Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

1078

A. J. E. M. Janssen
P. van der Steen

Integration Theory

Springer-Verlag
Berlin Heidelberg New York Tokyo

Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

1078

A. J. E. M. Janssen
P. van der Steen

Integration Theory

Springer-Verlag
Berlin Heidelberg New York Tokyo 1984

Authors

A. J. E. M. Janssen
Philips' Research Laboratories
P.O.Box 80.000, 5600 JA Eindhoven, The Netherlands

P. van der Steen
Department of Mathematics, University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

AMS Subject Classification (1980): 28-01, 26-01

ISBN 3-540-13386-0 Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-13386-0 Springer-Verlag New York Heidelberg Berlin Tokyo

Printing and binding: Beltz Offsetdruck, Hemsbach / Bergstr.
2146/3140-543210

PREFACE

Presenting yet another book on integration theory requires some justification. When writing the present material, we had in mind to explain what Lebesgue integration is and how it can be developed. An important point was to reconcile the various methods to introduce the integral. Many of the ideas used occur already in papers by Stone in 1948-1950. But the form we present them in, and much else as well, springs from a series of lectures by N.G. de Bruijn, around 1964. The General Introduction extensively explains our intentions.

Thanks are due to N.G. de Bruijn: without him the book would never have been written; to J.W. Nienhuys, who critically read portions of the manuscript; and to David Klarner, who read the whole manuscript and suggested many improvements. Parts of the manuscript were typed at the Mathematical Department of the Technological University Eindhoven. The final typing was done by Mrs. Elsina Baselmans-Weijers, who did a superb job, as usual.

A.J.E.M. Janssen

P. van der Steen

CONTENTS

GENERAL INTRODUCTION

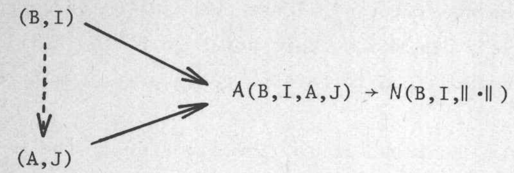
The main purpose of this book is to present and compare various ways to introduce Lebesgue integration. The underlying observation is that the usual methods (catchwords: Carathéodory, Bourbaki, Daniell) from a certain point on follow similar courses. We shall show that these methods can in fact go together a considerable part of the way.

This common part can be roughly described as follows. In each of the methods one obtains somehow a space B of basic functions on a set S, together with a positive linear functional I on B that has a certain continuity property. Then one enlarges B to a complete space L and extends I to a positive linear functional (again denoted by I) on L. In the situation thus obtained, denoted symbolically by $\mathcal{L}(L,I)$ or $\mathcal{L}$, one has Lebesgue's dominated convergence theorem, which is really at the heart of any theory of integration.

How one gets at the class B depends on the point of departure chosen. In the text the following three possibilities are discussed. First there is Carathéodory's method, the starting point of which is denoted by $\mathcal{R}$. In this case one has a ring (or a semiring) Γ of subsets of S and a measure μ defined on Γ. From the pair (Γ,μ) the space B and the appropriate functional I are readily constructed. Then one has Bourbaki's point of view, here denoted by $\mathcal{T}$. In this case B consists of the real-valued continuous functions of compact support on a locally compact Hausdorff space S, and I is assumed to be given from the outset as a positive linear functional on B (without further requirements). Finally, in Daniell's method as extended by Stone both B and I are assumed to be given a priori. As noted before, I must satisfy some continuity condition. There are two feasible possibilities for this condition; they are denoted by $\mathcal{D}$ or $\mathcal{D}'$, where $\mathcal{D}'$ is the stronger of the two. It turns out that $\mathcal{R}$ leads to $\mathcal{D}$, and $\mathcal{T}$ to $\mathcal{D}'$, so in a way the first two methods are concrete, and the last one is an abstraction of the first two.

Suppose we have a class B of basic functions on S with a positive linear functional I and an appropriate continuity condition $\mathcal{D}$ or $\mathcal{D}'$, how do we get at $\mathcal{L}$? As said before, B must be completed and I suitably extended. For this completion we use a norm defined on the class Φ of the $\mathbb{R}^*$-valued functions on the underlying set S. One of the requirements for such a norm is that it coincides with I for the non-negative members of B, and another is that it satisfies a strong triangle inequality. A situation where we have B, I and a suitable norm $\|\cdot\|$ is called a station $\mathcal{N}(B,I,\|\cdot\|)$, or $\mathcal{N}$. (There will appear many such stations; they always consist of function classes, functionals, and certain relationships between these objects. The reason for using the term "station" will be clarified later.)

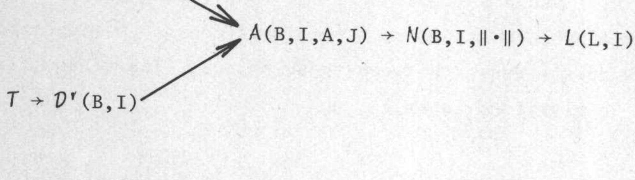
Now the norm can be defined in two ways, depending on whether the starting point was $\mathcal{D}$ or $\mathcal{D}'$, but both are instances of a more general construction, indicated in the diagram below.



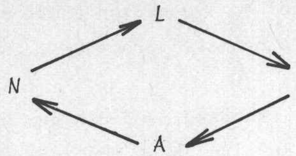
In this construction we assume that besides the class B of basic functions and the functional I, we are given a class A of auxiliary functions (certain non-negative $\mathbb{R}^*$-valued functions on S) together with a positive functional J on A. Such a situation is denoted by $\mathcal{A}$, or $\mathcal{A}$(B,I,A,J), and it is easy to get from $\mathcal{A}$ to $\mathcal{N}$. Of course, A and J must be connected in some way with B and I, and cannot be chosen arbitrarily. One of the restrictions is that the values of the norm and the functional I should coincide for non-negative basic functions. The dotted arrow in the diagram indicates that in some situations the pair (A,J) may be derived from the pair (B,I). In the $\mathcal{D}$- or $\mathcal{D}'$-case this can in fact be done. Due to the restriction mentioned above there is in the $\mathcal{D}$-case virtually only one possible choice for A and J: here A must consist of the non-negative members of B, and J must be a restriction of I. This is equivalent to Daniell's original approach. Since $\mathcal{D}'$ is stronger than $\mathcal{D}$, so that any station $\mathcal{D}'$(B,I) is also a station $\mathcal{D}$(B,I), the same procedure may be applied in the $\mathcal{D}'$-case. In this case, however, there is a second possibility (an abstraction of part of the Bourbaki version of integration theory), which quite often furnishes a bigger class of auxiliary functions and a richer theory of integration.

Departing from $\mathcal{N}$, the class of integrable functions is defined: it consists of all real-valued functions on S that can be approximated arbitrarily closely (with respect to the norm) by basic functions. The integral itself is then derived from I on B by arguments of continuity, and the result is a station $\mathcal{L}$(L,I) where Lebesgue's dominated convergence theorem holds, which was the ultimate aim.

The diagram below depicts the stations described up to now and their connections.



Eventually all paths lead to L, from which point the theory may be developed further. This is not the whole truth, however, for it will turn out that every station L is also a D, that is, an arrow may be added from L to D. The new diagram then contains the following substructure:



The way it is drawn suggests the name "circle line" for this structure and the term "station" for D, A, N and L.

The name "circle line" itself suggests several questions. For instance, one can start from D and develop the theory by means of $D \to A \to N \to L$. What happens if one makes the transition $L \to D$, and repeats the process? The answer is simple: nothing changes. In particular, if one starts from T or R and makes the transitions indicated by the arrows, then the class of integrable functions and the integral will be definitively fixed at the first confrontation with L.

Now we give a short description of the contents of each of the seven chapters. Chapter 0 contains some preliminary material: the fundamental notion of a Riesz function space, enough topology to read the chapter on integration on locally compact Hausdorff spaces, something about Riemann-Stieltjes integration, and unordered summation. Most of this will be familiar to most readers. In Chapter 1 the stations of the circle line are developed, the connections between stations are described, and some questions are studied that arise from the possibility of travelling more than once in the circle line. In Chapter 2 the theory is further developed; here L or N is the starting point; one meets measurability, L^p-spaces and local null functions. The next chapter is devoted to measure theory. It contains the description of R and the connection between R and D. Since R has more structure than D (or A, N, or L), there are more refined results connected with approximation of integrable or measurable functions and sets. Since there is also a simple connection $L \to R$, the circle line is present in this chapter, too. Chapter 4 is about station T and the connections $T \to D$, $T \to D'$. The results of the two ways to derive a norm are compared, and the theory is developed further for the Bourbaki method. As in the R-case, there are approximation results for integrable or measurable functions. Moreover, the important Riesz representation theorem is discussed, which establishes a connection between T and R.

The final two chapters are not related to the circle line. Chapter 5 is about signed measures. The main result is the Radon-Nikodym theorem; an important application of this theorem is the identification of continuous linear functionals on L^p $(1 \le p < \infty)$. Classical questions about the relationship between differentiation and integration on $\mathbb{R}$ are treated here, and something is said about change of variables in integrals. Product spaces are the subject of Chapter 6. First appears Stone's version of the Fubini and Tonelli theorems, which is then specialized for the $\mathcal{R}$-case and the $\mathcal{T}$-case. There is a section on Fourier theory in $L^2(\mathbb{R})$ to show how useful the Fubini theorems are. The final section contains an application of the Radon-Nikodym theorem together with product measures to a question about stochastic processes. (The reader need not know anything about stochastic processes, though.)

Much of the contents of this book is standard, and can be found in many other textbooks as well (except, perhaps, the section about the relation between differentiation and integration, and Section 6.5). It is the presentation of the material, which allows us to describe and compare the various approaches to Lebesgue integration, that distinguishes this work from many other books. The central idea of relating the various approaches to integration in the circle line is due to N.G. de Bruijn. Lecture notes by him were the starting point for this book.

NOTATION

$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ have their usual meaning, denoting the sets of natural numbers, integers, rational numbers, real numbers and complex numbers, respectively. The symbol $\mathbb{R}^*$ denotes the set of extended real numbers, that is $\mathbb{R}$ together with ∞ and $-\infty$, while $\mathbb{R}^+$ and $\mathbb{R}^*_+$ denote the non-negative members of $\mathbb{R}$ and $\mathbb{R}^*$, respectively.

":=" means "is defined by". For instance, $p := x^2$ defines p as the square of x.

Let S and T be non-empty sets, and let $f : S \to T$ be a function. For any subset B of T the inverse image of B under f is written as $f^{\leftarrow}(B) := \{s \in S \mid f(s) \in B\}$. If f is an injection (that is, if f is one-to-one), then the inverse function of f, which is defined on the range of f, is also denoted by $f^{\leftarrow}$.

Let S and T be sets. If for each $s \in S$ there is in some way given an $f(s) \in T$, then this defines a function $f : S \to T$. We write $f := \curlyvee_{s \in S} f(s)$, which is read as: f is defined to be that function that takes at S the value f(s). (Note that the range space T need not be mentioned explicitly.) For instance, we may write $s := \curlyvee_{x \in \mathbb{R}^2} x^2$, which defines s on $\mathbb{R}$ as the squaring function, or $\cos := \curlyvee_{x \in \mathbb{R}} \sum_{n=0}^{\infty} (-1)^n x^{2n}/(2n)!$, which defines the cosine function on $\mathbb{R}$.

Composition of functions is denoted by means of $\circ$. Thus $f \circ g := \curlyvee_{s \in S} f(g(s))$, if S , is the domain of g and g maps S into the domain of f.

CHAPTER ZERO
PRELIMINARIES

This chapter contains some preliminary material, most of which will be familiar. Not everything presented is necessary for each of the following chapters. We indicate briefly what the reader should minimally do.

The first section tells what a vector space is, and in particular what a Riesz function space is. Every reader should know the contents of this section. The second section is about topology. Except for the notion of completeness, its contents are not needed before Chapter 3, and the product topology is used only in Section 6.3. Then there is a short section giving the definition of a normed space and of an inner product space; the terminology suffices and specific results are not needed. The next section deals with summation. In a certain sense it exemplifies what happens in Chapter 1, but only the facts about change of order of summation in series are really necessary; these may as well be taken on faith. The final section, about Riemann and Riemann-Stieltjes integration, may be read through quickly. In order to follow our development of integration the knowledge of calculus is almost sufficient. What one further needs is knowledge of the real numbers, and pencil and paper.

0.1. Algebraic preliminaries

Most readers will meet only one new concept in this section, namely that of a Riesz function space in 0.1.6. Since this concept is fundamental in our approach to integration, one should at least read its definition.

0.1.1. In integration theory it is convenient to extend the set $\mathbb{R}$ of real numbers with the symbols ∞ and $-\infty$; the resulting set is denoted by $\mathbb{R}^*$ and called the *extended real number system*. The algebraic operations are partially extended to $\mathbb{R}^*$ in an obvious way, roughly by thinking of ∞ as a very large positive number. For instance, if $x \in \mathbb{R}$, then $x + \infty = \infty + x = \infty$, and if in addition $x > 0$, then $x \cdot \infty = \infty \cdot x = \infty$, while if $x < 0$, then $x \cdot \infty = \infty \cdot x = -\infty$. Moreover, we use the convention $0 \cdot \infty = \infty \cdot 0 = 0 \cdot (-\infty) = (-\infty) \cdot 0 = 0$, and this will turn up quite often in the following. Expressions like $\infty - \infty$ remain undefined, because there seems to be no way to handle them consistently. The notions of order, and of supremum and infimum are similarly extended. For instance, if S is a subset of $\mathbb{R}$ which is not bounded above (so there is no $a \in \mathbb{R}$ such that $s \le a$ for all $s \in S$), then we write $\sup S = \infty$.

0.1.2. A *linear space* (or a *vector space*) V *over* $\mathbb{R}$ is a non-empty set V supplied with two operations, called *addition* and *scalar multiplication*. The members of $\mathbb{R}$ are called *scalars*, and $\mathbb{R}$ is the *scalar field* of V. Addition is an operation that takes an element $(x,y) \in V \times V$ into the *sum* $x + y \in V$, while scalar multiplication takes an element $(\alpha,x) \in \mathbb{R} \times V$ into the *scalar multiple* $\alpha x \in V$. These operations are assumed to satisfy the following conditions. If $x,y,z \in V$ and $\alpha,\beta \in \mathbb{R}$, then

$$x + y = y + x \,, \qquad (\alpha + \beta)x = \alpha x + \beta x \,,$$

$$(x + y) + z = x + (y + z) \,, \qquad \alpha(x + y) = \alpha x + \alpha y \,,$$

$$\alpha(\beta x) = (\alpha\beta)x \,, \qquad 1 \cdot x = x \,,$$

while there is a unique element $0 \in V$ such that $u + 0 = u$ for all $u \in V$. This element 0, which behaves neutrally with respect to addition, is called the zero element of V. It follows that $-x$, which is an abbreviation for $-1 \cdot x$, satisfies $x + (-x) = 0$, and that $0 \cdot x = 0$, whenever $x \in V$. All this looks a little bit formal, but it is just the natural generalization of the way we operate with vectors in vector calculus.

What we have defined is commonly called a real vector space, because the scalars are taken from $\mathbb{R}$. In one or two places in the following we shall have occasion to use complex vector spaces, where $\mathbb{C}$ acts as the scalar field, that is, where multiplication by complex numbers is allowed. The formal definition of a complex vector space is the same as that for a real vector space, except for the larger scalar field.

0.1.3. Let V be a vector space (real or complex). A *linear subspace* W of V is a non-empty subset W of V such that $x + y \in W$ and $\alpha x \in W$ for all $x,y \in W$ and all scalars α. Obviously, W is a vector space in its own right with the operations inherited from V.

0.1.4. The nicest functions on a linear space are the linear functions. Let V and W be vector spaces (both real or both complex). A mapping $f : V \to W$ is called *linear* if $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ for all $x,y \in V$ and all scalars α,β. Since the field of scalars is itself a vector space, it may be taken in the role of W, and in this particular case f is called a *linear functional*.

0.1.5. Most vector spaces that we shall consider are of a special kind, which we now describe. Let S be a non-empty set, and let W be a vector space (real or complex). Denote the scalar field of W by K. Let f and g be two functions from S into W. The *sum function* $f + g$ is defined by

$$f + g := \forall_{s \in S}(f(s) + g(s)) \,,$$

and if $\alpha \in K$, then αf is defined by

$$\alpha f := \Upsilon_{s \in S}\ \alpha f(s)\ .$$

If V is a non-empty set of functions from S into W such that $f + g \in V$, $\alpha f \in V$ whenever $f,g \in V$ and $\alpha \in K$, then V is called a *function space*. It is a little bit clumsy, but not difficult, to check that with these operations V is a vector space over K. The zero element of V is the function that is identically zero on S.

0.1.6. The most important function spaces in the sequel consist of real-valued functions, and have additional structure. Once again, let S be a non-empty set. If f and g are real-valued functions on S, then the functions $\sup(f,g)$, $\inf(f,g)$, $|f|$ are defined by

$$\sup(f,g) := \Upsilon_{s \in S}\ \sup(f(s),g(s))\ ,$$

$$\inf(f,g) := \Upsilon_{s \in S}\ \inf(f(s),g(s))\ ,$$

$$|f| := \Upsilon_{s \in S} |f(s)|\ .$$

A function space V over $\mathbb{R}$ is called a *Riesz function space* if $\sup(f,g) \in V$ and $\inf(f,g) \in V$ for every $f,g \in V$.

If V is a function space over $\mathbb{R}$ and $f \in V$, then $|f| \in V$, since $|f| = \sup(f,-f)$. Conversely, if V is a function space over $\mathbb{R}$, and $|f| \in V$ for every $f \in V$, then V is a Riesz function space, as one sees by noting that

$$\sup(a,b) = \tfrac{1}{2}(a + b) + \tfrac{1}{2}|a - b|\ ,$$

$$\inf(a,b) = \tfrac{1}{2}(a + b) - \tfrac{1}{2}|a - b|$$

for all $a,b \in \mathbb{R}$.

If f and g are elements of a Riesz function space V, then we write $f \geq g$ (also $g \leq f$) if $\sup(f,g) = f$. For any $W \subset V$ we put $W^+ := \{f \in W \mid f \geq 0\}$. In particular, $\mathbb{R}^+ := \{x \in \mathbb{R} \mid x \geq 0\}$. A linear functional I defined on V is called *positive* if $I(f) \geq 0$ for all $f \in V^+$. A subset W of the Riesz function space V is said to be *directed* if for every $f,g \in W$ there is an $h \in W$ such that $h \geq \sup(f,g)$. Obviously, if $\{f_1,f_2,\ldots,f_n\}$ is a finite subset of the directed set W, then there exists $h \in W$ such that $h \geq \sup(f_1,f_2,\ldots,f_n)$.

Exercises Section 0.1

1. Let B be a Riesz function space consisting of functions on the set S. Let I be a linear functional on B such that

$$\sup\{|I(\psi)| \mid 0 \le \psi \le \varphi\} < \infty$$

for $\varphi \in B^+$. Show that I can be decomposed as $I = I_+ - I_-$, where I_+ and I_- are positive linear functionals on B. Use the following steps.

(i) For $\varphi \in B^+$ put

$$I_+(\varphi) := \sup\{I(\psi) \mid 0 \le \psi \le \varphi\} ,$$

$$I_-(\varphi) := \sup\{-I(\psi) \mid 0 \le \psi \le \varphi\} .$$

Show that I_+ is a non-negative function on B^+ satisfying $I_+(\alpha\varphi_1 + \beta\varphi_2) = \alpha I_+(\varphi_1) + \beta I_+(\varphi_2)$ for $\alpha,\beta \ge 0$, $\varphi_1,\varphi_2 \in B^+$, and that $I_+(\varphi_1) \ge I_+(\varphi_2)$ if $\varphi_1 \ge \varphi_2$ and $\varphi_1,\varphi_2 \in B^+$. Ditto for I_-.

(ii) Show that $I(\varphi) = I_+(\varphi) - I_-(\varphi)$ for $\varphi \in B^+$. (Hint. Use that $I(\varphi) + I_-(\varphi) = \sup\{I(\varphi - \psi) \mid 0 \le \psi \le \varphi\}$.)

(iii) For $\varphi \in B$ put $\varphi_+ := \sup(\varphi,0)$, $\varphi_- := \sup(-\varphi,0)$, $I_+(\varphi) := I_+(\varphi_+) - I_+(\varphi_-)$, $I_-(\varphi) := I_-(\varphi_+) - I_-(\varphi_-)$. Show that both I_+ and I_- are positive linear functionals on B and that $I(\varphi) = I_+(\varphi) - I_-(\varphi)$ for $\varphi \in B$. (Hint. Show first that $I_+(\varphi_1) - I_+(\varphi_2) = I_+(\varphi_3) - I_+(\varphi_4)$ if $\varphi_1,\varphi_2,\varphi_3,\varphi_4 \in B^+$ and $\varphi_1 - \varphi_2 = \varphi_3 - \varphi_4$.)

2. Let B, I, I_+ and I_- be as in the preceding exercise. Show that the decomposition $I = I_+ - I_-$ has the following extremal property. If $I = I'_+ - I'_-$ for positive linear functionals I'_+ and I'_-, then $I_+(\varphi) \le I'_+(\varphi)$, $I_-(\varphi) \le I'_-(\varphi)$ for all $\varphi \in B^+$.

0.2. Topological preliminaries

This section contains what we need from general topology. The simpler proofs have been left out, but the three results of fundamental importance (Dini's theorem on uniform convergence in 0.2.12, Urysohn's separation lemma in 0.2.16, and the theorem on the existence of partitions of unity in 0.2.17) are proved in full.

0.2.1. A *metric space* (S,d) is a non-empty set S and a function $d : S \times S \to \mathbb{R}^+$ such that $d(x,y) = d(y,x)$, $d(x,y) \le d(x,z) + d(z,y)$ for all $x,y,z \in S$, and such that $d(x,y) = 0$ if and only if $x = y$. The function d is called a *metric* on S. Sometimes we speak of the metric space S, with metric d. In a metric space (S,d) one has a notion of convergence of sequences. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in S, let $x \in S$ and assume that $\lim_{n\to\infty} d(x_n,x) = 0$. Then the sequence $(x_n)_{n\in\mathbb{N}}$ is said to be *convergent to* x (it *tends to* x or *converges to* x), and that x is *limit* of the sequence. Notation: $x = \lim_{n\to\infty} x_n$, or $x_n \to x$ $(n \to \infty)$, or just $x_n \to x$. It is easy to see that in a metric space a sequence has at most one limit, so we can speak about *the* limit of a convergent sequence.

If $x \in S$ and $\delta > 0$, then the *open ball* with *center* x and *radius* δ is the set $B_{x,\delta} := \{y \in S \mid d(x,y) < \delta\}$. It is easy to prove that $x = \lim_{n\to\infty} x_n$ if and only if for every $\delta > 0$ one has $x_n \in B_{x,\delta}$ for all but finitely many $n \in \mathbb{N}$. A subset of S is called *bounded* if it is contained in some open ball.

A sequence $(x_n)_{n\in\mathbb{N}}$ is called a *fundamental sequence* or a *Cauchy sequence* if $d(x_n,x_m) \to 0$ $(n \to \infty, m \to \infty)$; that is, if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_n,x_m) < \varepsilon$ whenever $n > N$, $m > N$. A convergent sequence is a fundamental sequence, but a fundamental sequence need not converge. If in a metric space every fundamental sequence is convergent, then it is called *complete*. Important examples of complete metric spaces are the spaces $\mathbb{R}^n$ (where $n \in \mathbb{N}$) with the usual metric d given by $d(x,y) := (\sum_{i=1}^n |x_i - y_i|^2)^{\frac{1}{2}}$ for $x = (x_1,\dots,x_n) \in \mathbb{R}^n$, $y = (y_1,\dots,y_n) \in \mathbb{R}^n$.

0.2.2. A *topological space* (S,T) is a non-empty set S with a collection T of subsets of S satisfying the following conditions:

(i) $\emptyset \in T$, $S \in T$.

(ii) If $U,V \in T$, then $U \cap V \in T$.

(iii) If $\{U_\alpha \mid \alpha \in A\}$ is a collection of members of T, then $\bigcup_{\alpha\in A} U_\alpha \in T$.

The members of T are called *open* sets, and T is called a *topology* for S. The complements in S of open sets are called *closed* sets. The conditions for a topology imply that the intersection of any collection of closed sets is closed, and that the union of finitely many closed sets is closed.

If A is a subset of the topological space S, then its *closure* $\bar{A}$ is the smallest closed set containing A; that is, $\bar{A}$ is the intersection of the closed sets that contain A. The *interior* A° of A is the largest open set contained in A, or also, A° is the union of the open sets contained in A. A *neighborhood* U of a point $x \in S$ is a subset of S such that $x \in U^\circ$. A subset A of S is called *dense in* S if $\bar{A} = S$. A topological space is called *separable* if it has a countable dense subset.

In a topological space (S,T) convergence of sequences is defined as follows. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in S and x a point of S. We say that the sequence $(x_n)_{n\in\mathbb{N}}$ is *convergent* (*converges* or *tends*) *to* x (with respect to the topology T) if for every neighborhood U of x we have $x_n \in U$ for all but finitely many $n \in \mathbb{N}$. Again, x is called *limit* of the sequence. (It may occur that a sequence has more than one limit, see Exercise 0.2.8(ii).)

0.2.3. Any metric space is also a topological space. For let (S,d) be a metric space. Call a subset O of S open if for every $x \in O$ there exists an open ball $B_{x,\delta} \subset O$ (the δ may of course depend on x). It is now not difficult to see that the collection of open sets thus defined is a topology for S; it is called the *metric topology*. The open balls are open sets for the topology.

Different metrics can give rise to the same topology. For example, the usual metric for $\mathbb{R}^n$ generates what we call the usual topology for $\mathbb{R}^n$, but this topology is also generated by the metric $d'(x,y) := \max\{|x_k - y_k| \mid 1 \le k \le n\}$, or by $d''(x,y) := \sum_{k=1}^{n} |x_k - y_k|$.

In a metric space convergence of a sequence in the metric sense is equivalent to convergence in the topological sense.

0.2.4. Let S be a set, and let $\mathcal{B}$ be a class of subsets of S. If T is the class of all unions of elements of $\mathcal{B}$, we say that $\mathcal{B}$ is a *base* for T, and that $\mathcal{B}$ *generates* T. It is not difficult to show that $\mathcal{B}$ generates a topology of S if and only if the following conditions are satisfied:

(i) If $A \in \mathcal{B}$, $B \in \mathcal{B}$, and $x \in A \cap B$, then there is a $C \in \mathcal{B}$ with $x \in C \subset A \cap B$.

(ii) $S = \bigcup_{A \in \mathcal{B}} A$.

In a metric space the collection of open balls is a base for the metric topology.

A topological space is said to satisfy the *second axiom of countability* if its topology has a countable base.

0.2.5. If (S,T) is a topological space, and A is a non-empty subset of S, there is a natural topology on A: just take $\{A \cap U \mid U \in T\}$ as the collection of open sets. This topology is called the *relative topology* induced on A; it is denoted by $T|A$.

0.2.6. A topological space (S,T) is called a *Hausdorff space* if for any two distinct points $x \in S$, $y \in S$, there exist neighborhoods U of x and V of y such that $U \cap V = \emptyset$. In a Hausdorff space a sequence has at most one limit. A metric space is a Hausdorff space.

0.2.7. Let A be a subset of the topological space (S,T). An *open covering* of A is a collection of open sets whose union contains A. We call A *compact* if every open covering of A contains a finite *subcovering* (a finite subclass of the covering whose union still contains A).

In $\mathbb{R}^n$ with the usual topology we have the important theorem of Heine-Borel: a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded (see Exercise 0.2-10).

0.2.8. Proposition. Let (S,T) be a topological space, C a compact subset of S, and F a closed subset of S. Then $C \cap F$ is compact.

Proof. Exercise 0.2-9. □

0.2.9. Compact subsets of a topological space behave more or less like finite sets. An example of this phenomenon is the following result, which will be needed in the proof of Urysohn's lemma.

Proposition. Let (S,T) be a Hausdorff space, C and D disjoint compact subsets of S. Then there exist disjoint open sets U and V in S with $C \subset U$, $D \subset V$.

Proof. First assume that D consists of one point only, x say. For every $y \in C$ there exist disjoint open sets U_y and V_y such that $y \in U_y$, $x \in V_y$. Now $Q := \{U_y \mid y \in C\}$ is an open covering of the compact set C. Hence Q contains a finite subcovering $\{U_y \mid y \in E\}$, where E is a finite subset of C. Now $U := \bigcup_{y \in E} U_y$ and $V := \bigcap_{y \in E} V_y$ are open sets that are clearly disjoint, and they cover C and $\{x\}$, respectively.

To handle the general case, apply the result just proved as follows. For every $x \in D$ there exist disjoint open sets U_x and V_x with $C \subset U_x$, $x \in V_x$. Now $\{V_x \mid x \in D\}$ is an open covering of the compact set D, which therefore contains a finite subcovering $\{V_x \mid x \in F\}$ where F is a finite subset of D. The sets $U := \bigcap_{x \in F} U_x$ and $V := \bigcup_{x \in F} V_x$ satisfy the conditions. □

0.2.10. It is worth noting that the first part of the preceding proof shows that in a Hausdorff space compact sets are closed.

0.2.11. Let (S,T) and (S',T') be topological spaces. A mapping $\varphi : S \to S'$ is called *continuous* if $\overleftarrow{\varphi}(U) \in T$ for every $U \in T'$. If (S'',T'') is a third topological space, and $\varphi : S \to S'$ and $\psi : S' \to S''$ are both continuous, then the composite function $\psi \circ \varphi$ is continuous.

If the topology T on S is generated by a metric d, then continuity of $\varphi : S \to S'$, where S' is a topological space, is equivalent to the following condition: for every $x \in S$ and every sequence $(x_n)_{n \in \mathbb{N}}$ in S with $x_n \to x$, one has $\varphi(x_n) \to \varphi(x)$.

Let (S,T) be a topological space, and let $\varphi : S \to \mathbb{R}$. The *support* of φ is the closure of $\{s \in S \mid \varphi(s) \neq 0\}$; we denote it by supp φ. We say that φ has *compact support* if supp φ is compact.

0.2.12. The next result is known as Dini's theorem.

Theorem (Dini's theorem). Let (S,T) be a topological space. Let $(\varphi_n)_{n \in \mathbb{N}}$ be a sequence of continuous real-valued functions on S that decreases to zero pointwise,