MEASURES AND DISTRIBUTIONS[1]

JOSE-LUIS MENALDI[2]
(e-mail: jlm@math.wayne.edu)

Version of October 24, 2007[3]

---

[2] Wayne State University, Department of Mathematics, Detroit, MI 48202, USA

[3] This book is being progressively updated and expanded. If you discover any errors or you have suggested improvements please e-mail the author.

Contents

Part I

Measure and Integration Theory

As indicated by the title, this first part is an introduction to abstract measure and integration. The objective is the d-dimensional Lebesgue integral, but in the process, we develop some general properties valid for measures in metric spaces. Instead of taking a direct way to reach our goal (after a background chapter), we prefer a more systematic approach in which family of sets and measurable functions are presented first, without a direct motivation. Chapter 2 (abstract measures) begins with the classic Caratheodory construction and even if the typical example is the Lebesgue measure, we continue with Borel measures before establishing more specific properties on the Lebesgue measure. Basic on integrals is developed in Chapter 3, where the convergence theorem are obtained, and in Chapter 4, we give some complement on measure theory. Finally, in Chapter 5, we consider most of the useful results applicable to Euclidean d-dimensional spaces.

Therefore, before going in more details and as a preview of what is to come, we discuss discrete measures in a quick way, including the typical convergence theorems.

Recall that for any series of nonnegative real numbers $\sum_{i=1}^{\infty} a_i$, with $a_i \geq 0$, we have: (1) if ι is a bijective function from the positive integers into themselves then $\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} a_{\iota(i)}$; (2) if $I_1, I_2, \ldots$ is a partition (finite or non) of the positive integers then $\sum_{i=1}^{\infty} a_i = \sum_n \sum_{i \in I_n} a_i$. Now, we use a discrete example to motivate the discussion.

Let Ω be a non empty set and denote by 2^Ω the family of all the subsets of Ω. Suppose we choose a finite or at most countable subset I of Ω and a sequence of strictly positive real numbers $\{a_i : i \in I\}$. Consider $m : 2^\Omega \to [0, \infty]$ defined by $m(A) = \sum_{i \in I} a_i \mathbb{1}_A(i)$, where $\mathbb{1}_A(i) = \mathbb{1}_{\{i \in A\}}$ is equal to 1 only if $i \in A$ and zero otherwise. We have by definition (1) $m(\emptyset) = 0$ and we remark the following property, so-called σ-additivity:

(2) if $A = \bigcup_{i=1}^{\infty} A_i$ with $A_i \cap A_j = \emptyset$ for any $i \neq j$, then $m(A) = \sum_{i=1}^{\infty} m(A_i)$.

Such a function (on sets) m is called a *discrete measure*, the set I is the *set of atoms* and a_i is the measure (or weight) of the atom i. Clearly, to define m we need only to know the values $m(\{i\})$ for any i in the finite or countable set I. An element N of 2^Ω is called *negligible* with respect to m if $m(N) = 0$. In the case of discrete measures, any subset of N of $\Omega \smallsetminus I$ is negligible. If $m(\Omega) = 1$ then we say that m is a discrete probability measure.

A function $f : \Omega \to \mathbb{R}$ is called *integrable* with respect to m if the series $\sum_{i \in I} |f(i)| \, m(\{i\}) = \sum_{i \in I} |f(i)| \, a_i$ converges, and in this case the integral with respect to m is defined as the following real number:

$$\int_\Omega f \mathrm{d}m = \int f \mathrm{d}m = \sum_{i \in I} f(i) \, m(\{i\}) = \sum_{i \in I} f(i) \, a_i.$$

Even if the series diverges, if f is nonnegative then we can define the integral as above (a nonnegative number when the series converges and the symbol $+\infty$ otherwise). Next, a function f is called *quasi-integrable* with respect to m if either the positive part f^+ or the negative part f^- is integrable. The class of integrable functions is denoted by $\mathcal{L}$ or $\mathcal{L}(\Omega, 2^\Omega, m)$ if necessary.

A couple of properties are immediately proved for the integral:
(1) if $c \in \mathbb{R}$ and $f, g \in \mathcal{L}$ then $cf + g \in \mathcal{L}$ and $\int (cf + g)\mathrm{d}m = c \int f\mathrm{d}m + \int g\mathrm{d}m$;
(2) if $f \leq g$, quasi-integrable then $\int f\mathrm{d}m \leq \int g\mathrm{d}m$;
(3) $f \in \mathcal{L}$ if and only if $|f| \in \mathcal{L}$ and in this case $\big| \int f\mathrm{d}m \big| \leq \int |f|\mathrm{d}m$;
(4) if $f \geq 0$ and $\int f\mathrm{d}m = 0$ then $f = 0$ except in a negligible set.

There are three main ways of taking limit inside the integral for a sequence f_n of functions:
(a) Beppo Levi's monotone convergence: if $0 \leq f_n \leq f_{n+1}$ and $f(x) = \lim_n f_n(x)$ for any $x \in \Omega$ then $\int f\mathrm{d}m = \lim_n \int f_n\mathrm{d}m$;
(b) Fatou's lemma: suppose that $f_n \geq 0$ and $f(x) = \liminf_n f_n(x)$ then $\int f\mathrm{d}m \leq \liminf_n \int f_n\mathrm{d}m$;
(c) Lebesgue's dominated convergence: if $|f_n| \leq g$ with $g \in \mathcal{L}$ and $f(x) = \lim_n f_n(x)$ exists for any $x \in \Omega$ then $\int f\mathrm{d}m = \lim_n \int f_n\mathrm{d}m$.
Essentially, any one of the three theorems can be deduced from any of the others. For instance, use (a) on $g_n(x) = \inf_{k \geq n} f_k(x)$ to get (b) and use (b) on $g \pm f_n$ to deduce (c). To prove (a), for any number $C < \int f\mathrm{d}m$ there exists a finite set $J \subset I$ of atoms such that $\sum_{i \in J} f(i)\, m(\{i\}) > C$. Since J is finite, for every $\varepsilon > 0$ there exists $N = N(\varepsilon, C, J)$ such that $\sum_{i \in J} f_n(i)\, m(\{i\}) > C - \varepsilon$ for every $n \geq N$. Because $f_n \geq 0$ and C, ε are arbitrary we deduce $\int f\mathrm{d}m \leq \lim_n \int f_n\mathrm{d}m$, and the equality follows.

For instance, if $\Omega = \{1, 2, \ldots\}$ is the set of strictly positive integers then $m(\{i\}) = 2^{-i}$ defines a probability measure. Consider the sequence of functions $\{f_n\}$ given by $f_n(x) = 2^n \mathbb{1}_{\{n=x\}}$. We can verify that f_n is integrable, converges to the function identically zero but $\int f_n\mathrm{d}m = 1$ for every n, i.e., we cannot use (a) or (c) and the inequality in (b) is strict.

Negligible sets (or sets of measure zero) do not play an essential role for discrete measures since there is a largest set of measure zero, namely $\Omega \smallsetminus I$, i.e., for a given discrete measure m with atoms I we may ignore the complement of I. However, in general, negligible sets are a fundamental part, i.e., almost everything happens except a set of measure zero, referred to as *almost everywhere* (a.e.) or *almost surely* (a.s.) when a probability is used.

The generalization of these arguments is the basis of measure theory. However, before being able to reach the classic example of the Lebesgue-Borel measure and the extension of the Riemann-Stieltjes integral, several points should be discussed.

Chapter 0

Background

Before going into some details and without really discussion set theory (e.g. see Halmos [5]) we have to consider a couple of issues. On the other hand, we need to recall some basic topology, e.g. most of the material presented in the preliminaries section of Yosida [10, Chapter 0, pp. 1–22]. Alternatively, the reader may check part of the material in Royden [8, Chapters 1,2,7,8 and 9] or Dshalalow [2, Part I, pp. 1–200]. Reading what follows for the first time could be very dense, so that the reader should have some acquaintance with most of the concepts discussed in this preliminary chapter. Certainly, not every aspect (of this chapter) is need later, but it is preferred to face these possible difficulties now and not later, when the actual focus of interest is reveled.

0.1 Cardinality

We want to *count* the number of element of any set. If a set is finite, then the number of elements (also called the cardinal of the set) is a natural or nonnegative integer number (zero if the set is empty). However, if the set is infinite then some consideration should be made. To define the *cardinal* of a set in general, we say that two sets have the same cardinal or are *equipotent* if there exits a bijection between them. Since equipotent is a reflexive, symmetric and transitive relation, we have equivalence classes of sets with the same cardinal. Also we say that the cardinal of a set A is not greater than the cardinal of another set B if there is a injective function from A into B, usually denoted by $\mathrm{card}(A) \leq \mathrm{card}(B)$, and we can show that $\mathrm{card}(A) \leq \mathrm{card}(B)$ and $\mathrm{card}(B) \leq \mathrm{card}(A)$ imply that A and B have the same cardinal. Similarly, if A and B have the same cardinal then we write $\mathrm{card}(A) = \mathrm{card}(B)$, and also $\mathrm{card}(A) < \mathrm{card}(B)$ with obvious meaning.

We use the symbol $\aleph_0$ (aleph-nought) for the cardinal of $\mathbb{N}$ the set of natural (or nonnegative integer) numbers, and we show that $\aleph_0$ is first nonfinite cardinal. Any set in the class $\aleph_0$ is called *countable* infinite, while any countable set, finite or infinite, is called *denumerable*. With time, the two names countable and denumerable are used indistinctly and if necessary, we have to specify finite or infinite for countable sets. It can be shown that the integer numbers $\mathbb{Z}$ and the rational numbers $\mathbb{Q}$ are both countable sets, moreover, the Cartesian product or the union does not change the cardinality of infinite sets, i.e., if $\{A_i : i \in I\}$ is a family (or collection) of sets with A_i and I all with cardinal not greater than some $\aleph$ then $\prod_{i \in I} A_i$ and $\bigcup_{i \in I} A_i$ have cardinal not greater than $\aleph$. However, we can also show that the cardinal of 2^A, the set of the parts of a nonempty set A (i.e., the set of all subsets of A) has cardinal strictly greater than $\mathrm{card}(A)$. Nevertheless, if A is an infinite set then the set composed by all subsets of A having a finite number of element (called the finite-parts of A) have the same cardinal as A. Indeed, if $A = \{1, 2, \ldots\}$ then the set $2^A_{_F}$ of the finite-parts of A can be represented (omitting the empty set) as finite sequences $a = \{a_1, \ldots, a_n\}$ of elements a_i in A. Thus, if $\{2, 3, 5, 7, \ldots, p_i, \ldots\}$ is the sequence of all prime numbers then for each a there is a unique positive integer $m = 2^{a_1} 3^{a_2} 5^{a_3} 7^{a_4} \cdots p_n^{a_n}$, and because the factorization in term of the prime numbers is unique, the mapping $a \mapsto m$ is one-to-one, i.e., $2^A_{_F}$ is countable.

Representing real numbers in binary form, we observe that $2^{\{0,1,\ldots\}}$ has the same cardinal as the real numbers $\mathbb{R}$, which is strictly greater than $\aleph_0$. The cardinality of $\mathbb{R}$ is called cardinality of *continuum* and denoted by $2^{\aleph_0}$. However, we do not know whether or not there exists a set A with cardinal $\aleph$ such that $\aleph_0 < \aleph$ and $\aleph < 2^{\aleph_0}$. Any way, we have $\aleph_1 = 2^{\aleph_0}$, $\aleph_2 = 2^{\aleph_1}$, and so on.

The *continuum hypothesis* states that for any infinite set A *there is no set with cardinal strictly between the cardinal of* A *and the cardinal of* 2^A. In particular for $\aleph_0$ and $2^{\aleph_0}$, this assumption has an equivalent formulation as follows: the set $\mathbb{R}$ can be well-ordered in such a way that each element of $\mathbb{R}$ is preceded by only countably many elements, i.e., there is a relation $\preceq$ satisfies (a) for any two real numbers x and y we have $x \preceq y$ or $y \preceq x$ or $x = y$ (linear order), (b) for every real number x we have $x \preceq x$ (reflexive), and (c) every nonempty subset of real numbers A has a first number, i.e., there exist a_0 in A such that $a_0 \preceq a$ for any a in A (well-ordered), and the extra condition (d) for every real number x the set of real numbers $y \preceq x$ is a countable set.

0.2 Frequent Axioms

Sometimes, we need to differentiate sets with the same cardinal based on other characteristics, e.g., a natural order of numbers or the natural inclusion for collection or family of sets.

A *partially* ordered set $(X, \preceq)$ is a set X (family of sets) and a relation $\preceq$, which is transitive ($a \preceq b$ and $b \preceq c$ imply $a \preceq c$) and antisymmetric ($a \preceq b$ and $b \preceq a$ imply $a = b$). An order $\preceq$ (on a set X) is called (1) *linear* (or total, and the set X is linearly ordered) if for every a and b in X we have either $a \preceq b$ or $b \preceq a$, and (2) *well-order* (and the set X is well-ordered) if (1) holds and any nonempty subset A of X has a minimum element, i.e., there is $\underline{a}$ in A such that $\underline{a} \preceq a$, for every a in A.

Certainly (several) well-order $\preceq$ can be given to a finite set, and any subset of integer numbers with a finite infimum inhered a well-order from the integer numbers. A typical situation is to partially order a collection of sets with the inclusion. Note that the natural order of the real numbers $\mathbb{R}$ is a linear order, but not a well-order. Also, the set $\mathbb{R}^I$ of all real-valued functions on a set I (of more than one element), with the natural partial order $(x_i) \preceq (y_i)$ if $x_i \leq y_i$ for every i in I, is not a linear order.

In a partially ordered set not all elements are comparable, i.e., we may have two elements a and b such that neither $a \preceq b$ nor $b \preceq a$. Thus, given a partially ordered set $(X, \preceq)$ and a subset $A \subset X$, we say that x in X is an *upper bound* of A if $a \preceq x$ for every a in A, and if x belongs to A then x is the maximum element of A. Maximum has little use for partially ordered set, instead, we say that m in A is a *maximal element* of A if for any a in A such that $m \preceq a$ we have $a = m$ (i.e., m is larger or equal to any other element a in A that can be compared with m). A *chain* in X is a subset $C \subset X$ such that $\preceq$ becomes a linear order on C.

There are several equivalent ways of expressing the well-ordering principle (or axioms), e.g.,

Hausdorff Maximal Principle: Every partially ordered set $(X, \preceq)$ has a maximal chain, i.e., a subset C of X such that $(C, \preceq)$ is a linearly ordered set and $(C', \preceq)$ is not a linearly ordered set, for any subset C' strictly containing C. □

Zorn's Lemma: Every nonempty partially ordered set has a maximal element if any chain has an upper bound. □

Zermelo's Axiom: Every set can be well-ordered, i.e., if X is a set, then there is some well-order $\preceq$ on X, i.e., $\preceq$ is a linear order (all elements in X are comparable) and every nonempty subset of X has a first element. □

A typical use is when a construction of sets satisfying some properties is partially ordered (e.g., by the inclusion), and we deduce the existence of a maximal set satisfying those properties.

Related to the above assumptions, but independent from other axioms of the set theory, is the so-called *Axiom of Choice* (AoC), which can also be expressed in various equivalent ways, e.g.,

AoC Form (a): The Cartesian product of any nonempty family of nonempty sets must be a nonempty set, i.e., if $\{A_i : i \in I\}$ is a family of sets such that $I \neq \emptyset$ and $A_i \neq \emptyset$, for any $i \in I$ then there exits at least one choice $a_i \in A_i$, for any $i \in I$. □

AoC Form (b): If $\{A_i : i \in I\}$ is a family of arbitrary nonempty disjoint sets indexed by a set I, then there exists a set consisting of exactly one element form each A_i, with $i \in I$. □

All these axioms come into play when dealing with uncountable sets.

Beside using cardinality to classify sets (mainly sets involving numbers), we may push further and classify well-ordered sets. Thus, similarly to the cardinality, we say that two well-ordered sets $(X, \preceq)$ and $(Y, \preceq)$ have the same *ordinal* if there is a bijection between them preserving the order. Thus, to each well-ordered set we associate an ordinal (an equivalence class). It clear that finite ordinals are the sets of natural numbers $\{1, 2, \ldots, n\}$, $n = 1, 2, \ldots$, with the natural order, and the first infinite ordinal is the set of natural numbers $\mathbb{N} = \{1, 2, \ldots\}$ (or equivalently any infinite subset of integer number with a finite infimum). Each ordinal has a next ordinal, i.e., given an ordinal $(X, \preceq)$ we define $(X+1, \preceq)$ to be $X+1 = X \cup \{\infty\}$, with $\infty \notin X$ and $x \preceq \infty$, for every x in X. However, each ordinal not necessarily has a previous (or precedent) ordinal, e.g., there is not an ordinal X such that $X + 1 = \mathbb{N}$.

Similarly to cardinals, we say that the ordinal of a well-ordered set $(X, \preceq)$ is not greater than the ordinal of another well-ordered set $(Y, \preceq)$ if there is a injective function from X into Y preserving the linear order, usually denoted by $\mathrm{ord}(X) \leq \mathrm{ord}(Y)$. Thus, we can show that if $\mathrm{ord}(X) \leq \mathrm{ord}(Y)$ and $\mathrm{ord}(Y) \leq \mathrm{ord}(X)$ then $(X, \preceq)$ and $(Y, \preceq)$ have the same ordinal. Moreover, there are many properties satisfied by the ordinal (e.g., see Kelley [7, Appendix, 240–281]), we state for future reference

Ordinal Order: The set (actually class or family of sets) of all ordinals is well-ordered (the above order denoted by $\leq$ and the strict order by $<$, i.e., the $\leq$ and the $\neq$), namely, every nonempty subset (subfamily) of ordinals has a first ordinal. Moreover, given an ordinal x, the set $\{y < x\}$ of all ordinal strictly precedent to x has the same ordinal as x. □

For instance, we may call ω_0 the first uncountable ordinal, i.e., any ordinal $\omega < \omega_0$ is countable (finite or infinite). It is clear that there are plainly of ordinals between $\mathbb{N}$ and ω_0. Thus, transfinite induction and recursion can used with ordinals, i.e., first for any element a of a well-ordered set $(X, \preceq)$ we define the initial segment of X determined by a, i.e., $I(a) = \{x \in X : x \preceq a,\ x \neq a\}$, and we have

Transfinite Induction Principle: If a subset A of a well-ordered set X satisfies (for every a in X) the condition $I(a) \subset A \Rightarrow a \in A$ then A is indeed the whole set, i.e., $A = X$. □

It is clear that if a is the minimum element in $X \smallsetminus A$ then by definition $I(a) \subset A$ and therefore a belongs to A. A neat case is when the well-ordered set X is actually the natural number, i.e., the ordinary mathematical induction. Similarly to the ordinary recursion argument, where a function f on the natural numbers can be defined by specifying $f(0)$ and then defining $f(n)$ in terms of $f(0), \ldots, f(n-1)$, we have the recursion principle for well-ordered set, e.g., see Dudley [3, Section I.3, pp. 12–15]. Alternatively, the reader may check Folland [4, Chapter 0, pp. 1–17], for a short and clean discussion on the above points.

0.3 Metrizable Spaces

First, remark that we use only with topological spaces where points are separated closed sets, i.e., Hausdorff spaces, so that these properties are implicitly assumed everywhere (even if it is seldom restated) in the text, unless explicitly mentioned otherwise.

A metric on a set X is a function $\mathrm{d} : X \times X \to \mathbb{R}$ satisfying for every x and y in X the following conditions: (a) $\mathrm{d}(x, y) \geq 0$ and $\mathrm{d}(x, y) = 0$ if and only if $x = y$, (b) $\mathrm{d}(x, y) = \mathrm{d}(y, x)$ and (c) $\mathrm{d}(x, y) \leq \mathrm{d}(x, z) + \mathrm{d}(z, y)$ for every z in X. The couple (X, d) is called a metric space, which becomes a topological space with the open sets defined by means of open balls $B(x, r) = \{y \in X : \mathrm{d}(x, y) < r\}$, for any x in X and any $r > 0$. On the other hand, a metrizable space is a topological space X in which a metric can be defined (but not really used) so that (X, d) has a topology equivalent to the initial one given on X (where the topology has a simple characterization).

In a metric space, for every x in X the countable family of balls $B(x, 1/n)$, $n = 1, 2, \ldots$, forms a neighborhood-basis at x and so, the (X, d) topology is *first-countable* . In a first-countable topology (in particular in a metric space), we can use only convergent sequences to define its topology, i.e., a subset A of X is closed for the topology induced by the metric d if and only if for every sequence $\{a_n : n \geq 1\}$ of points in A such that $\mathrm{d}(a_n, a) \to 0$ as $n \to \infty$, for some a in X, we have that the limit point a belongs to A.

On the other hand, a topological space X is called *separable* if there exists a countable dense subset $Q \subset X$, i.e., Q is countable and its closure $\overline{Q}$ is the whole space X. Hence, a separable metric space (X, d) is *second-countable*, i.e., its contains a countable basis, namely, the family of balls $B(q, 1/n)$ with q in a countable dense set Q and $n \geq 1$. Actually the converse is also true, i.e., a metric space (X, d) is second-countable if and only if it is separable, and similarly, a topological space with a countable basis is separable. Recall that a topological space is called *sequentially compact* if every sequence admits a convergent subsequence. Any sequentially compact metric space is separable, and for a sequential space (i.e., where convergent sequences to define its topology), compactness and sequentially compactness are equivalent. Moreover any locally compact (or vector topological) space with a countable basis is metrizable, but certainly, the converse is false.

Given a family $\{(X_i, \mathrm{d}_i) : i \in I\}$ of metric spaces the product space $X = \prod_{i \in I} X_i$ is a topological space with the product topology which may not be metrizable. For a countable family $I = \{1, 2, \ldots\}$, we may define the metric

$$\mathrm{d}(x, y) = \sum_{i=1}^{\infty} 2^{-i} \frac{\mathrm{d}_i(x_i, y_i)}{1 + \mathrm{d}_i(x_i, y_y)}, \qquad \forall x = (x_i),\ y = (y_i),$$

which induces an equivalent topology on X, i.e., a countable product space $X = \prod_{i=1}^{\infty} X_i$ is indeed metrizable with the above metric d.

We may consider uniform continuity and Cauchy sequences in a metric space (X, d). Thus, (X, d) is *complete* if any Cauchy sequence has a limit, i.e., if $\mathrm{d}(x_n, x_m) \to 0$ as $n, m \to \infty$ then there exists x such that $\mathrm{d}(x_n, x) \to 0$ as $n \to \infty$. If a space (X, d) is not complete then we can completed it in the same way as we pass from the rational number to the real numbers. However, the concept of *completeness* is not a topological property, i.e., on a given space we may have two metrics yielding equivalent topologies but only one of them is complete. Anyway, every compact metric space is complete. A complete separable metrizable space is called a *Polish space*, i.e., a separable topological space X with a metric yielding a complete metric space (X, d). For instant, the space $C(\mathbb{R}^d)$ of all real-valued continuous functions is a Polish space with the metric

$$\mathrm{d}(f,g) = \sum_{n=1}^{\infty} 2^{-n} \frac{\|f-g\|_n}{1+\|f-g\|_n}, \qquad \|f\|_n = \sup_{|x|\le n} |f(x)|\}.$$

In probability, the sample spaces are Polish spaces, most of the times, we use the space of continuous functions from $\mathbb{R}$ into $\mathbb{R}^d$ or the space of all cad-lag functions from $\mathbb{R}$ into $\mathbb{R}^d$, i.e., functions continuous from the right and having limits from the left.

A subset K of a metric space (X, d) is called *totally bounded* if for every $\varepsilon > 0$ there exists a finite number of points $x_1, \ldots, x_n$ in K such that $K \subset \bigcup_{i=1}^{n} B(x_i, \varepsilon)$, i.e., any x in K is within a distance ε from the set $\{x_1, \ldots, x_n\}$. It is very instructive (but no simple) to show that a subset K (of a complete metric space) is totally bounded if and only if the closure of K is compact, e.g., Yosida [10, Section 0.2, pp. 13–15].

A *vector topological space* has a topology compatible with the vector structure, i.e., such that the addition and the scalar multiplication (of vectors) are continuous operations. An example is the so-called locally convex spaces, and better, a *normed space* X, which is vector space with a *norm*, i.e., a nonnegative function $\|\cdot\|$ defined on X such that (a) $\|\lambda x\| = |\lambda| \, \|x\|$, for every x in X and λ in $\mathbb{R}$, (b) $\|x+y\| \le \|x\| + \|y\|$, for every x and y in X, and (c) $\|x\| \ge 0$ for every x in X and $\|x\| = 0$ only if $x = 0$. Given a norm, we define a metric $\mathrm{d}(x,y) = \|x-y\|$ (but not any metric comes from a norm), which yields the topology.

In a normed space, any set that can be covered by a ball is called a *bounded set*. But, only on finite dimensional normed spaces, we show that closed and bounded sets are also compacts. A complete normed space is called a *Banach space*. The space $C_b(X)$ of all real-valued (or complex-valued) bounded continuous functions on a Hausdorff topological space X, with the sup-norm

$$\|f\| = \sup\big\{|f(x)| \,:\, x \in X\big\},$$

is a typical example of an infinite dimensional Banach space.

A topological space X is said to be *locally compact* if every point has a compact neighborhood, i.e., an open set with compact closure. This implies that for every point x in X and any open set U containing x there is another open set V containing x such that $\overline{V} \subset U$ and the closure $\overline{V}$ is compact. A locally compact Banach space is necessarily a space of finite dimension (i.e., homeomorphic to some $\mathbb{R}^d$, $d \ge 1$).

Again, better than a norm is an *inner* or *scalar product*, i.e., a bilinear maps $(\cdot,\cdot)$ from $X \times X$ into $\mathbb{R}$ satisfying (a) $(\lambda x + y, z) = \lambda(x,z) + (y,z)$, for every x, y in X and λ in $\mathbb{R}$, (b) $(x,y) = (y,x)$, for every x, y in X, and

(c) $(x, x) \geq 0$ and $(x, x) = 0$ only if $x = 0$. From an inner product we can define a norm $\|x\| = \sqrt{(x, x)}$, indeed, by considering the discriminant of the positive quadratic form $\lambda \mapsto (x + \lambda y, x + \lambda)$ we obtain the Cauchy inequality $|(x, y)| \leq \|x\| \, \|y\|$, for every x and y, which yields the triangular inequality (b) for the norm. Certainly, not any norm comes from an inner product, indeed, any norm derived form a inner product satisfies the parallelogram law, i.e., $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$, for every x and y in X. A complete normed space where the norm comes from an inner product is called a *Hilbert space.*

A typical infinite dimensional Hilbert space is ℓ^2, the space of real-valued sequences $a = \{a_n\}$ satisfying $\sum_n a_n^2 < \infty$, with the inner product $(a, b) = \sum_n a_n b_n$. A more elaborate example is space $L^2(K)$, with K a compact subset of $\mathbb{R}^d$, which is the completion of $C_b(K) = C(K)$, space of continuous functions, with the inner product

$$(f, g) = \int_K f(x) \, g(x) \, \mathrm{d}x.$$

By means of the theory of the integral we are able to study in great detail spaces similar to this one.

Moreover, we may have complex Banach and Hilbert spaces, i.e., the vector space is on the complex field $\mathbb{C}$, and $|\lambda|$ denotes the modulus (instead of the absolute value) when λ belongs to $\mathbb{C}$ for the condition (a) of norm. In the case of a inner product, the condition (b) becomes $(x, y) = \overline{(y, x)}$, where the over-line means complex conjugate, i.e., the application $(\cdot, \cdot)$ is sesquilinear (instead or bilinear) with complex values.

For instance, the reader may take a look at DiBenedetto[1, Chapters 1 and 2, pp. 1–64] or Hewitt and Stromberg [6, Chapter 2, pp. 53–103] or Royden [8, Chapters 1 and 2, pp. 1–53] or Rudin [9, Chapter 1, pp. 3–40], for a quick review on topology and continuous functions.

0.4 Some Basic Lemmas

There are a couple of useful results, such as

Lemma 0.1 (Urysohn)**.** *Let A and B be two nonempty, disjoint and closed sets in a metric space (X, d). Then the function*

$$x \mapsto g(x) = \frac{\mathrm{d}(x, A)}{\mathrm{d}(x, A) + \mathrm{d}(x, B)}$$

is continuous and $g(x) = 0$ for any x in A and $g(x) = 1$ for any x in B. □

Based on the previous Lemma, we have

Exercise 0.2 (Urysohn's Lemma)**.** Let A and B be two nonempty, disjoint and closed sets in a metric space (X, d). Now, given two real number a and b, construct a continuous functions $h : X \to [a, b]$ such that $h(x) = a$ for any x in A and $h(x) = b$ for any x in B. □

Proposition 0.3 (Tietze's Extension)**.** *Let f be a bounded real-valued function defined on a closed subset C of a metric space (X, d). Then there exists a continuous extension g of f to the entire space X.* □

Statements in Lemma 0.1 and Proposition 0.3 are also valid for more general topological spaces, e.g., Hausdorff locally compact spaces and normal spaces.

The proof of Tietze's Extension is based on the following construction, where we show that g is uniform limit in X of the series $\sum_k g_k$ of continuous functions defined as follow:

(First) with $a = \sup_C |f(x)|$ define $A = \{x \in C : f(x) \leq -a/3\}$ and $B = \{x \in C : f(x) \geq a/3\}$ to find a continuous function $g_1 : X \to [-a/3, a/3]$ satisfying $|f(x) - g_1(x)| \leq 2a/3$, for every x in C.

(Next) with $f_1 = f - g_1$, $a_1 = 2a/3$ define $A_1 = \{x \in C : f_1(x) \leq -a_1/3\}$ and $B_1 = \{x \in C : f_1(x) \geq a_1/3\}$ to find a continuous function $g_2 : X \to [-a_1/3, a_1/3]$ satisfying $|f_1(x) - g_2(x)| \leq 2a_1/3$, for every x in C.

(Show) inductively that $|f(x) - \sum_{k=1}^n g_k(x)| \leq 2^n a 3^{-n}$, for every x in C and that $|g_n(x)| \leq 2^{n-1} a 3^{-n}$, for any x.

A vector space $\mathtt{F}$ of real-valued functions, on an arbitrary nonempty X, is called an algebra if for any f and g in $\mathtt{F}$ the product function fg belongs also to $\mathtt{F}$. A class of real-valued functions $\mathtt{F}$ is said to separate points in X if for every $x \neq y$ in X there exists a function f in $\mathtt{F}$ such that $f(x) \neq f(y)$. If X is a Hausdorff topological space then $C(X)$ the space of all continuous real-valued functions defined on X is an algebra that separate points.

Theorem 0.4 (Stone-Weierstrauss)**.** *Let K be a compact Hausdorff space and $\mathtt{F}$ be an algebra of functions in $C(K)$ which separate points and contains constants functions. Then for every $\varepsilon > 0$ and any g in $C(K)$ there exists f in $\mathtt{F}$ such that*

$$\sup \big\{ |f(x) - g(x)| \,:\, x \in X \big\} = \|f - g\| < \varepsilon,$$

i.e., $\mathtt{F}$ is dense in $C(K)$ for the sup-norm $\|\cdot\|$. □

As a particular case (Weierstrauss approximation theorem) we have that the set of all polynomials in d variables is dense in $C(K)$ with the sup-norm, for every compact subset K of $\mathbb{R}^d$. For instance, a proof of Stone-Weierstrauss Theorem 0.4 can be found in DiBenedetto [1, Sections IV.16–18, pp. 199–203] or Yosida [10, Section 0.2, pp. 8–11].

Typical applications of Zorn's Lemma are the following:

Lemma 0.5. *Any linear vector space X contains a subset $\{x_i : i \in I\}$, so-called Hamel basis for X, of linear independent elements such that the linear subspace spanned by $\{x_i : i \in I\}$ coincides with X.*

Proof. Consider the set S whose elements are all the subsets of linearly independent elements in X, with the partial order given by the inclusion. If $\{\mathcal{A}_\alpha\}$

is a chain or totally ordered subset of S then $\mathcal{A} = \bigcup_\alpha \mathcal{A}_\alpha$ is an upper bound, since any finite number of element in $\mathcal{A}$ are linearly independent. Hence, Zorn's Lemma implies the existence of a maximal element, denoted by $\{x_i : i \in I\}$. Now, for any x in X the set $\{x_i : i \in I\} \cup \{x\}$ has to be linearly dependent and so, x is linear combination of some finite number of elements in $\{x_i : i \in I\}$, as desired. □

Recall that a real linear functional T on a (linear) vector space X is a linear mapping from X into $\mathbb{R}$, i.e., satisfying $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$, for every x, y in X and α, β in $\mathbb{R}$.

Lemma 0.6 (Hahn-Banach)**.** *Let X be a real linear vector space and p be a function from X into $[0, \infty)$ satisfying*

$$\mathrm{p}(x + y) \leq \mathrm{p}(x) + \mathrm{p}(y), \qquad \mathrm{p}(\lambda x) \leq \lambda \mathrm{p}(x), \qquad \forall \lambda \geq 0,\ x, y \in X.$$

If X_0 is a linear subspace X_0 of X and $T_0 : X_0 \to \mathbb{R}$ is a linear mapping satisfying $T_0(x) \leq \mathrm{p}(x)$ for every x in X_0, then there exits a real linear function T on X such that $T(x) \leq \mathrm{p}(x)$ for every x in X and $T(x) = T_0(x)$, for any x in X_0.

Proof. Consider the family $\mathcal{A}$ of all pairs (X_α, T_α) where X_α is a linear subspace containing X_0 and T_α is a real linear functional defined on X_α such that $T_\alpha(x) \leq \mathrm{p}(x)$ for every x in X_α and $T_\alpha(x) = T_0(x)$, for any x in X_α. The order relation in $\mathcal{A}$ is given by the condition $(X_\alpha, T_\alpha) \prec (X_\beta, T_\beta)$ if $X_\alpha \subset X_\beta$ and $T_\alpha = T_\beta$ on X_α.

Since any chain or totally ordered subset of $\{(X_\beta, T_\beta)\}$ has $(\bigcup_\beta X_\beta, T')$, with $T' = T_\beta$ on X_β, as an upper bound, we can use Zorn's Lemma to find a maximal element (X_1, T_1). We show that $X_1 = X$ by contradiction. Indeed, if x_1 belongs to $X \smallsetminus X_1$ then the subspace $X_2 = \{x + \lambda x_1 : x \in X_1,\ \lambda \in \mathbb{R}\}$ strictly contains X_1, and we can define $T_2(x + \lambda x_1) = T_1(x) + \lambda c$, for some suitable constant c satisfying

$$T_1(x) + \lambda c \leq \mathrm{p}(x + \lambda x_1), \quad \forall x \in X_1,\ \lambda \in \mathbb{R}, \tag{1}$$

we obtain a contradiction.

Now, we remark that $-\mathrm{p}(-y - x_1) - T_1(y) \leq \mathrm{p}(x + x_1) - T_1(x)$, for every x, y in X and we select any c such that

$$\sup_{y \in X_0} \big\{ -\mathrm{p}(-y - x_1) - T_1(y) \big\} \leq c \leq \inf_{x \in X_0} \big\{ \mathrm{p}(x + x_1) - T_1(x) \big\},$$

to obtain a constant c satisfying condition (1). □

For instance, the interested reader may check the books by DiBenedetto [1], Dudley [3], Dshalalow [2], Hewitt and Stromberg [6] and Royden [8], among many others.

Bibliography

[1] E. DiBenedetto. *Real analysis.* Birkhäuser Boston Inc., New York, 2002.

[2] J.H. Dshalalow. *Real Analysis.* Chapman & Hall / CRC Press, Boca Raton (FL), 2001.

[3] R.M. Dudley. *Real Analysis and Probability.* Cambridge University Press, Cambridge, 2002.

[4] G.B. Folland. *Real Analysis.* Wiley, New York, 2nd edition, 1999.

[5] P.R. Halmos. *Naive Set Theory.* Springer-Verlag, New York, 1974.

[6] E. Hewitt and K. Stromberg. *Real and Abstract Analysis.* Springer-Verlag, New York, 1965.

[7] J.L. Kelley. *General topology.* Springer-Verlag, New York, 1975. Reprint of the 1955 edition.

[8] H. Royden. *Real Analysis.* Prentice-Hall, Englewood Cliffs (NJ), 3rd edition, 1988.

[9] W. Rudin. *Functional analysis.* International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, second edition, 1991.

[10] K. Yosida. *Functional Analysis.* Springer-Verlag, New York, 1965.