44 Suppose that $(X,\mathcal{A},\mu)$ and $(Y,\mathcal{B},\nu)$ are complete σ-finite measure spaces, and let $(X\times Y,\mathcal{L},\lambda)$ be the completion of the product space. If $f:X\times Y\to\overline{\mathbb{R}}$ is a nonnegative or integrable $\mathcal{L}$-measurable function, then f_x is $\mathcal{B}$-measurable μ-a.e. in $x\in X$ and f^y is $\mathcal{A}$-measurable ν-a.e. in $y\in Y$. Furthermore, $I(x)=\int f_x(y)\,d\nu(y)$ and $J(y)=\int f^y(x)\,d\mu(x)$ are measurable, and

$$\int_{X\times Y}f\,d\lambda=\int_X\left(\int_Y f_x(y)\,d\nu(y)\right)\,d\mu(x)=\int_Y\left(\int_X f^y(x)\,d\mu(x)\right)\,d\nu(y).$$

12.6 The L^p spaces

The L^p-spaces consist of functions whose pth powers are integrable. The space L^∞ requires a separate definition.

Definition 12.45 Let $(X,\mathcal{A},\mu)$ be a measure space and $1\leq p<\infty$. The space $L^p(X,\mu)$ is the space of equivalence classes of measurable functions $f:X\to\mathbb{C}$, with respect to the equivalence relation of a.e.-equality, such that

$$\int |f|^p\,d\mu<\infty.$$

The L^p-norm of f is defined by

$$\|f\|_p=\left(\int_X |f|^p\,d\mu\right)^{1/p}. \tag{12.7}$$

The space $L^\infty(X,\mu)$ consists of equivalence classes of functions $f : X \to \mathbb{C}$ such that there is a finite constant M with

$$|f(x)| \le M \qquad \mu\text{-a.e.}$$

The L^∞-norm is defined by

$$\begin{aligned} \|f\|_\infty &= \text{ess sup}\,\{|f(x)| \mid x \in X\} \\ &= \inf\{M \mid |f(x)| \le M \ \mu\text{-a.e.}\}. \end{aligned}$$

We are mainly interested in the case when X is a Lebesgue measurable subset of $\mathbb{R}^n$, but most results below apply to arbitrary measure spaces. When the measure μ is understood, we often abbreviate $L^p(X,\mu)$ to $L^p(X)$, or simply L^p. For example, we write $L^p(\mathbb{R}^n)$ for $L^p(\mathbb{R}^n,\lambda)$, where λ is Lebesgue measure.

Theorem 12.46 If $(X,\mathcal{A},\mu)$ is a measure space and $1 \le p \le \infty$, then $L^p(X)$ is a Banach space.

Proof. We will only prove the result for $1 \le p < \infty$. We abbreviate $L^p(X)$ to L^p and $\|\cdot\|_p$ to $\|\cdot\|$. The verification that L^p is a linear space and that $\|\cdot\|$ is a norm is straightforward, with the exception of the triangle inequality, which we prove in Theorem 12.56 below. We therefore just have to show that L^p is complete.

From Exercise 1.20, a normed linear space is complete if and only if every absolutely convergent series converges. Suppose that $f_n \in L^p$ with $n = 1, 2, \ldots$ is a sequence of functions such that

$$\sum_{n=1}^\infty \|f_n\| = M,$$

where $M < \infty$. We need to show that there is a function $f \in L^p$ such that

$$\lim_{N\to\infty} \left\| f - \sum_{n=1}^N f_n \right\| = 0.$$

First, we consider the sequence of nonnegative functions g_N defined by

$$g_N = \sum_{n=1}^N |f_n|.$$

The sequence (g_N) is monotone increasing, so it converges pointwise to a nonnegative, measurable extended real-valued function $g : X \to [0,\infty]$. We have $\|g_N\| \le M$ for every $N \in \mathbb{N}$, so

$$\int |g_N|^p d\mu = \|g_N\|^p \le M^p.$$

The monotone convergence theorem, Theorem 12.33, implies that

$$\|g\|^p = \lim_{N\to\infty} \|g_N\|^p \le M^p.$$

In particular, $\|g\| < \infty$, so $g \in L^p$. Therefore, g is finite μ-a.e., which means that the sum $\sum_{n=1}^{\infty} f_n(x)$ is absolutely convergent μ-a.e. We can then define a function f pointwise-a.e. by

$$f(x) = \sum_{n=1}^{\infty} f_n(x).$$

The partial sums of this series satisfy

$$\left|\sum_{n=1}^{N} f_n(x)\right|^p \le g(x)^p.$$

Since $g \in L^p$ and $|f(x)| \le g(x)$, we have that $f \in L^p$. Furthermore, $g^p \in L^1$ and

$$\left|f(x) - \sum_{n=1}^{N} f_n(x)\right|^p \le (2g(x))^p.$$

The dominated convergence theorem, Theorem 12.35, implies that

$$\lim_{N\to\infty} \int \left|f - \sum_{n=1}^{N} f_n\right|^p d\mu = 0,$$

so the series converges to f in L^p, which proves that L^p is complete. □

The following example shows that the L^p-convergence of a sequence does not imply pointwise-a.e. convergence. One can prove, however, that if $f_n \to f$ in L^p, then there is a subsequence of (f_n) that converges pointwise-a.e. to f.

Example 12.47 For $2^k \le n \le 2^{k+1}-1$, where $k = 0, 1, 2, \ldots$, we define the interval I_n by

$$I_n = \left[(n-2^k)/2^k, (n+1-2^k)/2^k\right],$$

and the function $f_n : [0,1] \to \mathbb{R}$ by

$$f_n = \chi_{I_n}.$$

The sequence (f_n) consists of characteristic functions of intervals of width 2^{-k} that sweep across the interval $[0,1]$. We have $f_n \to 0$ in $L^p([0,1])$, for $1 \le p < \infty$, but $f_n(x)$ does not converge for any $x \in [0,1]$. The subsequence (f_{2^k}) converges pointwise-a.e. to zero.

As we have seen, it is often useful to approximate an arbitrary element in some space as the limit of a sequence of elements with special properties. Every L^p-function may be approximated by simple functions.

Theorem 12.48 Suppose that $(X, \mathcal{A}, \mu)$ is a measure space and $1 \leq p \leq \infty$. If $f \in L^p(X)$, then there is a sequence (φ_n) of simple functions $\varphi_n : X \to \mathbb{C}$ such that

$$\lim_{n\to\infty} \|f - \varphi_n\|_p = 0.$$

Proof. It is sufficient to prove the result for nonnegative functions, since we may approximate a general function in L^p by approximating its positive and negative parts. We consider only the case $1 \leq p < \infty$ for simplicity. If $f \geq 0$, then from Theorem 12.26 there is a monotone increasing sequence of nonnegative simple functions (φ_n) that converges pointwise to f. The sequence (g_n) defined by

$$g_n = f^p - (f - \varphi_n)^p$$

is a monotone increasing sequence of nonnegative functions. The monotone convergence theorem implies that

$$\int_X g_n \, d\mu \to \int_X f^p \, d\mu$$

as $n \to \infty$, from which it follows that $\varphi_n \to f$ in L^p. □

As an application of this theorem, we prove that $L^p(\mathbb{R}^n)$ is separable for $p < \infty$.

Theorem 12.49 If $1 \leq p < \infty$, then $L^p(\mathbb{R}^n)$ is a separable metric space.

Proof. We have to show that $L^p(\mathbb{R}^n)$ contains a countable dense subset. The set S of simple functions whose values are complex numbers with rational real and imaginary parts is dense in the space of simple functions. Hence, Theorem 12.48 implies that S is dense in $L^p(\mathbb{R}^n)$. The set S is not countable because there are far too many measurable sets, but we can approximate every simple function by a simple function of the form (12.3) in which each set A_i is chosen from a suitable countable collection $\mathcal{F}$ of measurable sets. For example, we can use the collection of cubes of the form $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, where $a_1, b_1, a_2, b_2, \ldots, a_n, b_n \in \mathbb{Q}$. We omit a detailed proof. The result then follows. □

This theorem also applies to $L^p(\Omega)$, with $1 \leq p < \infty$, where Ω is an arbitrary measurable subset of $\mathbb{R}^n$. The space $L^\infty(\mathbb{R}^n)$ is not separable (see Exercise 12.13).

For $p < \infty$, we can approximate functions in $L^p(\mathbb{R}^n)$ by continuous functions with compact support.

Theorem 12.50 The space $C_c(\mathbb{R}^n)$ of continuous functions with compact support is dense in $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$.

Proof. If $f \in L^p(\mathbb{R}^n)$, then the Lebesgue dominated convergence theorem implies that the sequence (f_n) of compactly supported functions,

$$f_n(x) = \begin{cases} f(x) & \text{if } |x| \leq n, \\ 0 & \text{if } |x| > n, \end{cases}$$

converges to f in L^p as $n \to \infty$. Each f_n may be approximated by simple functions, which are a finite linear combination of characteristic functions. It therefore suffices to prove that the characteristic function χ_A of a bounded, measurable set A in $\mathbb{R}^n$ may be approximated by continuous functions with compact support. From Theorem 12.10, for every $\epsilon > 0$ there is an open set G and a compact set $K \subset G$ such that $K \subset A \subset G$ and $\lambda(G \setminus K) < \epsilon^p$. By Urysohn's lemma (see Exercise 1.16), there is a continuous, real-valued function f such that $0 \le f(x) \le 1$, $f = 1$ on K, and $f = 0$ on G^c. Then

$$\|f - \chi_A\|_p = \left(\int_{G\setminus K} |f - \chi_A|^p \, dx\right)^{1/p} \le \lambda(G \setminus F)^{1/p} < \epsilon.$$

□

We can use an approximate identity to smooth out C_c-approximations, thus obtaining C_c^∞-approximations.

Theorem 12.51 If $1 \le p < \infty$, then $C_c^\infty(\mathbb{R}^n)$ is a dense subspace of $L^p(\mathbb{R}^n)$.

Proof. For $\epsilon > 0$, let $\varphi_\epsilon \in C_c^\infty(\mathbb{R}^n)$ be an approximate identity. If $f \in C_c(\mathbb{R}^n)$, we define $f_\epsilon = \varphi_\epsilon * f$. Then $f_\epsilon \in C_c^\infty(\mathbb{R}^n)$ for every $\epsilon > 0$. Moreover, $f_\epsilon \to f$ uniformly, and hence in L^p, as $\epsilon \to 0^+$. Since $C_c(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$, the result follows. □

12.7 The basic inequalities

It has been said that analysis is the art of estimating. In this section, we give the basic inequalities of L^p theory.

Many inequalities are based on convexity arguments. We first prove Jensen's inequality, which states that the mean of the values of a convex function is greater than or equal to the value of the convex function at the mean. We recall from Definition 8.47 that a function $\varphi : C \to \mathbb{R}$ on a convex set C is convex if

$$\varphi\left(tx + (1-t)y\right) \le t\varphi(x) + (1-t)\varphi(y)$$

for every $x, y \in C$ and $t \in [0, 1]$. We define the *mean* of an integrable function f on a finite measure space (X, μ) by

$$\langle f \rangle_\mu = \frac{1}{\mu(X)} \int_X f \, d\mu. \tag{12.8}$$

Theorem 12.52 (Jensen) Let (X, μ) be a finite measure space. If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $f : X \to \mathbb{R}$ belongs to $L^1(X, \mu)$, then

$$\varphi\left(\langle f \rangle_\mu\right) \le \langle \varphi \circ f \rangle_\mu. \tag{12.9}$$

Proof. From Exercise 12.9, there is a constant $c \in \mathbb{R}$ such that

$$\varphi(y) \geq \varphi(\langle f \rangle_\mu) + c(y - \langle f \rangle_\mu) \qquad \text{for every } y \in \mathbb{R}. \tag{12.10}$$

Setting $y = f(x)$ in this inequality and integrating the result over X, we obtain that

$$\begin{aligned} \int_X \varphi \circ f \, d\mu &\geq \varphi(\langle f \rangle_\mu) \int_X d\mu + c \left(\int_X f \, d\mu - \langle f \rangle_\mu \int_X d\mu \right) \\ &= \varphi\left(\langle f \rangle_\mu\right) \mu(X). \end{aligned}$$

Dividing this equation by $\mu(X) < \infty$, we obtain (12.9). □

The right-hand side of (12.9) may be infinite. Although $\varphi \circ f$ need not be in L^1, its negative part is integrable from (12.10), so its integral is well defined.

Example 12.53 Suppose that $\{x_1, x_2, \ldots, x_n\}$ is a finite subset of $\mathbb{R}$, and μ is a discrete probability measure on $\mathbb{R}$ with $\mu(\{x_i\}) = \lambda_i$, where $0 \leq \lambda_i \leq 1$ and

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1.$$

Then Jensen's inequality, with $f(x) = x$, implies that for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$ we have

$$\varphi\left(\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n\right) \leq \lambda_1 \varphi(x_1) + \lambda_2 \varphi(x_2) + \ldots + \lambda_n \varphi(x_n).$$

Hölder's inequality is one of the most important inequalities for proving estimates in L^p-spaces. We say that two numbers $1 \leq p, p' \leq \infty$ are *Hölder conjugates* or *conjugate exponents* if they satisfy

$$\frac{1}{p} + \frac{1}{p'} = 1, \tag{12.11}$$

with the convention that $1/\infty = 0$. For example, $p = 1$ and $p' = \infty$ are Hölder conjugates, and $p = 2$ is conjugate to itself. Hölder's inequality applies to a pair of functions, one in L^p and one in $L^{p'}$.

Theorem 12.54 (Hölder) Let $1 \leq p, p' \leq \infty$ satisfy (12.11). If $f \in L^p(X, \mu)$ and $g \in L^{p'}(X, \mu)$, then $fg \in L^1(X, \mu)$ and

$$\left| \int fg \, d\mu \right| \leq \|f\|_p \|g\|_{p'}. \tag{12.12}$$

Proof. For the conjugate pair of exponents $(p, p') = (1, \infty)$, Hölder's inequality is the obvious inequality

$$\left| \int_X fg \, d\mu \right| \leq \|g\|_\infty \int_X |f| \, d\mu.$$

We therefore assume that $1 < p, p' < \infty$.

The set $Y = \{x \in X \mid g(x) \neq 0\}$ is measurable, and

$$\left| \int_X f(x) g(x) \, d\mu(x) \right| = \left| \int_Y f(x) g(x) \, d\mu(x) \right| \leq \int_Y |f(x)| \, |g(x)| \, d\mu(x).$$

We also have

$$\|f\|_p^p = \int_X |f|^p \, d\mu \geq \int_Y |f|^p \, d\mu.$$

Therefore, replacing X by Y if necessary, it is sufficient to prove (12.12) under the assumption that $f(x) \geq 0$ and $g(x) > 0$ for every $x \in X$.

We define a new measure ν on X by

$$\nu(A) = \int_A g^{p'} \, d\mu.$$

The function $\varphi : \mathbb{R} \to \mathbb{R}$ defined by $\varphi(x) = |x|^p$ is convex for $p \geq 1$. An application of Jensen's inequality (12.9) to the function $h : X \to \mathbb{R}$ defined by

$$h = \frac{f}{g^{p'/p}}$$

implies that

$$\varphi\left(\langle h \rangle_\nu\right) \leq \langle \varphi \circ h \rangle_\nu .$$

Writing out this equation explicitly, we obtain

$$\left| \frac{1}{\nu(X)} \int_X \frac{f}{g^{p'/p}} \, d\nu \right|^p \leq \frac{1}{\nu(X)} \int_X \frac{f^p}{g^{p'}} \, d\nu.$$

Rewriting the integrals with respect to ν as integrals with respect to μ, and using the assumption that p and p' are dual exponents, we obtain that

$$\left| \frac{\int_X fg \, d\mu}{\int_X g^{p'} \, d\mu} \right|^p \leq \frac{1}{\int_X g^{p'} \, d\mu} \int_X f^p \, d\mu.$$

Taking the pth root of this equation, rearranging the result, and using the fact that p and p' are conjugate exponents again, we get (12.12). Since the right-hand side of this inequality is finite, it follows that $h \in L^1(X, \nu)$ and $fg \in L^1(X, \mu)$. $\square$

In the special case when $p = p' = 2$, Hölder's inequality is the Cauchy-Schwarz inequality for L^2-spaces. As an application of Hölder's inequality, we prove a result about the inclusion of L^p-spaces.

Proposition 12.55 Suppose that (X, μ) is a finite measure space, meaning that $\mu(X) < \infty$, and $1 \leq q \leq p \leq \infty$. Then

$$L^1(X, \mu) \supset L^q(X, \mu) \supset L^p(X, \mu) \supset L^\infty(X, \mu).$$

Proof. We define $1 \leq r \leq \infty$ by

$$\frac{1}{p} + \frac{1}{r} = \frac{1}{q}.$$

Then p/q and r/q are Hölder conjugates. Moreover, if $f \in L^p$, then $|f|^{p/q} \in L^q$ and $1 \in L^{r/q}$, since $\mu(X) < \infty$. Hölder's inequality therefore implies that

$$\begin{aligned}
\|f\|_q^q &= \int_X |f|^q \cdot 1 \, d\mu \\
&\leq \left(\int_X (|f|^q)^{p/q} \, d\mu \right)^{q/p} \left(\int_X d\mu \right)^{q/r} \\
&= (\mu(X))^{p/(p-q)} \, \|f\|_p^q.
\end{aligned}$$

Hence, $\|f\|_q$ is finite for every $f \in L^p$, which proves the claimed inclusion. □

For these inclusions to hold, it is crucial that $\mu(X) < \infty$, as illustrated in Exercise 12.14. Minkowski's inequality is the triangle inequality for the L^p-norm.

Theorem 12.56 (Minkowski) If $1 \leq p \leq \infty$, and $f, g \in L^p(X, \mu)$, then $f + g \in L^p(X, \mu)$, and

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \tag{12.13}$$

Proof. We have

$$\begin{aligned}
|f + g|^p &\leq (|f| + |g|)^p \\
&\leq 2^p \max\left(|f|^p, |g|^p\right) \\
&\leq 2^p \left(|f|^p + |g|^p\right).
\end{aligned}$$

Hence, $f + g \in L^p$ if $f, g \in L^p$.

If $f + g \in L^p$, then $|f + g|^{p-1} \in L^{p'}$, where p' is the Hölder conjugate of p, and

$$\left\| |f + g|^{p-1} \right\|_{p'} = \|f + g\|_p^{p-1}.$$

Hence, using Hölder's inequality followed by this result, we find that

$$\begin{aligned}
\|f + g\|_p^p &= \int |f + g| \, |f + g|^{p-1} \, d\mu \\
&\leq \int |f| \, |f + g|^{p-1} \, d\mu + \int |g| \, |f + g|^{p-1} \, d\mu \\
&\leq \|f\|_p \left\| |f + g|^{p-1} \right\|_{p'} + \|g\|_p \left\| |f + g|^{p-1} \right\|_{p'} \\
&\leq \left(\|f\|_p + \|g\|_p \right) \|f + g\|_p^{p-1}.
\end{aligned}$$

If $\|f + g\|_p \neq 0$, then division of this inequality by $\|f + g\|_p^{p-1}$ gives (12.13). If $\|f + g\|_p = 0$, then the result is trivial. □

For $0 \leq p \leq 1$, nonnegative functions satisfy the reverse triangle inequality

$$\|f+g\|_p \geq \|f\|_p + \|g\|_p,$$

which explains why L^p is only a normed linear space for $p \geq 1$.

Tchebyshev's inequality is an elementary inequality that is especially useful in probability theory.

Theorem 12.57 (Tchebyshev) Suppose that $f \in L^p(X)$, where $0 < p < \infty$. For every $\epsilon > 0$, we have

$$\mu\left(\{x \in X \mid |f(x)| > \epsilon\}\right) \leq \frac{1}{\epsilon^p}\|f\|_p^p.$$

Proof. Define $S_\epsilon \in \mathcal{A}$ by

$$S_\epsilon = \{x \in X \mid |f(x)| > \epsilon\}.$$

Then

$$\|f\|_p^p = \int_X |f|^p \, d\mu \geq \int_{S_\epsilon} |f|^p \, d\mu \geq \epsilon^p \mu\left(S_\epsilon\right),$$

which is what we had to prove. □

The following inequality for the L^p-norm of convolutions is called *Young's inequality.* This inequality shows that convolution is a continuous operation when defined on an appropriate choice of Banach spaces.

Theorem 12.58 (Young) Suppose that $1 \leq p, q, r \leq \infty$ satisfy

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}. \tag{12.14}$$

If $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$, then $f * g \in L^r(\mathbb{R}^n)$, and

$$\|f * g\|_r \leq \|f\|_p \, \|g\|_q.$$

Proof. We leave it to the reader to check that it is sufficient to prove the result for nonnegative functions f, g such that $\|f\|_p = \|g\|_q = 1$.

We first consider the special case $p = q = r = 1$. Using Fubini's theorem to exchange the order of integration, we have

$$\begin{aligned}
\|f * g\|_1 &= \int \left[\int f(y) g(x-y) \, dy\right] dx \\
&= \left[\int f(y) \, dy\right] \left[\int g(z) \, dz\right] \\
&= \|f\|_1 \|g\|_1, && (12.15)
\end{aligned}$$

which proves the result in this case.

For general values of p, q, we observe from (12.14) and the definition of the Hölder conjugate that

$$\frac{1}{p'} + \frac{1}{q'} + \frac{1}{r} = 1.$$

An application of the generalized Hölder inequality in Exercise 12.12 therefore implies that

$$\begin{aligned} f * g(x) &= \int \left[f(y)^{p/r} g(x-y)^{q/r}\right] \left[f(y)^{1-p/r}\right] \left[g(x-y)^{1-q/r}\right] dy \\ &\leq \left[\int f(y)^p g(x-y)^q \, dy\right]^{1/r} \left[\int f(y)^{(1-p/r)q'} \, dy\right]^{1/q'} \\ &\qquad \left[\int g(x-y)^{(1-q/r)p'} \, dy\right]^{1/p'}. \end{aligned}$$

Since $(1 - p/r)q' = p$, $(1 - q/r)p' = q$, and $\|f\|_p = \|g\|_q = 1$, it follows that

$$f * g(x)^r \leq \int f(y)^p g(x-y)^q \, dy,$$

meaning that $(f * g)^r \leq f^p * g^q$. The use of this inequality and (12.15) then implies that

$$\|f * g\|_r^r = \|(f * g)^r\|_1 \leq \|f^p * g^q\| = \|f^p\|_1 \|g^q\|_1 = \|f\|_p^p \|g\|_q^q = 1,$$

which proves the theorem. □

Two common special cases of this result are:

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1, \qquad \|f * g\|_\infty \leq \|f\|_2 \|g\|_2.$$

12.8 The dual space of L^p

In Section 5.6, we gave the general definition of the dual space of a Banach space. In this section, we describe the dual space $L^p(X)^*$ of bounded, linear functionals on $L^p(X)$, where X is a measure space equipped with a measure μ.

Suppose that $1 \leq p \leq \infty$ and $g \in L^{p'}(X)$, where p' is the Hölder conjugate of p. We define $\varphi_g : L^p(X) \to \mathbb{C}$ by

$$\varphi_g(f) = \int_X f g \, d\mu \qquad \text{for every } f \in L^p(X). \tag{12.16}$$

Hölder's inequality implies that φ_g is a bounded linear functional on L^p, with

$$\|\varphi_g\|_{(L^p)^*} \leq \|g\|_{L^q}.$$

Here, $\|\cdot\|_{(L^p)^*}$ is the norm of a bounded linear function defined in (5.23). The next theorem states that if $1 \leq p < \infty$, then all linear functionals on L^p are of the form (12.16).

Theorem 12.59 If $1 < p < \infty$, then every $\varphi \in L^p(X)^*$ is of the form

$$\varphi(f) = \int_X fg\,d\mu$$

for some $g \in L^{p'}(X)$, where $1/p + 1/p' = 1$. If μ is σ-finite the same conclusion holds when $p = 1$ and $p' = \infty$. Moreover,

$$\|\varphi\|_{(L^p)^*} = \|g\|_{L^{p'}}.$$

We will not give the proof. According to Theorem 12.59, we may identify the dual of L^p with $L^{p'}$. When $p = p' = 2$, we recover the result of the Riesz representation theorem that the dual of the Hilbert space L^2 may be identified with itself. The dual of L^1 is L^∞, but the dual of L^∞ is strictly larger than L^1 (except in trivial cases, such as when X is a finite set). The full description of $(L^\infty)^*$ is complicated and rarely useful, so we will not give it here. If $1 < p < \infty$, then $(L^p)^{**} = L^p$ and L^p is reflexive, but L^1 and L^∞ are not reflexive.

The continuous linear functionals define the weak topology. From Definition 5.59, Definition 5.60, and Theorem 12.59, we have the following definition of weak L^p-convergence.

Definition 12.60 Suppose that $1 \leq p < \infty$. A sequence (f_n) *converges weakly* to f in L^p, written $f_n \rightharpoonup f$, if

$$\lim_{n\to\infty} \int f_n g\,d\mu = \int fg\,d\mu \qquad \text{for every } g \in L^{p'}, \tag{12.17}$$

where p' is the Hölder conjugate of p. When $p = \infty$ and $p' = 1$, the condition in (12.17) corresponds to weak-$*$ convergence $f_n \overset{*}{\rightharpoonup} f$ in L^∞.

As in the case of Hilbert spaces, discussed in Section 8.6, weak L^p-convergence does not imply strong L^p-convergence, meaning convergence in the L^p-norm. The following example illustrates three typical ways in which a weakly convergent sequence of functions can fail to be strongly convergent.

Example 12.61 Let $g \in L^p(\mathbb{R})$ be a fixed nonzero function, where $1 < p < \infty$. For each of the following three sequences, we have $f_n \rightharpoonup 0$ weakly as $n \to \infty$, but not $f_n \to 0$ strongly, in $L^p(\mathbb{R})$.

(a) $f_n(x) = g(x)\sin nx$ (oscillation);
(b) $f_n(x) = n^{1/p}g(nx)$ (concentration);
(c) $f_n(x) = g(x - n)$ (escape to infinity).

The Banach-Alaoglu theorem, in Theorem 5.61, leads to the following weak compactness result for L^p.

Theorem 12.62 Suppose that (f_n) is a bounded sequence in $L^p(X)$, meaning that there is a constant M such that $\|f_n\| \le M$ for every $n \in \mathbb{N}$. If $1 < p < \infty$, then there is a subsequence (f_{n_k}) and a function $f \in L^p(X)$ with $\|f\| \le M$ such that $f_{n_k} \rightharpoonup f$ as $k \to \infty$ weakly in $L^p(X)$.

12.9 Sobolev spaces

Many problems in applied analysis involve differentiable functions. *Sobolev spaces* are Banach spaces of functions whose weak derivatives belong to L^p spaces. They provide the simplest and most useful setting for the application of functional analytic methods to the study of differential equations. We have already discussed several special cases of Sobolev spaces in the chapters on Fourier series and unbounded linear operators. Here, we give more general definitions of Sobolev spaces, and describe some of their main properties. We use the multi-index notation introduced in Section 11.1, and consider real-valued functions for simplicity.

We will define Sobolev spaces of functions whose domain is an open subset Ω of $\mathbb{R}^n$, equipped with n-dimensional Lebesgue measure. In particular, we could have $\Omega = \mathbb{R}^n$. As usual, $L^p(\Omega)$ denotes the space of Lebesgue measurable functions $f : \Omega \to \mathbb{R}$ whose pth powers are integrable. We also introduce the *local L^p spaces*, denoted by $L^p_{\mathrm{loc}}(\Omega)$. A function f belongs to $L^p_{\mathrm{loc}}(\Omega)$ if it is measurable and

$$\int_K |f|^p \, dx < \infty$$

for every compact subset K of Ω. For example, $1/x$ belongs to $L^1_{\mathrm{loc}}((0,1))$, but not to $L^1((0,1))$ or $L^1_{\mathrm{loc}}(\mathbb{R})$. For every $1 \le p \le \infty$, we have the inclusions

$$L^1_{\mathrm{loc}}(\Omega) \supset L^p_{\mathrm{loc}}(\Omega) \supset L^p(\Omega).$$

Thus, $L^1_{\mathrm{loc}}(\Omega)$ is the largest space of integrable functions. We adopt the following definition of a test function for the purposes of this chapter.

Definition 12.63 A *test function* $\varphi : \Omega \to \mathbb{R}$ on an open subset Ω of $\mathbb{R}^n$ is a function with continuous partial derivatives of all orders whose support is a compact subset of Ω. We denote the set of test functions on Ω by $C_c^\infty(\Omega)$.

Definition 12.64 If $f, g_\alpha \in L^1_{\mathrm{loc}}(\Omega)$ are such that

$$\int_\Omega g_\alpha \varphi \, dx = (-1)^{|\alpha|} \int_\Omega f \partial^\alpha \varphi \, dx \qquad \text{for all } \varphi \in C_c^\infty(\Omega),$$

then we say that $g_\alpha = \partial^\alpha f$ is the αth *weak partial derivative* of f.

The weak derivative is only defined pointwise up to a set of measure zero.

Example 12.65 The function

$$f(x) = \frac{1}{|x|^a}$$

belongs to $L^1_{\mathrm{loc}}(\mathbb{R}^n)$ if and only if $a < n$. The weak derivative of f with respect to x_i is given by

$$g_i(x) = -a\frac{x_i}{|x|}\frac{1}{|x|^{a+1}}$$

provided that g_i is locally integrable which is the case when $a < n-1$. For example, in one space dimension, if $f' \in L^p_{\mathrm{loc}}(\mathbb{R})$ for some $p > 1$, then $a < 0$ and f is continuous. In fact, the Sobolev embedding theorem below implies that any function on $\mathbb{R}$ whose weak derivative belongs to $L^p_{\mathrm{loc}}(\mathbb{R})$ for some $p > 1$ is continuous. In higher space dimensions, a function may be weakly differentiable but discontinuous. The strength of an allowable singularity in a weakly differentiable function increases with the number of space dimensions n.

Definition 12.66 Let k be a positive integer, $1 \leq p \leq \infty$, and Ω an open subset of $\mathbb{R}^n$. The *Sobolev space* $W^{k,p}(\Omega)$ consists of all functions $f : \Omega \to \mathbb{R}$ such that $\partial^\alpha f \in L^p(\Omega)$ for all weak partial derivatives of order $0 \leq |\alpha| \leq k$. We define a norm on $W^{k,p}(\Omega)$ by

$$\|f\|_{W^{k,p}(\Omega)} = \left(\sum_{0\leq|\alpha|\leq k}\int_\Omega |\partial^\alpha f|^p \, dx\right)^{1/p}$$

when $1 \leq p < \infty$, and by

$$\|f\|_{W^{k,\infty}(\Omega)} = \max_{0\leq|\alpha|\leq k}\left\{\sup_{x\in\Omega}|\partial^\alpha f(x)|\right\}$$

when $p = \infty$. Here, the supremum is to be interpreted as an essential supremum. For $p = 2$, corresponding to the case of square integrable functions, we write $W^{k,2}(\Omega) = H^k(\Omega)$, and define an inner product on $H^k(\Omega)$ by

$$\langle f, g\rangle_{H^k(\Omega)} = \sum_{0\leq|\alpha|\leq k}\int_\Omega \partial^\alpha f \partial^\alpha g \, dx.$$

The space $W^{k,p}(\Omega)$ is a Banach space and $H^k(\Omega)$ is a Hilbert space.

Next, we define a Sobolev space of functions that "vanish on the boundary of Ω."

Definition 12.67 The closure of $C_c^\infty(\Omega)$ in $W^{k,p}(\Omega)$ is denoted by

$$W_0^{k,p}(\Omega) = \overline{C_c^\infty(\Omega)}.$$

We also define

$$H_0^k(\Omega) = W_0^{k,2}(\Omega).$$

Informally, we can think of $W_0^{k,p}(\Omega)$ as the $W^{k,p}(\Omega)$-functions whose derivatives of order less than or equal to $k-1$ vanish on the boundary $\partial\Omega$ of Ω. Compactly supported functions are dense in $W^{k,p}(\mathbb{R}^n)$, so that $W_0^{k,p}(\mathbb{R}^n) = W^{k,p}(\mathbb{R}^n)$.

The definition of the boundary values of Sobolev functions which do not vanish on the boundary is non-trivial. The boundary of a smooth set has measure zero, but Sobolev functions are not necessarily continuous and they are defined pointwise only up to sets of measure zero. The trace theorem, in Theorem 12.76 below, gives a way to assign boundary values to Sobolev functions.

Sobolev spaces of negative orders may be defined by duality.

Definition 12.68 Let k be a positive integer, $1 \le p < \infty$, and p' the Hölder conjugate of p. The Sobolev space $W^{-k,p}(\Omega)$ is the dual space of $W_0^{k,p'}(\Omega)$. That is, $f \in W^{-k,p}(\Omega)$ is a continuous linear map

$$f : W_0^{k,p'}(\Omega) \to \mathbb{R}, \qquad f : u \mapsto \langle f, u\rangle.$$

We define a norm on $W^{-k,p}(\Omega)$ by

$$\|f\|_{W^{-k,p}} = \sup_{\substack{u \in W_0^{k,p'} \\ u \neq 0}} \frac{\langle f, u\rangle}{\|u\|}.$$

In particular, $H^{-k}(\Omega)$ is the dual space of $H_0^k(\Omega)$. Elements of $W^{-k,p}(\Omega)$ are distributions whose action on test functions extends continuously to an action on functions in $W_0^{k,p'}(\Omega)$. The dual space of $W^{k,p'}(\Omega)$ is not a space of distributions, because the action of a continuous linear functional on functions in $W^{k,p'}(\Omega)$ depends on the values of the $W^{k,p'}(\Omega)$-functions on the boundary $\partial\Omega$, and therefore it is not determined by its action on compactly supported test functions. It is possible to show that any distribution $f \in W^{-k,p}(\Omega)$ may be written (nonuniquely) as

$$f = \sum_{|\alpha| \le k} \partial^\alpha g_\alpha$$

where $g_\alpha \in L^p(\Omega)$. The action of this distribution f on a $W_0^{k,p'}$-function u is given by

$$\langle f, u\rangle = \sum_{|\alpha| \le k} (-1)^{|\alpha|} \int_\Omega g_\alpha \partial^\alpha u \, dx.$$

More generally, it is possible to define Sobolev spaces $W^{s,p}$ of fractional order for any $s \in \mathbb{R}$ and $p \in [1, \infty]$. These spaces arise naturally in connection with the trace theorem below, but we will not describe them here. For $p = 2$, the Hilbert spaces $H^s(\mathbb{R}^n)$ may be defined by use of the Fourier transform (see Definition 11.38).

12.10 Properties of Sobolev spaces

In this section, we summarize the main properties of Sobolev spaces without proof. These properties include the approximation of Sobolev functions by smooth functions (density theorems), the integrability or continuity properties of Sobolev functions (embedding theorems), compactness conditions (the Rellich-Kondrachov theorem), and the definition of boundary values of Sobolev functions (trace theorems). Depending on the context, there are many different regularity conditions that the domain Ω on which the Sobolev functions are defined must satisfy, and the differences between them are sometimes quite subtle. We will say that a domain is *regular* if it satisfies an appropriate regularity condition, without stating the precise condition that is required. Any domain bounded by a smooth hypersurface (meaning that the boundary is locally the zero set of a smooth function with nonzero derivative) satisfies the regularity conditions for all the results stated in this section. Domains with outward pointing cusps or needle-shaped protrusions are typical examples of domains with insufficient regularity, and some of the properties stated below do not hold on such domains.

We use $C(\overline{\Omega})$ to denote the space of uniformly continuous functions on Ω, and $C_0(\mathbb{R}^n)$ to denote the space of continuous functions on $\mathbb{R}^n$ that tend to zero as $x \to \infty$. This space is the closure of $C_c^\infty(\mathbb{R}^n)$ in $L^\infty(\mathbb{R}^n)$. The space $C^k(\overline{\Omega})$ consists of functions whose partial derivatives of order less than or equal to k are uniformly continuous in Ω, and $C^\infty(\overline{\Omega})$ consists of functions with uniformly continuous derivatives of all orders in Ω. If a function is uniformly continuous in the open set Ω, then it has a unique continuous extension to the closure $\overline{\Omega}$.

In the theorems stated below, we consider two types of domains: $\Omega = \mathbb{R}^n$; and Ω a regular, bounded, open subset of $\mathbb{R}^n$ with boundary $\partial\Omega$. It is frequently possible to consider more general domains, but this complicates the statements of some of the theorems. The order k is a positive integer and $1 \leq p \leq \infty$, unless stated otherwise.

Theorem 12.69 (Density) The space $C_c^\infty(\mathbb{R}^n)$ is dense in $W^{k,p}(\mathbb{R}^n)$. If Ω is an open subset of $\mathbb{R}^n$, then $C_c^\infty(\Omega)$ is dense in $W_0^{k,p}(\Omega)$, and if Ω is regular, then $C^\infty(\overline{\Omega})$ is dense in $W^{k,p}(\Omega)$.

Meyers and Serrin proved for general domains that $C^\infty(\Omega) \cap W^{k,p}(\Omega)$ is dense in $W^{k,p}(\Omega)$, but functions in $C^\infty(\Omega)$ need not be smooth up to the boundary.

Among the most important properties of Sobolev spaces are the embedding theorems, which provide information about the integrability or continuity of a function given information about the integrability of its derivatives. To motivate the embedding theorems, we first consider functions $u : \mathbb{R}^n \to \mathbb{R}$, and ask when it is possible to have an estimate of the form

$$\|u\|_{L^q} \leq C\|\nabla u\|_{L^p} \tag{12.18}$$

for a constant C that is independent of u. Here,

$$\nabla u = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \dots, \frac{\partial u}{\partial x_n}\right)$$

is the gradient, or derivative, of u and

$$\|\nabla u\|_{L^p} = \left(\left\|\frac{\partial u}{\partial x_1}\right\|_{L^p}^p + \left\|\frac{\partial u}{\partial x_2}\right\|_{L^p}^p + \dots + \left\|\frac{\partial u}{\partial x_n}\right\|_{L^p}^p\right)^{1/p}.$$

We also use the notation Du for ∇u.

For $t > 0$, we define the rescaled function

$$u_t(x) = u(tx).$$

A simple calculation shows that

$$\|u_t\|_{L^q} = t^{-n/q}\|u\|_{L^q}, \qquad \|\nabla u_t\|_{L^p} = t^{1-n/p}\|\nabla u\|_{L^p}. \tag{12.19}$$

These norms must scale according to the same exponent if the estimate in (12.18) is to hold. This occurs if and only if $p < n$ and $q = p^*$, where

$$p^* = \frac{np}{n-p}.$$

This equation may also be written as

$$\frac{1}{p^*} = \frac{1}{p} - \frac{1}{n}. \tag{12.20}$$

We call p^* the *Sobolev conjugate* of p. The inequality in (12.18) does in fact hold for every $u \in C_c^\infty(\mathbb{R}^n)$ when $q = p^*$, and it follows by a density argument that every function in $W^{1,p}(\mathbb{R}^n)$ belongs to $L^{p^*}(\mathbb{R}^n)$ when $p < n$.

The inclusion $W^{1,p}(\mathbb{R}^n) \subset L^{p^*}(\mathbb{R}^n)$ is equivalent to the existence of an embedding

$$J : W^{1,p}(\mathbb{R}^n) \to L^{p^*}(\mathbb{R}^n),$$

where $Jf = f$. The estimate (12.18) implies the continuity of J.

If $\Omega \subset \mathbb{R}^n$ is a bounded domain, then we have $W^{1,p}(\Omega) \subset L^{p^*}(\Omega)$ and $L^{p^*}(\Omega) \subset L^q(\Omega)$ for $1 \leq q \leq p^*$. Thus, there is a continuous embedding $J : W^{1,p}(\Omega) \to L^q(\Omega)$. Summarizing these results, we get the following theorem.

Theorem 12.70 (Embedding) Suppose that Ω is a regular, bounded, open set in $\mathbb{R}^n$, $p < n$, and p^* is the Sobolev conjugate of p, defined in (12.20).

(a) If $u \in W^{1,p}(\mathbb{R}^n)$, then $u \in L^{p^*}(\mathbb{R}^n)$. There is a constant $C = C(p, n)$ such that

$$\|u\|_{L^{p^*}(\mathbb{R}^n)} \leq C\|\nabla u\|_{L^p(\mathbb{R}^n)}.$$

(b) If $u \in W^{1,p}(\Omega)$ and $1 \leq q \leq p^*$, then $u \in L^q(\Omega)$. There is a constant $C = C(p, q, \Omega)$ such that

$$\|u\|_{L^q(\Omega)} \leq C\|u\|_{W^{1,p}(\Omega)}.$$

To prove this theorem, one uses the Hölder inequality to show that the estimate holds for test functions. The result follows for arbitrary Sobolev functions by the density of test functions in Sobolev spaces (see Adams [1] for complete proofs).

The above embedding theorem applies if $p < n$. If $p > n$, then functions in $W^{1,p}$ are continuous and one can estimate their uniform norm in terms of their $W^{1,p}$-norm.

Theorem 12.71 (Embedding) Suppose that $n < p < \infty$, and Ω is a regular bounded open subset of $\mathbb{R}^n$.

(a) If $u \in W^{1,p}(\mathbb{R}^n)$, then $u \in C_0(\mathbb{R}^n)$. There exists a constant $C = C(p, n)$ such that

$$\|u\|_{L^\infty(\mathbb{R}^n)} \leq C\|\nabla u\|_{L^p(\mathbb{R}^n)}.$$

(b) If $u \in W^{1,p}(\Omega)$ then $u \in C(\overline{\Omega})$ and there exists a constant $C = C(p, \Omega)$ such that

$$\|u\|_{L^\infty(\Omega)} \leq C\|u\|_{W^{1,p}(\Omega)}.$$

A more refined version of this embedding theorem states that the functions are *Hölder continuous.*

Definition 12.72 A function $u : \Omega \to \mathbb{R}$ is *Hölder continuous* in the open set Ω, with exponent $0 < r \leq 1$, if

$$\sup_{\substack{x,y\in\Omega \\ x\neq y}} \frac{|u(x) - u(y)|}{|x - y|^r} < \infty.$$

If u is Hölder continuous with exponent $r = 1$, then u is Lipschitz continuous. Any Hölder continuous function is continuous, but not conversely. The Banach space $C^{0,r}(\overline{\Omega})$ consists of all bounded Hölder continuous functions in Ω with the norm

$$\|u\|_{C^{0,r}(\overline{\Omega})} = \sup_{x\in\Omega} |u(x)| + \sup_{\substack{x,y\in\Omega \\ x\neq y}} \frac{|u(x) - u(y)|}{|x - y|^r}.$$

For each positive integer k and $0 < r \leq 1$, we define $C^{k,r}(\overline{\Omega})$ to be the space of functions that are k times continuously differentiable in Ω, with uniformly continuous derivatives whose kth-order derivatives are Hölder continuous with exponent r. This space is a Banach space with the norm

$$\|u\|_{C^{k,r}(\overline{\Omega})} = \max_{0\leq|\alpha|\leq k}\left\{\sup_{x\in\Omega}|\partial^\alpha u(x)|\right\} + \max_{|\alpha|=k}\sup_{\substack{x,y\in\Omega\\ x\neq y}}\frac{|\partial^\alpha u(x) - \partial^\alpha u(y)|}{|x-y|^r}.$$

Theorem 12.73 (Morrey) Suppose that $n < p < \infty$. Let

$$r = 1 - \frac{n}{p}.$$

(a) If $u \in W^{1,p}(\mathbb{R}^n)$, then $u \in C^{0,r}(\mathbb{R}^n)$ and there exists a constant $C = C(p,n)$ such that

$$\|u\|_{C^{0,r}(\mathbb{R}^n)} \leq C\|\nabla u\|_{L^p(\mathbb{R}^n)}.$$

(b) If $u \in W^{1,p}(\Omega)$, then $u \in C^{0,r}(\Omega)$ and there exists a constant $C = C(p,\Omega)$ such that

$$\|u\|_{C^{0,r}(\Omega)} \leq C\|u\|_{W^{1,p}(\Omega)}.$$

Spaces of continuous functions form an algebra with respect to the pointwise product, since the pointwise product of continuous functions is continuous, but the L^p-spaces do not form an algebra; for example, the product of two L^2-functions belongs to L^1, but not in general to L^2. The Sobolev spaces form an algebra when they consist of continuous functions.

Theorem 12.74 Suppose that Ω is an open subset of $\mathbb{R}^n$, including the possibility $\Omega = \mathbb{R}^n$. If $kp > n$, then $W^{k,p}(\Omega)$ is an algebra and there is a constant C such that

$$\|uv\|_{W^{k,p}} \leq C\|u\|_{W^{k,p}}\|v\|_{W^{k,p}} \qquad \text{for all } u, v \in W^{k,p}(\Omega).$$

It is a general principle that a set of functions whose derivatives are uniformly bounded is compact. The Sobolev-space version of this principle is the Rellich-Kondrachov theorem which states that $W^{k,p}(\Omega)$ is compactly embedded in $L^q(\Omega)$ for $q < p^*$. The boundedness of the domain Ω and the condition that q is strictly less than the Sobolev conjugate p^* of p are both essential for compactness. In the critical case, $q = p^*$, the embedding is continuous but not compact.

Theorem 12.75 (Rellich-Kondrachov) Let Ω be a regular bounded domain in $\mathbb{R}^n$.

(a) Suppose that $1 \leq p < n$ and $1 \leq q < p^*$. Then bounded sets in $W^{1,p}(\Omega)$ are precompact in $L^q(\Omega)$.

(b) Suppose that $p > n$. Then bounded sets in $W^{1,p}(\Omega)$ are precompact in $C(\overline{\Omega})$.

In particular, suppose that (u_k) is a sequence of functions in $W^{1,p}(\Omega)$ such that

$$\|u_k\|_{W^{1,p}} \leq C$$

for a constant C that is independent of k. If $p < n$ and $1 \leq q < p^*$, then there is a subsequence of (u_k) that converges strongly in $L^q(\Omega)$. If $p > n$, then there is a uniformly convergent subsequence.

If $p > n$ and $0 < r < 1 - n/p$, then the embedding of $W^{1,p}(\Omega)$ into $C^{0,r}(\overline{\Omega})$ is compact. General compactness theorems follow by repeated application of this result. For example, if $kp < n$ then $W^{k,p}(\Omega)$ is compactly embedded in $L^q(\Omega)$ for any $1 \leq q < np/(n - kp)$, while if $kp > n$ then $W^{k,p}(\Omega)$ is compactly embedded in $C(\overline{\Omega})$.

There is no sensible way to assign boundary values $u|_{\partial\Omega}$ to a general function $u \in L^p(\Omega)$. Functions in L^p are defined only pointwise-a.e., and the boundary $\partial\Omega$ of a regular domain has measure zero. The situation is different for Sobolev functions. If $u \in W^{k,p}(\Omega)$, then one can assign boundary values to the derivatives of u of order less than or equal to $k - 1/p$. It is not possible to define boundary values of kth order derivatives, however, since they are just L^p functions.

Theorem 12.76 (Trace) There is a surjective bounded linear operator

$$\gamma : W^{1,p}(\Omega) \to W^{1-1/p,p}(\partial\Omega)$$

such that

$$\gamma u = u|_{\partial\Omega} \qquad \text{if } u \in W^{1,p}(\Omega) \cap C(\overline{\Omega}).$$

There is a "loss of $1/p$ derivatives" in restricting a Sobolev function to the boundary. For example, the boundary values of a function in $H^1(\Omega)$ belong to $H^{1/2}(\partial\Omega)$. Conversely given an element of $H^{1/2}(\partial\Omega)$, there is a function in $H^1(\Omega)$ which takes those boundary values.

Our last result is the Poincaré inequality, which has many variants. The common theme is that, after removing nonzero constant functions, one can estimate the L^p-norm of a function in terms of the L^p-norm of its derivative.

Theorem 12.77 (Poincaré) Suppose that Ω is a bounded domain. Then there is a constant C such that

$$\|u\|_{L^p} \leq C\|\nabla u\|_{L^p}$$

for every $u \in W_0^{1,p}(\Omega)$.

More generally, this estimate holds if Ω is bounded in one direction. The Poincaré estimate is false for nonzero constant functions, so the assumption that

$u \in W_0^{1,p}$, rather than $u \in W^{1,p}$, is essential. A useful consequence of this estimate is that $\|\nabla u\|_{L^p}$ provides an equivalent norm on $W_0^{1,p}(\Omega)$. When $p = 2$, it follows that we can use

$$\langle u, v\rangle = \int_\Omega \nabla u(x) \cdot \nabla v(x)\, dx$$

as an inner product on $H_0^1(\Omega)$. Another Poincaré inequality is the following.

Theorem 12.78 (Poincaré) Suppose that Ω is a smooth connected bounded domain. There exists a constant C such that

$$\|u - \langle u\rangle\|_{L^p} \leq C\|\nabla u\|_{L^p} \tag{12.21}$$

for every $u \in W^{1,p}(\Omega)$, where $\langle u\rangle$ is the mean of u over Ω,

$$\langle u\rangle = \frac{1}{|\Omega|}\int_\Omega u(x)\, dx,$$

and $|\Omega|$ is the volume of Ω.

12.11 Laplace's equation

The Dirichlet problem for the Laplacian is

$$\begin{aligned} -\Delta u = f \qquad & x \in \Omega, \\ u(x) = 0 \qquad & x \in \partial\Omega. \end{aligned} \tag{12.22}$$

Here, $f : \Omega \to \mathbb{R}$ is a given function (or distribution) and Ω is a smooth bounded open set in $\mathbb{R}^n$. We assume homogeneous boundary conditions for simplicity; nonhomogeneous boundary conditions may be transferred to the PDE in the usual way.

To formulate any PDE problem in a precise way, we have to specify what function space solutions should belong to. We also have to specify how the derivatives are defined and in what sense the solution satisfies the boundary conditions and any other side conditions. There is often a great deal of choice in how this is done. A *classical solution* of (12.22) is a twice-continuously differentiable function u that satisfies the PDE pointwise, wheras a ***weak solution*** satisfies it in a distributional sense.

To motivate the definition of a weak solution, we suppose that u is a smooth solution. Let $\varphi \in C_c^\infty(\Omega)$ be any test function. Then multiplication of (12.22) by φ and an integration by parts imply that

$$\int_\Omega \nabla u(x) \cdot \nabla \varphi(x)\, dx = \int_\Omega f(x)\varphi(x)\, dx. \tag{12.23}$$

Conversely, if u is a smooth function that vanishes on $\partial\Omega$ and satisfies (12.23) for all test functions φ, then u is a classical solution of the original boundary value problem.

Let us require that the solution u and the test function φ belong to the same space. Then ∇u and $\nabla\varphi$ must both be square-integrable, so it is natural to look for solutions in the space $H_0^1(\Omega)$. Since $\varphi \in H_0^1(\Omega)$, we can make sense of the right-hand side of (12.23) provided that $f \in H^{-1}(\Omega)$. This leads to the following definition.

Definition 12.79 Given a distribution $f \in H^{-1}(\Omega)$, we say that u is a *weak solution* of (12.22) if $u \in H_0^1(\Omega)$ and

$$\int_\Omega \nabla u \cdot \nabla\varphi \, dx = \langle f, \varphi\rangle \qquad \text{for every } \varphi \in H_0^1(\Omega),$$

where $\langle\cdot,\cdot\rangle$ denotes the duality pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$.

If we define a quadratic functional $I : H_0^1(\Omega) \to \mathbb{R}$ by

$$I(u) = \frac{1}{2}\int_\Omega |\nabla u|^2 \, dx - \langle f, u\rangle,$$

then, as we will see in Section 13.9, a function u that minimizes I is a weak solution of (12.22). As a result of this connection, the present approach to the study of the Laplacian is often called the variational method. The boundary condition $u = 0$ on $\partial\Omega$ is replaced by the condition that $u \in H_0^1(\Omega)$. The precise sense in which weak solutions satisfy boundary conditions or initial conditions often requires careful attention. Definition 12.79 is not the most general definition of weak solutions. For example, we could consider distributional solutions of (12.22) when $f \notin H^{-1}(\Omega)$. The definition given is the natural one for the following existence theorem.

Theorem 12.80 There is a unique weak solution $u \in H_0^1(\Omega)$ of (12.22) for every $f \in H^{-1}(\Omega)$. There is a constant $C = C(\Omega)$ such that

$$\|u\|_{H^1} \le C\|f\|_{H^{-1}} \qquad \text{for all } f \in H^{-1}(\Omega).$$

Proof. By the Poincaré inequality, we can use

$$(u, v) = \int_\Omega \nabla u \cdot \nabla v \, dx$$

as an inner product on $H_0^1(\Omega)$. Since $f \in H^{-1}(\Omega) = H_0^1(\Omega)^*$, and $H_0^1(\Omega)$ is a Hilbert space, the Riesz representation theorem implies that there is a unique $u \in H_0^1(\Omega)$ such that

$$(u, \varphi) = \langle f, \varphi\rangle$$

for every $\varphi \in H_0^1(\Omega)$. This function u is the unique weak solution of (12.22). Moreover, we have $\|u\|_{H_0^1} = \|f\|_{H^{-1}}$. □

This theorem implies that

$$-\Delta : H_0^1(\Omega) \to H^{-1}(\Omega)$$

is a Hilbert space isomorphism; in fact, it is the isomorphism between $H_0^1(\Omega)$ and its dual space $H^{-1}(\Omega)$, in which the dual space is represented concretely as a space of distributions.

The proof of Theorem 12.80 gives a solution u of (12.22) that belongs to $H_0^1(\Omega)$. This is the best regularity one can hope for in the case of a general right hand side $f \in H^{-1}$. If, however, $f \in H^k$ is smooth, then elliptic regularity theory shows that the solution $u \in H^{k+2}$ and that

$$\|u\|_{H^{k+2}} \leq C\|f\|_{H^k}.$$

This gain of derivatives is typical of elliptic equations. One can estimate the L^2-norm of all second derivatives of u in terms of the L^2-norm of the single combination of second derivatives Δu. If $f \in H^k(\Omega)$ with $k > n/2$, then it follows from the Sobolev embedding theorem that $u \in H^{k+2}(\Omega) \subset C^2(\overline{\Omega})$, so u is a classical solution, and if $f \in C^\infty(\overline{\Omega})$, then $u \in C^\infty(\overline{\Omega})$.

If $f \in C(\overline{\Omega})$, then it is not necessarily true that $u \in C^2(\overline{\Omega})$. If, however, $f \in C^{k,r}(\overline{\Omega})$ is Hölder continuous, where $0 < r \leq 1$, then there is a unique Hölder continuous solution $u \in C^{k+2,r}(\overline{\Omega})$, and one can estimate the Hölder norms of the second derivatives of u in terms of the Hölder norm of f. Analogous existence, uniqueness, and regularity results hold in $L^p(\Omega)$ for $1 < p < \infty$, but not for $p = 1$ or $p = \infty$.

The idea in Theorem 12.80 of using the Riesz representation theorem to prove the existence and uniqueness of a weak solution applies to more general linear PDEs. Consider a linear equation that can be written in the abstract form

$$Au = f, \tag{12.24}$$

where $A : \mathcal{H} \to \mathcal{H}^*$ is a bounded linear operator from a Hilbert space $\mathcal{H}$ to its dual space $\mathcal{H}^*$, and $f \in \mathcal{H}^*$. In the case of Laplace's equation, we had $A = -\Delta$, $\mathcal{H} = H_0^1(\Omega)$, and $\mathcal{H}^* = H^{-1}(\Omega)$. Evaluation of (12.24) on a test function $v \in \mathcal{H}$, gives an equivalent weak formulation:

$$a(u, v) = \langle f, v \rangle \qquad \text{for all } v \in \mathcal{H},$$

where $a : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is defined by

$$a(u, v) = \langle Au, v \rangle.$$

This bilinear form a is called the *Dirichlet form* associated with A. In the case of Laplace's equation, we have

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v \, dx.$$

The Dirichlet form of a Sturm-Liouville ordinary differential operator is given in (10.38). If a is a symmetric, positive definite, sesquilinear form, and there exists a constant $\alpha > 0$ such that

$$a(u, u) \geq \alpha \|u\|^2 \qquad \text{for all } u \in \mathcal{H},$$

then the *energy norm*

$$\|u\|_A = \sqrt{a(u, u)}$$

is equivalent to the original norm on $\mathcal{H}$, and the Riesz representation theorem implies the existence and uniqueness of a weak solution u of (12.24) for every $f \in \mathcal{H}^*$.

If a is not symmetric, then it does not define an inner product on $\mathcal{H}$, and the Riesz representation cannot be used directly to establish the existence of a weak solution. Nevertheless, a similar result, called the *Lax-Milgram lemma*, still applies. The proof is outlined in Exercise 12.23.

Theorem 12.81 (Lax-Milgram) Suppose that $a : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is a sesquilinear form on a Hilbert space $\mathcal{H}$, and there are constants $\alpha > 0$, $\beta > 0$ such that

$$\alpha \|x\|^2 \leq |a(x, x)|, \qquad |a(x, y)| \leq \beta \|x\| \, \|y\|,$$

for all $x, y \in \mathcal{H}$. Then for every bounded linear functional $F : \mathcal{H} \to \mathbb{C}$ there is a unique element $x \in \mathcal{H}$ such that

$$a(x, y) = F(y) \qquad \text{for all } y \in \mathcal{H}.$$

Finally, we mention that (12.23) is a useful starting point for numerical methods of solving Laplace's equation, especially the *finite element method*.

12.12 References

Jones [25] gives a clear and well-motivated introduction to the Lebesgue integral. For a detailed account of measure theory, see Folland [12]. A concise, concrete introduction to the subject, including a discussion of L^p-spaces, is in Lieb and Loss [32]. For a detailed account of Sobolev spaces, see Adams [1]. For Sobolev spaces and elliptic PDEs, see Evans [11] and Gilbarg and Trudinger [15]. An extensive discussion of Sobolev spaces, variational problems, and related analysis of linear PDEs is contained in Dautry and Lions [7].

12.13 Exercises

Exercise 12.1 Prove that the Borel σ-algebra on $\mathbb{R}$ is generated by the following families of sets:

$$\{(a, b] \mid a < b\}, \qquad \{[a, b) \mid a < b\}, \qquad \{[a, b] \mid a < b\}, \qquad \{(a, \infty) \mid a \in \mathbb{R}\}.$$

Exercise 12.2 Let $\mathcal{A}$ be a σ-algebra of subsets of Ω, and suppose μ is a measure on Ω. Prove the following properties:

(a) If $A, B \in \mathcal{A}$, then $A \setminus B \in \mathcal{A}$;
(b) If $A, B \in \mathcal{A}$, and $A \subset B$, then $\mu(A) \leq \mu(B)$;
(c) If $A, B \in \mathcal{A}$, then $\mu(A \cup B) \leq \mu(A) + \mu(B)$.

Exercise 12.3 If (A_i) is an increasing sequence of measurable sets, meaning that

$$A_1 \subset A_2 \subset \ldots \subset A_i \subset A_{i+1} \subset \ldots,$$

then prove that

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} \mu\left(A_i\right).$$

If (A_i) is a decreasing sequence of measurable sets, meaning that

$$A_1 \supset A_2 \supset \ldots \supset A_i \supset A_{i+1} \supset \ldots,$$

and $\mu(A_1) < \infty$, prove that

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} \mu\left(A_i\right).$$

Give a counterexample to show that this result need not be true if $\mu(A_i)$ is infinite for every i.

Exercise 12.4 Give an example of a monotonic decreasing sequence of nonnegative functions converging pointwise to a function f such that the equality in Theorem 12.33 does not hold.

Exercise 12.5 Check that the counting measure defined in Example 12.6 is a measure.

Exercise 12.6 Use the dominated convergence theorem to prove Corollary 12.36 for differentiation under an integral sign.

Exercise 12.7 Prove that $f \sim g$ if and only if $f = g$ pointwise-a.e. defines an equivalence relation on the space of all measurable functions.

Exercise 12.8 Let $f_n : X \to \mathbb{C}$ be a sequence of measurable functions converging to f pointwise-a.e. Suppose there exists $g \in L^p(X)$ such that $|f_n| \leq g$ a.e. Prove that $f_n \to f$ in the L^p-norm.

Exercise 12.9 Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function. Prove the following properties:

(a) for all $x \in \mathbb{R}$ the left- and right-derivatives, $\varphi'_-(x)$ and $\varphi'_+(x)$ exist and satisfy

$$\varphi'_-(x) \leq \varphi'_+(x);$$

(b) φ is continuous on $\mathbb{R}$;

(c) for all $x \in \mathbb{R}$, there exists a constant $c \in \mathbb{R}$ such that

$$\varphi(y) \geq \varphi(x) + c(y - x) \qquad \text{for all } y \in \mathbb{R}. \tag{12.25}$$

The graph of the function $y \mapsto c(y - x)$ satisfying (12.25) is called a *support line* of φ at x.

Exercise 12.10 If $x, y \geq 0$ and $\epsilon > 0$ is any positive number, show that

$$xy \leq \frac{\epsilon}{2}x^2 + \frac{1}{2\epsilon}y^2.$$

This estimate is sometimes called the *Peter-Paul inequality.*

Exercise 12.11 Let $p_1, p_2, \dots, p_n$ be positive numbers whose sum is equal to one. Prove that for any nonnegative numbers $x_1, x_2, \dots, x_n$, we have the inequality

$$x_1^{p_1} x_2^{p_2} \dots x_n^{p_n} \leq p_1 x_1 + p_2 x_2 + \dots + p_n x_n.$$

Exercise 12.12 Prove the following generalization of Hölder's inequality: if $1 \leq p_i \geq \infty$, where $i = 1, \dots, n$, satisfy

$$\sum_{i=1}^{n} \frac{1}{p_i} = 1,$$

and $f_i \in L^{p_i}(X, \mu)$, then $f_1 \cdots f_n \in L^1(X, \mu)$ and

$$\left| \int f_1 \cdots f_n \, d\mu \right| \leq \|f_1\|_{p_1} \cdots \|f_n\|_{p_n}.$$

Exercise 12.13 Prove that $L^\infty([0, 1])$ is not separable, and that $C([0, 1])$ is not dense $L^\infty([0, 1])$.

Exercise 12.14 Prove that for any pair of distinct exponents $1 \leq p, q \leq \infty$, we have $L^p(\mathbb{R}) \not\subset L^q(\mathbb{R})$. Show that the function

$$f(x) = \frac{1}{|x|^{1/2}\sqrt{1 + \log^2 |x|}}$$

belongs to $L^2(\mathbb{R})$, but not to $L^p(\mathbb{R})$ for any $1 \leq p \leq \infty$ that is different from 2.

Exercise 12.15 If $f \in L^p(\mathbb{R}^n) \cap L^q(\mathbb{R}^n)$, where $p < q$, prove that $f \in L^r(\mathbb{R}^n)$ for any $p < r < q$, and show that

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

This result is one of the simplest examples of an *interpolation inequality.*

Exercise 12.16 A function $f : \mathbb{R}^n \to \mathbb{C}$ is said to be *L^p-continuous* if $\tau_h f \to f$ in $L^p(\mathbb{R}^n)$ as $h \to 0$ in $\mathbb{R}^n$, where $\tau_h f(x) = f(x-h)$ is the translation of f by h. Prove that, if $1 \leq p < \infty$, every $f \in L^p(\mathbb{R}^n)$ is L^p-continuous. Give a counter-example to show that this result is not true when $p = \infty$.
HINT. Approximate an L^p-function by a C_c-function.

Exercise 12.17 Prove that the unit ball in $L^p([0,1])$, where $1 \leq p \leq \infty$, is not strongly compact.

Exercise 12.18 Give an example of a bounded sequence in $L^1([0,1])$ that does not have a weakly convergent subsequence. Why doesn't this contradict the Banach-Alaoglu theorem?

Exercise 12.19 Let $1 \leq p \leq \infty$. Prove that if $f \in L^p(\mathbb{R})$, and its weak derivative is identically zero, then f is a constant function.

Exercise 12.20 Which of the following functions belongs to $H^1([-1,1])$?

(a) $f(x) = |x|$;
(b) $g(x) = \operatorname{sgn} x$;
(c) $h(x) = \sum_{n=1}^{\infty} (1/n)^{3/2} \sin nx$.

Exercise 12.21 Prove a Poincaré inequality of the form (12.21) for periodic functions defined on the n-dimensional torus $\mathbb{T}^n$.

Exercise 12.22 Use the Riesz representation theorem to prove that there is a unique weak solution $u \in H^1(\mathbb{R}^n)$ of the equation

$$-\Delta u + u = f$$

for every $f \in H^{-1}(\mathbb{R}^n)$. Show that $(-\Delta + I) : H^1(\mathbb{R}^n) \to H^{-1}(\mathbb{R}^n)$ is an isomorphism. Is $-\Delta : H^1(\mathbb{R}^n) \to H^{-1}(\mathbb{R}^n)$ an isomorphism?

Exercise 12.23 Suppose that $a : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is a sesquilinear form on a Hilbert space $\mathcal{H}$ that satisfies the hypotheses of the Lax-Milgram lemma in Theorem 12.81.

(a) Show that there is a bounded linear map $J : \mathcal{H} \to \mathcal{H}$ such that Jx is the unique element satisfying

$$a(x,y) = \langle Jx, y \rangle \qquad \text{for all } y \in \mathcal{H},$$

where $\langle \cdot, \cdot \rangle$ denote the inner product on $\mathcal{H}$.

(b) Show that $\alpha\|x\| \leq \|Jx\|$. Deduce that J is one-to-one and onto.
HINT. Show that $\operatorname{ran} J$ is closed and use the projection theorem to show that J is onto.

(c) Show that there is a unique solution x of the equation

$$\langle Jx, y\rangle = F(y) \qquad \text{for all } y \in \mathcal{H},$$

and prove the Lax-Milgram lemma.