Measure, Lebesgue Integrals, and Hilbert Space

A. N. KOLMOGOROV AND S. V. FOMIN

Measure, Lebesgue Integrals, and Hilbert Space

A. N. KOLMOGOROV AND S. V. FOMIN
Moscow State University
Moscow, U.S.S.R.

TRANSLATED BY

NATASCHA ARTIN BRUNSWICK
and
ALAN JEFFREY
Institute of Mathematical Sciences
New York University
New York, New York

ACADEMIC PRESS New York and London 196

ACADEMIC PRESS INC.
111 FIFTH AVENUE
NEW YORK 3, N. Y.

United Kingdom Edition
Published by
ACADEMIC PRESS INC. (LONDON) LTD.
BERKELEY SQUARE HOUSE, LONDON W. 1

Library of Congress Catalog Card Number 61-12279

First Printing, 1960
Second Printing, 1962

PRINTED IN THE UNITED STATES OF AMERICA

Translator's Note

This book is a translation of A. N. Kolmogorov and S. V. Fomin's book "Elementy Teorii Funktsii i Funktsional'nogo Analiza., II. Mera, Integral Lebega i Prostranstvo Hilberta" (1960).

An English translation of the first part of this work was prepared by Leo F. Boron and published by Graylock Press in 1957 and is mentioned in [A] of the suggested reading matter added at the end of the present book.

The approach adopted by the Russian authors should be of great interest to many students since the concept of a semiring is introduced early on in the book and is made to play a fundamental role in the subsequent development of the notions of measure and integral. Of particular value to the student is the initial chapter in which all the ideas of measure are introduced in a geometrical way in terms of simple rectangles in the unit square. Subsequently the concept of measure is introduced in complete generality, but frequent back references to the simpler introduction do much to clarify the more sophisticated treatment of later chapters.

A number of errors and inadequacies of treatment noted by the Russian authors in their first volume are listed at the back of their second book and have been incorporated into our translation. The only change we have made in this addenda is to re-reference it in terms of the English Translation [A].

In this edition the chapters and sections have been renumbered to make them independent of the numbering of the first part of the book and to emphasize the self-contained character of the work.

New York, January 1961

NATASCHA ARTIN BRUNSWICK
ALAN JEFFREY

Foreword

This publication is the second book of the "Elements of the Theory of Functions and Functional Analysis," the first book of which ("Metric and Normed Spaces") appeared in 1954. In this second book the main role is played by measure theory and the Lebesgue integral. These concepts, in particular the concept of measure, are discussed with a sufficient degree of generality; however, for greater clarity we start with the concept of a Lebesgue measure for plane sets. If the reader so desires he can, having read §1, proceed immediately to Chapter II and then to the Lebesgue integral, taking as the measure, with respect to which the integral is being taken, the usual Lebesgue measure on the line or on the plane.

The theory of measure and of the Lebesgue integral as set forth in this book is based on lectures by A. N. Kolmogorov given by him repeatedly in the Mechanics-Mathematics Faculty of the Moscow State University. The final preparation of the text for publication was carried out by S. V. Fomin.

The two books correspond to the program of the course "Analysis III" which was given for the mathematics students by A. N. Kolmogorov.

At the end of this volume the reader will find corrections pertaining to the text of the first volume.

A. N. Kolmogorov
S. V. Fomin

Contents

CHAPTER I

Measure Theory

CHAPTER II

Measurable Functions

CHAPTER III

The Lebesgue Integral

CHAPTER IV

Functions Which Are Square Integrable

CHAPTER V

The Abstract Hilbert Space. Integral Equations with a Symmetric Kernel

List of Symbols

A	set
A^*	linear operator adjoint to (the linear operator) A, p. 129.
$A \subset B$	B is a proper subset of A.
$A \subseteq B$	A is a subset of B.
$a \in A$	element a is a member of set A.
$A \setminus B$	difference, complement of B with respect to A
$A \triangle B$	symmetric difference of sets A and B defined by the expression $A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$.
$B(f, g)$	bilinear functional, p. 128.
$\mathfrak{B}(\mathfrak{S})$	Borel algebra containing $\mathfrak{S}$, p. 25.
δ-*ring*	p. 24
δ_{ik}	Kronecker delta, p. 107.
ε	arbitrary positive number.
$\varnothing$	null set
$\|f\|$	norm of f, p. 95.
(f, g)	scalar product of functions in L_2, p. 94.
$\lambda(A)$	continuation of measure m defined for A, p. 39.
$\bar{M}$	closure of the set M, p. 105.
$M \oplus M'$	direct sum of spaces M and M', p. 124.
$\mathfrak{M}(A)$	system of sets A, p. 20.
$m(P)$	measure of rectangular set P, p. 2.
$\mu(A)$	Lebesgue measure of set A, p. 7.
$\mu^*(A)$	outer measure of set A, p. 6.
$\mu_*(A)$	inner measure of set A, p. 6.

$\bar{\mu}(A)$	outer measure of set A, p. 31.
$\underline{\mu}(A)$	inner measure of set A, p. 31.
$\mathfrak{N}$	system of sets, p. 25.
$\mathfrak{P}$	system of sets, p. 20.
$\boxed{\times} X_k$	direct product of sets $X_1, X_2, \cdots, X_n$, p. 78.
$Q(f)$	quadratic form, p. 129.
$\mathfrak{R}$	ring, p. 19.
$\mathfrak{S}$	system of sets, p. 19.
σ-ring	p. 24
$\{x:f(x)<c\}$	set of elements x with property $f(x)<c$, p. 49.
$\bigcup, \cup$	union
$\bigcap, \cap$	intersection, product.
$\{x_1, x_2, \cdots\}$	set of elements $x_1, x_2, \cdots$

CHAPTER I

MEASURE THEORY

The concept of the measure $\mu(A)$ of a set A is a natural generalisation of the following concepts:

1) the length $l(\Delta)$ of a segment Δ;
2) the area $S(F)$ of a plane figure F;
3) the volume $V(G)$ of a figure G in space;
4) the increment $\varphi(b) - \varphi(a)$ of a non-decreasing function $\varphi(t)$ on the half open line interval $[a, b)$;
5) the integral of a non-negative function taken over some line, plane surface or space region, etc.

The concept of the measure of a set, which originated in the theory of functions of a real variable, has subsequently found applications in probability theory, the theory of dynamic systems, functional analysis and other fields of mathematics.

In the first section of this chapter we shall introduce the notion of measure for sets in a plane, starting from the concept of the area of a rectangle. The general theory of measures will be given in §§3–7. The reader will however easily notice that all arguments used in §1 are of a general character and can be re-phrased for the abstract theory without essential changes.

1. Measure of Plane Sets

Let us consider a system $\mathfrak{S}$ of sets in the (x, y)-plane, each of which is given by one of the inequalities of the form

$$a \leq x \leq b,$$
$$a < x \leq b,$$
$$a \leq x < b,$$
$$a < x < b,$$

together with one of the inequalities

$$c \leq y \leq d,$$

$$c < y \leq d,$$

$$c \leq y < d,$$

$$c < y < d,$$

where a, b, c and d are arbitrary numbers. The sets belonging to this system we shall call "rectangles". The closed rectangle, given by the inequalities

$$a \leq x \leq b, \qquad c \leq y \leq d$$

is the usual rectangle (including the boundary) if $a < b$ and $c < d$, or a segment if $a = b$ and $c < d$, or a point if $a = b$ and $c = d$. The open rectangle

$$a < x < b, \qquad c < y < d$$

is, depending on the relations among a, b, c and d, respectively, a rectangle without boundaries, or the empty set. Each of the rectangles of the other types (let us call them half-open) is either a real rectangle without one, two or three sides, or an interval, or a half-interval, or, finally, an empty set.

We shall define for each of the rectangles a measure corresponding to the concept of area, well known from elementary geometry, in the following way:

a) the measure of an empty set is equal to zero;

b) the measure of a non-empty rectangle (closed, open or half-open) given by the numbers a, b, c and d is equal to

$$(b - a)(d - c).$$

In this way, to each rectangle P there corresponds a number $m(P)$—the measure of this rectangle; moreover the following conditions are obviously satisfied:

1) the measure $m(P)$ takes on real non-negative values;

2) the measure is additive, i.e., if $P = \bigcup_{k=1}^{n} P_k$ and $P_i \cap P_k = \varnothing$

for $i \neq k$, then

$$m(P) = \sum_{k=1}^{n} m(P_k).$$

Our task now is to generalise the measure $m(P)$, which up to now has been defined only for rectangles, to a wider class of measures, preserving the properties 1) and 2).

The first step in this direction consists of generalising the concept of measure to so called elementary sets. We shall call a plane set *elementary* if it can be represented, at least in one way, as a union of a finite number of pairwise non-intersecting rectangles.

For what follows we shall use the following

Theorem 1. *The union, intersection, difference and symmetric difference of two elementary sets is also an elementary set.*

Proof. It is clear that the intersection of two rectangles is again a rectangle. Therefore, if

$$A = \bigcup_k P_k \quad \text{and} \quad B = \bigcup_j Q_j$$

are two elementary sets, then

$$A \cap B = \bigcup_{k,j} (P_k \cap Q_j)$$

is also an elementary set.

The difference of two rectangles is, as is easily checked, an elementary set. Hence, taking away from the rectangle some elementary set we again obtain an elementary set (as the intersection of elementary sets). Now let A and B be two elementary sets. One can obviously find a rectangle P containing each of them. Then

$$A \cup B = P \setminus [(P \setminus A) \cap (P \setminus B)]$$

is an elementary set by what has been said above. From this and the equalities

$$A \setminus B = A \cap (P \setminus B)$$

and

$$A \triangle B = (A \cup B) \setminus (A \cap B)$$

it follows that the difference and the symmetric difference of elementary sets are also elementary sets. The theorem is thus proved.

Let us now define the measure $m'(A)$ for elementary sets in the following way: if

$$A = \bigcup P_k,$$

where the P_k are pairwise non-intersecting rectangles, then

$$m'(A) = \sum m(P_k).$$

Let us show that $m'(A)$ does not depend on the way the set A is represented as a sum of rectangles. Let

$$A = \bigcup_k P_k = \bigcup_j Q_j,$$

where the P_k and the Q_j are rectangles and $P_i \cap P_k = \varnothing$, $Q_i \cap Q_k = \varnothing$ for $i \neq k$. Since the intersection of two rectangles is a rectangle we have, because the measures of rectangles are additive,

$$\sum_k m(P_k) = \sum_{k,j} m(P_k \cap Q_j) = \sum_j m(Q_j).$$

It is easy to see that the measure of elementary sets defined in this way is non-negative and additive.

Let us establish the following property of elementary measures which will be important in what follows.

Theorem 2. *If A is an elementary set and $\{A_n\}$ is a finite or countable system of elementary sets such that*

$$A \subseteq \bigcup_n A_n,$$

then

$$m'(A) \leq \sum_n m'(A_n). \qquad (*)$$

Proof. For an arbitrary $\varepsilon > 0$ and a given A we can obviously find a closed elementary set $\bar{A}$, which is contained in A and satisfies the conditions

$$m'(\bar{A}) \geq m'(A) - \frac{\varepsilon}{2}.$$

(For this it is sufficient to replace each of the k rectangles P_i which form A by the closed rectangle which is wholly contained in it and having an area larger than $m(P_i) \; - \varepsilon/2^{k+1}$.)

Furthermore, for each A_n one can find an open elementary set $\tilde{A}_n$ containing A_n and satisfying the condition

$$m'(\tilde{A}_n) \leq m'(A_n) + \frac{\varepsilon}{2^{n+1}}.$$

It is clear that

$$\bar{A} \subseteq \bigcup_n \tilde{A}_n.$$

From $\{\tilde{A}_n\}$ we can (by the Borel-Lebesgue lemma) select a system $A_{n_1}, A_{n_2}, \cdots, A_{n_s}$ which covers $\bar{A}$. Here moreover, obviously

$$m'(\bar{A}) \leq \sum_{i=1}^{s} m'(\tilde{A}_{n_i})$$

(since otherwise $\bar{A}$ could be covered by a finite number of rectangles having an added area which is less than $m'(\bar{A})$, which is obviously impossible). Therefore,

$$m'(A) \leq m'(\bar{A}) + \frac{\varepsilon}{2} \leq \sum_{i=1}^{s} m'(\tilde{A}_{n_i}) + \frac{\varepsilon}{2} \leq \sum_n m'(\tilde{A}_n) + \frac{\varepsilon}{2}$$

$$\leq \sum_n m'(A_n) + \sum_n \frac{\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2} = \sum_n m'(A_n) + \varepsilon,$$

yielding $(*)$, since ε is arbitrary and positive.

The set of elementary sets does not exhaust all those sets which were considered in elementary geometry and in classical analysis. Therefore it is natural to ask the question how to generalise the concept of a measure to a class of sets which is wider than the finite combinations of rectangles with sides parallel to the coordinate axes, preserving its basic properties.

The final solution of this problem was given by H. Lebesgue at the beginning of the twentieth century.

In presenting the theory of Lebesgue's measure we will have to consider not only finite but also infinite combinations of rectangles.

In order to avoid dealing with infinite values for measures we shall limit ourselves to sets which fully belong to the square $E = \{0 \leq x \leq 1; 0 \leq y \leq 1\}$.

On the set of all these sets we shall define two functions $\mu^*(A)$ and $\mu_*(A)$ in the following way.

Definition 1. We shall call the number

$$\inf_{A \subset \cup P_k} \sum m(P_k)$$

the *outer measure* $\mu^*(A)$ of the set A; the lower bound is taken over all possible coverings of the set A by finite or countable rectangles.

Definition 2. We call the number

$$1 - \mu^*(E \setminus A)$$

the *inner measure* $\mu_*(A)$ of the set A. It is easy to see that always

$$\mu_*(A) \leq \mu^*(A).$$

Indeed, suppose that for some $A \subset E$

$$\mu_*(A) > \mu^*(A),$$

i.e.,

$$\mu^*(A) + \mu^*(E \setminus A) < 1.$$

Then, by definition of the exact lower bound, one can find systems of rectangles $\{P_i\}$ and $\{Q_k\}$ covering A and $E \setminus A$, respectively, such that

$$\sum_i m(P_i) + \sum_k m(Q_k) < 1.$$

The union of the systems $\{P_i\}$ and $\{Q_k\}$ we shall denote by

$\{R_j\}$; we obtain

$$E \subseteq \bigcup_j R_j \quad \text{and} \quad m(E) > \sum_j m(R_j),$$

which contradicts Theorem 2.

Definition 3. We call a set A *measurable* (in the sense of Lebesgue), if

$$\mu_*(A) = \mu^*(A).$$

The common value $\mu(A)$ of the outer and inner measure for a countable set A is called its *Lebesgue measure.*

Let us find the basic properties of the Lebesgue measure and of countable sets.

Theorem 3. *If*

$$A \subseteq \bigcup_n A_n,$$

where A_n is a finite or countable system of sets, then

$$\mu^*(A) \leq \sum_n \mu^*(A_n).$$

Proof. By the definition of the outer measure, we can find for each A_n a system of rectangles $\{P_{nk}\}$, finite or countable, such that $A_n \subseteq \bigcup_k P_{nk}$ and

$$\sum_k m(P_{nk}) \leq \mu^*(A_n) + \frac{\varepsilon}{2^n},$$

where $\varepsilon > 0$ is selected arbitrarily. Then

$$A \subseteq \bigcup_n \bigcup_k P_{nk},$$

and

$$\mu^*(A) \leq \sum_n \sum_k m(P_{nk}) \leq \sum_n \mu^*(A_n) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this establishes the theorem.

We have already introduced above the concept of measure for sets which we called elementary. The theorem below shows that for elementary sets Definition 3 leads to exactly the same result.

Theorem 4. *Elementary sets are measurable, and for them the Lebesgue measure coincides with the measure $m'(A)$ constructed above.*

Proof. If A is an elementary set and $P_1, P_2, \cdots, P_k$ are the pairwise non-intersecting rectangles comprising it, then, by definition,

$$m'(A) = \sum_{i=1}^{k} m(P_i).$$

Since the rectangles P_i cover all of A,

$$\mu^*(A) \leq \sum_i m(P_i) = m'(A).$$

But, if $\{Q_j\}$ is an arbitrary finite or countable system of rectangles covering A, then, by Theorem 2, $m'(A) \leq \sum m(Q_j)$. Hence $m'(A) \leq \mu^*(A)$. Therefore, $m'(A) = \mu^*(A)$.

Since $E \setminus A$ is also an elementary set, $m'(E \setminus A) = \mu^*(E \setminus A)$. But

$$m'(E \setminus A) = 1 - m'(A) \quad \text{and} \quad \mu^*(E \setminus A) = 1 - \mu_*(A),$$

yielding

$$m'(A) = \mu_*(A).$$

Therefore

$$m'(A) = \mu^*(A) = \mu_*(A) = \mu(A).$$

From the result obtained we see that Theorem 2 is a special case of Theorem 3.

Theorem 5. *In order that the set A be measurable, it is necessary and sufficient that the following condition be satisfied: for any $\varepsilon > 0$ there exists an elementary set B, such that,*

$$\mu^*(A \triangle B) < \varepsilon.$$

Thus those sets and only those sets are measurable which can be "approximated with an arbitrary degree of accuracy" by elementary sets. For the proof of Theorem 5 we shall need the following

Lemma. *For two arbitrary sets A and B*

$$|\mu^*(A) - \mu^*(B)| \leq \mu^*(A \triangle B).$$

Proof of the lemma. Since

$$A \subset B \cup (A \triangle B),$$

we have

$$\mu^*(A) \leq \mu^*(B) + \mu^*(A \triangle B).$$

This implies the lemma if $\mu^*(A) \geq \mu^*(B)$. If $\mu^*(A) \leq \mu^*(B)$, then the lemma follows from the inequality

$$\mu^*(B) \leq \mu^*(A) + \mu^*(A \triangle B),$$

which can be established in an analagous manner.

Proof of Theorem 5. Sufficiency. Let us assume that for any $\varepsilon > 0$ there exists an elementary set B, such that

$$\mu^*(A \triangle B) < \varepsilon.$$

Then

$$|\mu^*(A) - m'(B)| = |\mu^*(A) - \mu^*(B)| < \varepsilon, \tag{1}$$

and since

$$(E \setminus A) \triangle (E \setminus B) = A \triangle B,$$

we have analogously that

$$|\mu^*(E \setminus A) - m'(E \setminus B)| < \varepsilon. \tag{2}$$

From the inequalities (1) and (2) we have, taking into account

$$m'(B) + m'(E \setminus B) = m'(E) = 1,$$

that

$$|\mu^*(A) + \mu^*(E \setminus A) - 1| < 2\varepsilon,$$

and, since $\varepsilon > 0$ is arbitrary,

$$\mu^*(A) + \mu^*(E \setminus A) = 1,$$

i.e., the set A is measurable.

Necessity. Let A be measurable, i.e., let

$$\mu^*(A) + \mu^*(E \setminus A) = 1.$$

Selecting $\varepsilon > 0$ arbitrarily, we shall seek coverings

$$A \subseteq \bigcup_n B_n \quad \text{and} \quad E \setminus A \subseteq \bigcup_n C_n$$

for the sets A and $E \setminus A$ by systems of rectangles $\{B_n\}$ and $\{C_n\}$ for which

$$\sum_n m(B_n) \leq \mu^*(A) + \frac{\varepsilon}{3} \quad \text{and} \quad \sum_n m(C_n) \leq \mu^*(E \setminus A) + \frac{\varepsilon}{3}.$$

Since $\sum_n m(B_n) < \infty$, we can find an N such that

$$\sum_{n>N} m(B_n) < \frac{\varepsilon}{3}.$$

Set

$$B = \bigcup_{n=1}^{N} B_n.$$

It is clear that the set

$$P = \bigcup_{n>N} B_n$$

contains $A \setminus B$, and that the set

$$Q = \bigcup_n (B \cap C_n)$$

contains $B \setminus A$, and therefore that $A \triangle B \subseteq P \cup Q$. Hence,

$$\mu^*(P) \leq \sum_{n>N} m(B_n) < \frac{\varepsilon}{3}.$$

Let us evaluate $\mu^*(Q)$. For this let us note that

$$\left(\bigcup_n B_n\right) \cup \left(\bigcup_n (C_n \setminus B)\right) = E,$$

and hence

$$\sum_n m(B_n) + \sum_n m'(C_n \setminus B) \geq 1. \tag{3}$$

But by assumption,

$$\sum_n m(B_n) + \sum_n m(C_n) \le \mu^*(A) + \mu^*(E\setminus A) + \frac{2\varepsilon}{3} = 1 + \frac{2\varepsilon}{3}. \quad (4)$$

Subtracting (3) from (4) we have

$$\sum_n m(C_n) - \sum_n m'(C_n\setminus B) = \sum_n m'(C_n \cap B) < \frac{2\varepsilon}{3},$$

i.e.,

$$\mu^*(Q) < \frac{2\varepsilon}{3}.$$

Therefore,

$$\mu^*(A \;\triangle\; B) \le \mu^*(P) + \mu^*(Q) < \varepsilon.$$

Hence, if A is measurable, there exists for any arbitrary $\varepsilon > 0$, an elementary set B such that $\mu^*(A \triangle B) < \varepsilon$. Theorem 5 is thus established.

Theorem 6. *The union and intersection of a finite number of measurable sets are measurable sets.*

Proof. It is clear that it suffices to give the proof for two sets. Let A_1 and A_2 be measurable sets. Then, for any $\varepsilon > 0$, one can find elementary sets B_1 and B_2 such that,

$$\mu^*(A_1 \triangle B_1) < \frac{\varepsilon}{2}, \qquad \mu^*(A_2 \triangle B_2) < \frac{\varepsilon}{2}.$$

Since

$$(A_1 \cup A_2) \triangle (B_1 \cup B_2) \subseteq (A_1 \triangle B_1) \cup (A_2 \triangle B_2),$$

we have

$$\mu^*[(A_1 \cup A_2) \triangle (B_1 \cup B_2)] \le \mu^*(A_1 \triangle B_1) + \mu^*(A_2 \triangle B_2) < \varepsilon. \quad (5)$$

$B_1 \cup B_2$ is an elementary set; hence, by Theorem 4, the set $A_1 \cup A_2$ is measurable. But, just by the definition of measurability, if A is measurable, then $E\setminus A$ is also measurable; hence the fact that the intersection of two sets is measurable follows from the relation

$$A_1 \cap A_2 = E\setminus[(E\setminus A_1) \cup (E\setminus A_2)].$$

Corollary. *The difference and symmetric difference of two measurable sets is measurable.*

This follows from Theorem 6 and the equations

$$A_1 \setminus A_2 = A_1 \cap (E \setminus A_2), \qquad A_1 \,\triangle\, A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1).$$

Theorem 7. *If $A_1, A_2, \cdots, A_n$ are pairwise non-intersecting measurable sets, then*

$$\mu\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mu(A_k).$$

Proof. As in Theorem 6 it suffices to consider the case $n = 2$. Let us select an arbitrary $\varepsilon > 0$ and elementary sets B_1 and B_2 such that

$$\mu^*(A_1 \,\triangle\, B_1) < \varepsilon, \tag{6}$$

$$\mu^*(A_2 \,\triangle\, B_2) < \varepsilon. \tag{7}$$

We set $A = A_1 \cup A_2$ and $B = B_1 \cup B_2$. The set A is measurable by Theorem 6. Since

$$B_1 \cap B_2 \subseteq (A_1 \,\triangle\, B_1) \cup (A_2 \,\triangle\, B_2),$$

we have

$$m'(B_1 \cap B_2) \leq 2\varepsilon. \tag{8}$$

By the lemma to Theorem 5, and from (6) and (7) we have

$$|m'(B_1) - \mu^*(A_1)| < \varepsilon, \tag{9}$$

$$|m'(B_2) - \mu^*(A_2)| < \varepsilon. \tag{10}$$

Since on the set of elementary sets the measure is additive, we obtain from (8), (9) and (10)

$$m'(B) = m'(B_1) + m'(B_2) - m'(B_1 \cap B_2) \geq \mu^*(A_1) + \mu^*(A_2) - 4\varepsilon.$$

Noting moreover that $A \,\triangle\, B \subseteq (A_1 \,\triangle\, B_1) \cup (A_2 \,\triangle\, B_2)$ we finally have

$$\mu^*(A) \geq m'(B) - \mu^*(A \,\triangle\, B) \geq m'(B) - 2\varepsilon$$

$$\geq \mu^*(A_1) + \mu^*(A_2) - 6\varepsilon.$$

Since 6ε can be made arbitrarily small,

$$\mu^*(A) \geq \mu^*(A_1) + \mu^*(A_2).$$

The reverse inequality

$$\mu^*(A) \leq \mu^*(A_1) + \mu^*(A_2)$$

is always true for $A = A_1 \cup A_2$, therefore, we finally obtain

$$\mu^*(A) = \mu^*(A_1) + \mu^*(A_2).$$

Since A_1, A_2 and A are measurable, we can replace μ^* by μ. The theorem is thus established.

Theorem 8. *The union and the intersection of a countable number of measurable sets are measurable sets.*

Proof. Let

$$A_1, A_2, \cdots, A_n, \cdots$$

be a countable system of measurable sets and $A = \bigcup_{n=1}^{\infty} A_n$. Let us set $A'_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k$. It is clear that $A = \bigcup_{n=1}^{\infty} A'_n$, where the sets A'_n are pairwise non-intersecting. By Theorem 6 and its corollary, all the sets A'_n are measurable. According to Theorems 7 and 3, for any n

$$\sum_{k=1}^{n} \mu(A'_k) = \mu\left(\bigcup_{k=1}^{n} A'_k\right) \leq \mu(A),$$

Therefore the series

$$\sum_{n=1}^{\infty} \mu(A'_n)$$

converges and hence for any $\varepsilon > 0$ one can find an N such that

$$\sum_{n>N} \mu(A'_n) < \frac{\varepsilon}{2}. \tag{11}$$

Since the set $C = \bigcup_{n=1}^{N} A'_n$ is measurable (being a union of a finite number of measurable sets), we can find an elementary set B such that

$$\mu^*(C \triangle B) < \frac{\varepsilon}{2}. \tag{12}$$

Since

$$A \triangle B \subseteq (C \triangle B) \cup \left(\bigcup_{n>N} A'_n\right),$$

(11) and (12) yield

$$\mu^*(A \triangle B) < \varepsilon.$$

Because of Theorem 5 this implies that the set A is measurable.

Since the complement of a measurable set is itself measurable, the statement of the theorem concerning intersections follows from the equality

$$\bigcap_n A_n = E \diagdown \bigcup_n (E \diagdown A_n).$$

Theorem 8 is a generalisation of Theorem 6. The following Theorem is a corresponding generalisation of Theorem 7.

Theorem 9. *If $\{A_n\}$ is a sequence of pairwise non-intersecting measurable sets, and $A = \bigcup_n A_n$, then*

$$\mu(A) = \sum_n \mu(A_n).$$

Proof. By Theorem 7, for any N,

$$\mu\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mu(A_n) \leq \mu(A).$$

Taking the limit as $N \to \infty$, we have

$$\mu(A) \geq \sum_{n=1}^{\infty} \mu(A_n). \tag{13}$$

On the other hand, according to Theorem 3,

$$\mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n). \tag{14}$$

Inequalities (13) and (14) yield the assertion of the Theorem.

The property of measures established in Theorem 9 is called its *countable additivity*, or *σ-additivity*. An immediate corollary of σ-additivity is the following property of measures, called *continuity*.

Theorem 10. *If $A_1 \supseteq A_2 \supseteq \cdots$ is a sequence of measurable sets, contained in each other, and $A = \bigcap_n A_n$, then*

$$\mu(A) = \lim_{n\to\infty} \mu(A_n).$$

It suffices, obviously, to consider the case $A = \varnothing$, since the general case can be reduced to it by replacing A_n by $A_n \setminus A$. Then

$$A_1 = (A_1 \setminus A_2) \cup (A_2 \setminus A_3) \cup \cdots$$

and

$$A_n = (A_n \setminus A_{n+1}) \cup (A_{n+1} \setminus A_{n+2}) \cup \cdots.$$

Therefore

$$\mu(A_1) = \sum_{k=1}^{\infty} \mu(A_k \setminus A_{k+1}), \tag{15}$$

and

$$\mu(A_n) = \sum_{k=n}^{\infty} \mu(A_k \setminus A_{k+1}); \tag{16}$$

since the series (15) converges, its remainder term (16) tends to zero as $n \to \infty$. Thus,

$$\mu(A_n) \to 0 \quad \text{for} \quad n \to \infty,$$

which was to be shown.

Corollary. *If $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing sequence of measurable sets, and*

$$A = \bigcup_n A_n,$$

then

$$\mu(A) = \lim_{n\to\infty} \mu(A_n).$$

For the proof it suffices to go over from the sets A_n to their complements and to use Theorem 10.

Thus we have generalised the concept of a measure from elementary sets to a wider class of sets, called measurable sets, which are closed with respect to the operations of countable unions and intersections. The measure constructed is σ-additive on this class of sets.

Let us make a few final remarks.

1. The theorems we have derived allow us to obtain an idea of the set of all Lebesgue measurable sets.

Since every open set belonging to E can be represented as a union of a finite or countable number of open rectangles, i.e., measurable sets, Theorem 8 implies that all open sets are measurable. Closed sets are complements of open sets and consequently are also measurable. According to Theorem 8 also all those sets must be measurable which can be obtained from open or closed sets by a finite or countable number of operations of countable unions and intersections. One can show, however, that these sets do not exhaust the set of all Lebesgue measurable sets.

2. We have considered above only those plane sets which are subsets of the unit square $E = \{0 \le x, y \le 1\}$. It is not difficult to remove this restriction, e.g., by the following method. Representing the whole plane as a sum of squares $E_{nm} = \{n \le x \le n + 1, m \le y \le m + 1\}$ (m, n integers), we shall say that the plane set A is measurable if its intersection $A_{nm} = A \cap E_{nm}$ with each of these squares is measurable, and if the series $\sum_{n,m} \mu(A_{nm})$ converges. Here we assume by definition that

$$\mu(A) = \sum_{n,m} \mu(A_{nm}).$$

All the properties of measures established above can obviously be carried over to this case.

3. In this section we have given the construction of Lebesgue measures for plane sets. Analogously Lebesgue measures may be constructed on a line, in a space of three dimensions or, in general, in a space of n dimensions. In each of these cases the measure is constructed by the same method: proceeding from a measure de-

fined earlier for some system of simple sets (rectangles in the case of a plane, intervals (a, b), segments $[a, b]$ and half-lines $(a, b]$ and $[a, b)$ in the case of a line, etc.) we first define a measure for finite unions of such sets, and then generalise it to a much wider class of sets—to sets which are Lebesgue measurable. The definition of measurability itself can be carried over word for word to sets in spaces of any dimension.

4. Introducing the concept of the Lebesgue measure we started from the usual definition of an area. The analogous construction for one dimension is based on the concept of length of an interval (segment, half-line). One can however introduce measure by a different and more general method.

Let $F(t)$ be some non-decreasing function, continuous from the left, and defined on a line. We set

$$m(a, b) = F(b) - F(a + 0),$$

$$m[a, b] = F(b + 0) - F(a),$$

$$m(a, b] = F(b + 0) - F(a + 0),$$

$$m[a, b) = F(b) - F(a).$$

It is easy to see that the function of the interval m defined in this way is non-negative and additive. Applying to it considerations which are analogous to the ones used in this section, we can construct some measure $\mu_F(A)$. Here the set of sets, which are measurable with respect to the given measure, is closed with respect to the operations of countable unions and intersections, and the measure μ_F is σ-additive. The class of sets, measurable with respect to μ_F, will, generally speaking, depend on the choice of the function F. However, for any choice of F, the open and closed sets, and therefore all their countable unions and intersections will obviously be measurable. Measures obtained from some function F are called Lebesgue-Stieltjes measures. In particular, to the function $F(t) = t$ there corresponds the usual Lebesgue measure on a line.

If the measure μ_F is such that it is equal to zero for any set whose normal Lebesgue measure is zero, then the measure μ_F is called completely continuous. If the measure is wholly concentrated on a finite or countable set of points (this happens when the set of values of the function $F(t)$ is finite or countable), then it is called discrete. The measure μ_F is called singular, if it is equal to zero for any one point set, and if there exists such a set M with Lebesgue measure zero, that the measure μ_F of its complement is equal to zero.

One can show that any measure μ_F can be represented as a sum of an absolutely continuous, a discrete and a singular measure.

• Existence of Non-measurable Sets

It has been shown above that the class of Lebesgue measurable sets is quite wide. Naturally the question arises as to whether there exist any non-measurable sets at all. We shall show that this problem is solved affirmatively. It is easiest to construct non-measurable sets on a circle.

Consider a circle with circumference C of unit length and let α be some irrational number. Let us consider those points of the circumference C which can be transformed into each other by a rotation of the circle through an angle $n\alpha$ (n an integer) as belonging to our class. Each of these classes will obviously consist of a countable number of points. Let us now select from each of these classes one point. We shall show that the set obtained in this manner, let us call it Φ, is non-measurable. Let us call the set, obtained from Φ by a rotation through angle $n\alpha$, Φ_n. It is easily seen that all the sets Φ_n are pairwise nonintersecting, and that they add up to the whole circular arc C. If the set Φ were measurable, then the sets Φ_n, congruent to it, would also be measurable. Since

$$C = \bigcup_{n=-\infty}^{\infty} \Phi_n, \qquad \Phi_n \cap \Phi_m = \varnothing \quad \text{for} \quad n \neq m,$$

this would imply, on the strength of the σ-additivity of measures that,

$$1 = \sum_{n=-\infty}^{\infty} \mu(\Phi_n). \tag{17}$$

But congruent sets must have the same measure:

$$\mu(\Phi_n) = \mu(\Phi).$$

This shows that equation (17) is impossible, since the sum of the series on the right-hand side of equation (17) is equal to zero if $\mu(\Phi) = 0$, and is infinite if $\mu(\Phi) > 0$. Thus the set Φ (and hence also each Φ_n) is non-measurable.

2. Systems of Sets

Before proceeding to the general theory of sets we shall first give some information concerning systems of sets, which supplements the elements of set theory discussed in Chapter I of Volume I.

We shall call any set, the elements of which are again certain sets, a *system of sets*. As a rule we shall consider systems of sets, each of which is a subset of some fixed set X. We shall usually denote systems of sets by Gothic letters. Of basic interest to us will be systems of sets satisfying, with respect to the operations introduced in Chapter I, §1 of Volume I, certain definite conditions of closure.

Definition 1. A non-empty system of sets $\mathfrak{R}$ is called a *ring* if it satisfies the conditions that $A \in \mathfrak{R}$ and $B \in \mathfrak{R}$ implies that the sets $A \triangle B$ and $A \cap B$ belong to $\mathfrak{R}$.

For any A and B

$$A \cup B = (A \triangle B) \triangle (A \cap B),$$

and

$$A \setminus B = A \triangle (A \cap B);$$

hence $A \in \mathfrak{R}$ and $B \in \mathfrak{R}$ also imply that the sets $A \cup B$ and $A \setminus B$ belong to $\mathfrak{R}$. Thus a ring of sets is a system which is invariant with respect to the operations of union and intersection, subtraction and the formation of a symmetric difference. Obviously the ring is also invariant with respect to the formation of any finite number of unions and intersections of the form

$$C = \bigcup_{k=1}^{n} A_k, \qquad D = \bigcap_{k=1}^{n} A_k.$$

Any ring contains the empty set $\varnothing$, since always $A \setminus A = \varnothing$. The system consisting of only the empty set is the smallest possible ring of sets.

A set E is called the *unit* of the system of sets $\mathfrak{S}$, if it belongs to $\mathfrak{S}$ and if, for any $A \in \mathfrak{S}$, the equation

$$A \cap E = A$$

holds.

Hence the unit of the system of sets $\mathfrak{S}$ is simply the maximal set of this system, containing all other sets which belong to $\mathfrak{S}$.

A ring of sets with a unit is called an *algebra* of sets.

Examples. 1) For any set A the system $\mathfrak{M}(A)$ of all its subsets is an algebra of sets with the unit $E = A$.

2) For any non-empty set A the system $\{\varnothing, A\}$ consisting of the set A and the empty set $\varnothing$, forms an algebra of sets with the unit $E = A$.

3) The system of all finite subsets of an arbitrary set A forms a ring of sets. This ring is then and only then an algebra if the set A itself is finite.

4) The system of all bounded subsets of the line forms a ring of sets which does not contain a unit.

From the definition of a ring of sets there immediately follows

Theorem 1. *The intersection* $\mathfrak{R} = \bigcap_\alpha \mathfrak{R}_\alpha$ *of any set of rings is also a ring.*

Let us establish the following simple result which will however be important in the subsequent work.

Theorem 2. *For any non-empty system of sets* $\mathfrak{S}$ *there exists one and only one ring* $\mathfrak{R}(\mathfrak{S})$, *containing* $\mathfrak{S}$ *and contained in an arbitrary ring* $\mathfrak{R}$ *which contains* $\mathfrak{S}$.

Proof. It is easy to see that the ring $\mathfrak{R}(\mathfrak{S})$ is uniquely defined by the system $\mathfrak{S}$. To prove its existence let us consider the union $X = \bigcup_{A \in \mathfrak{S}} A$ of all sets A contained in $\mathfrak{S}$ and the ring $\mathfrak{M}(X)$ of all subsets of the set X. Let Σ be the set of all rings of sets contained in $\mathfrak{M}(X)$ and containing $\mathfrak{S}$. The intersection

$$\mathfrak{P} = \bigcap_{\mathfrak{R} \in \Sigma} \mathfrak{R}$$

of all these rings will obviously be the required ring $\mathfrak{R}(\mathfrak{S})$.

Indeed, whatever the $\mathfrak{R}^*$ containing $\mathfrak{S}$, the intersection $\mathfrak{R} = \mathfrak{R}^* \cap \mathfrak{M}(X)$ is a ring of Σ and, therefore,

$$\mathfrak{S} \subseteq \mathfrak{P} \subseteq \mathfrak{R} \subseteq \mathfrak{R}^*,$$

i.e., $\mathfrak{P}$ really satisfies the requirement of being minimal. $\mathfrak{R}(\mathfrak{S})$ is called the *minimal ring over the system* $\mathfrak{S}$.

The actual construction of the ring $\mathfrak{R}(\mathfrak{S})$ for a given system $\mathfrak{S}$ is, generally speaking, fairly complicated. However, it becomes quite straightforward in the important special case when the system $\mathfrak{S}$ is a "semiring".

Definition 2. A system of sets $\mathfrak{S}$ is called a *semiring* if it contains the empty set, is closed with respect to the operation of intersection, and has the property that if A and $A_1 \subseteq A$ belong to $\mathfrak{S}$, then A can be represented in the form $A = \bigcup_{k=1}^{n} A_k$, where the A_k are pairwise non-intersecting sets of $\mathfrak{S}$, the first of which is the given set A_1.

In the following pages we shall call each system of non-intersecting sets $A_1, A_2, \cdots, A_n$, the union of which is the given set A, a *finite decomposition* of the set A.

Every ring of sets $\mathfrak{R}$ is a semiring since if A and $A_1 \subseteq A$ belong to $\mathfrak{R}$, then the decomposition

$$A = A_1 \cup A_2,$$

where

$$A_2 = A \setminus A_1 \in \mathfrak{R},$$

takes place.

As an example of a semiring which is not a ring of sets we can take the set of all intervals (a, b), segments $[a, b]$ and half segments $(a, b]$ and $[a, b)$ on the real axis.*

In order to find out how the ring of sets which is minimal over a given semiring is constructed, let us establish some properties of semirings of sets.

* Here, of course, the intervals include the "empty" interval (a, a) and the segments consisting of one point $[a, a]$.

Lemma 1. *Let the sets $A_1, A_2, \cdots, A_n, A$ belong to the semiring $\mathfrak{S}$, and let the sets A_i be pairwise non-intersecting and be subsets of the set A. Then, the sets $A_i (i = 1, 2, \cdots, n)$ can be included as the first n members of the finite decomposition*

$$A = \bigcup_{k=1}^{s} A_k, \qquad s \geq n,$$

of the set A, where all the $A_k \in \mathfrak{S}$.

Proof. The proof will be given by induction. For $n = 1$ the statement of the lemma follows from the definition of a semiring. Let us assume that the result is true for $n = m$ and let us consider $m + 1$ sets $A_1, A_2, \cdots, A_m, A_{m+1}$ satisfying the conditions of the lemma. By the assumptions made,

$$A = A_1 \cup A_2 \cup \cdots \cup A_m \cup B_1 \cup B_2 \cup \cdots \cup B_p,$$

where the sets $B_q (q = 1, 2, \cdots, p)$ belong to $\mathfrak{S}$. Set

$$B_{q1} = A_{m+1} \cap B_q.$$

By the definition of a semiring, we have the decomposition

$$B_q = B_{q1} \cup B_{q2} \cup \cdots \cup B_{qr_q},$$

where all the B_{qj} belong to $\mathfrak{S}$. It is easy to see that

$$A = A_1 \cup \cdots \cup A_m \cup A_{m+1} \cup \bigcup_{q=1}^{p} \bigcup_{j=2}^{r_q} B_{qj}.$$

Thus the assertion of the lemma is proved for $n = m + 1$, and hence for all n.

Lemma 2. *Whatever the finite system of sets $A_1, A_2, \cdots, A_n$ belonging to the semiring may be, one can find in $\mathfrak{S}$ a finite system of pairwise non-intersecting sets $B_1, B_2, \cdots, B_t$ such that each A_k can be represented as a union*

$$A_k = \bigcup_{s \in M_k} B_s$$

of some of the sets B_s.

Proof. For $n = 1$ the lemma is trivial since it suffices to set $t = 1$, $B_1 = A_1$. Assume that it is true for $n = m$ and consider in $\mathfrak{S}$ some system of sets $A_1, A_2, \cdots, A_{m+1}$. Let $B_1, B_2, \cdots, B_t$ be sets from $\mathfrak{S}$, satisfying the conditions of the lemma with respect to $A_1, A_2, \cdots, A_m$. Let us set

$$B_{s1} = A_{m+1} \cap B_s.$$

By Lemma 1, we have the decomposition

$$A_{m+1} = \bigcup_{s=1}^{t} B_{s1} \cup \bigcup_{p=1}^{q} B_p', \qquad B_p' \in \mathfrak{S}, \tag{1}$$

and by the definition of the semiring itself we have the decomposition

$$B_s = B_{s1} \cup B_{s2} \cup \cdots \cup B_{sf_s}, \qquad B_{sq} \in \mathfrak{S}.$$

It is easy to see that

$$A_k = \bigcup_{s \in M_k} \bigcup_{q=1}^{f_s} B_{sq}, \qquad k = 1, 2, \cdots, m,$$

and that the sets

$$B_{sq}, \qquad B_p'$$

are pairwise non-intersecting. Thus the sets B_{sq}, B_p' satisfy the conditions of the lemma with respect to $A_1, A_2, \cdots, A_m, A_{m+1}$. The lemma is thus proved.

Theorem 3. *If $\mathfrak{S}$ is a semiring, then $\mathfrak{R}(\mathfrak{S})$ coincides with the system $\mathfrak{Z}$ of the sets A, which admit of the finite decompositions*

$$A = \bigcup_{k=1}^{n} A_k$$

into the sets $A_k \in \mathfrak{S}$.

Proof. Let us show that the system $\mathfrak{Z}$ forms a ring. If A and B are two arbitrary sets in $\mathfrak{Z}$, the decompositions

$$A = \bigcup_{k=1}^{n} A_k, \qquad B = \bigcup_{k=1}^{m} B_k, \qquad A_k \in \mathfrak{S}, \qquad B_k \in \mathfrak{S},$$

take place. Since $\mathfrak{S}$ is a semiring, the sets

$$C_{ij} = A_i \cap B_j$$

also belong to $\mathfrak{S}$. By Lemma 1 we have the decompositions

$$A_i = \bigcup_j C_{ij} \cup \bigcup_{k=1}^{r_i} D_{ik}, \qquad B_j = \bigcup_i C_{ij} \cup \bigcup_{k=1}^{s_j} E_{jk}, \tag{2}$$

where D_{ik}, $E_{jk} \in \mathfrak{S}$. From equation (2) it follows that the sets $A \cap B$ and $A \triangle B$ admit of the decompositions

$$A \cap B = \bigcup_{i,j} C_{ij},$$

$$A \triangle B = \bigcup_{i,k} D_{ik} \cup \bigcup_{j,k} E_{jk},$$

and therefore belong to $\mathfrak{Z}$. Hence, $\mathfrak{Z}$ is indeed a ring; the fact that it is minimal among all the rings containing $\mathfrak{S}$ is obvious.

In various problems, in particular in measure theory, one has to consider unions and intersections of not only finite, but also infinite numbers of sets. Therefore it is useful to introduce, in addition to the concept of a ring of sets, the following concepts.

Definition 3. A ring of sets is called a *σ-ring,* if with each sequence of sets $A_1, A_2, \cdots, A_n, \cdots$ it also contains the union

$$S = \bigcup_n A_n.$$

Definition 4. A ring of sets is called a *δ-ring,* if in addition to each sequence of sets $A_1, A_2, \cdots, A_n, \cdots$ it also contains the intersection

$$D = \bigcap_n A_n.$$

It is natural to call a σ-ring with a unit a σ-algebra, and a δ-ring with a unit a δ-algebra. However, it is easy to see that these two concepts coincide: each σ-algebra is at the same time a δ-algebra, and each δ-algebra is a σ-algebra. This follows from the duality relations

$$\bigcup_n A_n = E \setminus \bigcap_n (E \setminus A_n),$$

$$\bigcap_n A_n = E \setminus \bigcup_n (E \setminus A_n)$$

(see Chapter 1, §1 of Volume I). The δ-algebras, or what is the same thing, the σ-algebras, are usually called *Borel algebras*, or simply *B-algebras*.

The simplest example of a B-algebra is the set of all subsets of some set A.

A theorem, analogous to Theorem 2 proved above for rings, holds for B-algebras.

Theorem 4. *For any non-empty system of sets $\mathfrak{S}$ there exists a B-algebra $\mathfrak{B}(\mathfrak{S})$, containing $\mathfrak{S}$ and contained in any B-algebra which contains $\mathfrak{S}$.*

The proof follows exactly the same lines as does the proof of Theorem 2. The B-algebra $\mathfrak{B}(\mathfrak{S})$ is called the *minimal B-algebra over the system* $\mathfrak{S}$ or the *Borel closure of the system* $\mathfrak{S}$.

So called *Borel sets* or B-sets play an important role in analysis. These sets can be defined as sets on the real axis belonging to the minimal B-algebra over the set of all segments $[a, b]$.

As a supplement to the information given in Chapter 1, §7 of Volume I, let us note the following facts which we shall need in Chapter II.

Let $y = f(x)$ be a function defined on the set M with values from the set N. Let us denote the system of all maps $f(A)$ of sets from the system $\mathfrak{M}$ (we assume that $\mathfrak{M}$ consists of subsets of the set M) by $f(\mathfrak{M})$ and the system of all subimages $f^{-1}(A)$ of sets from $\mathfrak{N}$ (we assume that $\mathfrak{N}$ consists of subsets of the set N) by $f^{-1}(\mathfrak{N})$. The following statements are true:

1) If $\mathfrak{N}$ is a ring, then $f^{-1}(\mathfrak{N})$ is a ring.
2) If $\mathfrak{N}$ is an algebra, then $f^{-1}(\mathfrak{N})$ is an algebra.
3) If $\mathfrak{N}$ is a B-algebra, then $f^{-1}(\mathfrak{N})$ is a B-algebra.
4) $\mathfrak{R}(f^{-1}(\mathfrak{N})) = f^{-1}(\mathfrak{R}(\mathfrak{N}))$.
5) $\mathfrak{B}(f^{-1}(\mathfrak{N})) = f^{-1}(\mathfrak{B}(\mathfrak{N}))$.

Let $\mathfrak{R}$ be some ring of sets. If in it we take the operation $A \triangle B$ to be "addition" and $A \cap B$ to be "multiplication", then $\mathfrak{R}$ is a ring in the usual algebraic sense of the word. All its elements will satisfy the conditions

$$a + a = 0, \qquad a^2 = a. \tag{$*$}$$

Rings whose elements satisfy condition $(*)$ are called "Boolian" rings. Each Boolian ring can be realised as a ring of sets with the operations $A \triangle B$ and $A \cap B$ (Stone).

3. Measures on Semirings. Continuation of a Measure from a Semiring to the Minimal Ring over it

In §1, when considering measure in the plane, we started from the measure of a rectangle (area) and then extended the concept of measure to a wider class of sets. The results as well as the methods given in §1 have a quite general character and can be generalized to measures defined on arbitrary sets without essential changes. The first step in constructing a measure on a plane consisted in generalizing the concept of measure from rectangles to elementary sets, i.e., to finite systems of pairwise non-intersecting rectangles.

In this section we shall consider the abstract analogue of this problem.

Definition 1. The set function $\mu(A)$ is called a *measure* if:

1) its domain of definition S_μ is a semiring of sets;

2) its values are real and non-negative;

3) it is additive, i.e., for any finite decomposition

$$A = \bigcup A_k$$

of the set $A \in S_\mu$ into sets $A_k \in S_\mu$, the equation

$$\mu(A) = \sum \mu(A_k)$$

holds.

Remark. From the decomposition $\varnothing = \varnothing \cup \varnothing$ it follows that $\mu(\varnothing) = 2\mu(\varnothing)$, i.e., $\mu(\varnothing) = 0$.

The following two theorems about measures on semirings will be frequently used in the subsequent pages.

Theorem 1. *Let μ be a measure defined on some semiring S_μ. If the sets $A_1, A_2, \cdots, A_n, A$ belong to S_μ, where the A_k are pairwise non-*

intersecting and all belong to A, then

$$\sum_{k=1}^{n} \mu(A_k) \leq \mu(A).$$

Proof. Since S_μ is a semiring there exists, according to Lemma 1, §2, the decomposition

$$A = \bigcup_{k=1}^{s} A_k, \qquad s \geq n, \qquad A_k \in S_\mu,$$

where the first n sets coincide with the given sets $A_1, A_2, \cdots, A_n$. Since the measure of any set is non-negative,

$$\sum_{k=1}^{n} \mu(A_k) \leq \sum_{k=1}^{s} \mu(A_k) = \mu(A).$$

Theorem 2. *If $A_1, A_2, \cdots, A_n, A$ belong to S_μ and $A \subseteq \bigcup_{k=1}^{n} A_k$, then*

$$\mu(A) \leq \sum_{k=1}^{n} \mu(A_k).$$

Proof. By Lemma 2, §2, one can find a system of pairwise non-intersecting sets $B_1, B_2, \cdots, B_t$ from S_μ, such that each of the sets $A_1, A_2, \cdots, A_n, A$ can be represented as a union of some of the sets B_s:

$$A = \bigcup_{s \in M_0} B_s, \qquad A_k = \bigcup_{s \in M_k} B_s, \qquad k = 1, 2, \cdots, n.$$

Moreover, each index $s \in M_0$ also belongs to some member of M_k. Therefore each term of the sum

$$\sum_{s \in M_0} \mu(B_s) = \mu(A)$$

enters once, or at most a few times, into the double sum

$$\sum_{k=1}^{n} \sum_{s \in M_k} \mu(B_s) = \sum_{k=1}^{n} \mu(A_k).$$

This yields

$$\mu(A) \leq \sum_{k=1}^{n} \mu(A_k).$$

In particular, for $n = 1$, we have the

Corollary. *If $A \subseteq A'$, then $\mu(A) \leq \mu(A')$.*

Definition 2. The measure $\mu(A)$ is called the *continuation of the measure $m(A)$* if $S_m \subseteq S_\mu$ and if, for every $A \in S_m$, the equality

$$\mu(A) = m(A)$$

holds.

The main aim of the present section is the proof of the following proposition.

Theorem 3. *Every measure $m(A)$ has one and only one continuation $\mu(A)$, having as its domain of definition the ring $\mathfrak{R}(S_m)$.*

Proof. For each set $A \in \mathfrak{R}(S_m)$ there exists a decomposition

$$A = \bigcup_{k=1}^{n} B_k, \qquad B_k \in S_m, \tag{1}$$

(Theorem 3, §2). Let us assume by definition

$$\mu(A) = \sum_{k=1}^{n} m(B_k). \tag{2}$$

It is easy to see that the quantity $\mu(A)$, given by equation (2), does not depend on the selection of the decomposition (1). Indeed, let us consider the two decompositions

$$A = \bigcup_{i=1}^{m} B_i = \bigcup_{j=1}^{n} C_j, \qquad B_i \in S_m, \qquad C_j \in S_m.$$

Since all intersections $B_i \cap C_j$ belong to S_m, we have, because of the additivity of measures,

$$\sum m(B_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} m(B_i \cap C_j) = \sum m(C_j),$$

which was to be proved. The fact that the function $\mu(A)$, given by equation (2) is non-negative and additive is obvious. Hence the existence of a continuation $\mu(A)$ of the measure $m(A)$ is shown. To show its uniqueness let us note that, by definition of continuation, if $A = \bigcup_{k=1}^{n} B_k$, where B_k are non-intersecting sets from S_m, then for any continuation μ^* of the measure m onto the ring $\mathfrak{R}(S_m)$

$$\mu^*(A) = \sum \mu^*(B_k) = \sum m(B_k) = \mu(A),$$

i.e., the measure μ^* coincides with the measure μ defined by equation (2). The theorem is proved.

The connection between this theorem and the constructions of §1 will be completely clear if we note that the set of rectangles in the plane is a semiring, the area of these rectangles is a measure in the sense of Definition 1, and the elementary plane sets form a minimal ring over the semiring of the rectangles.

4. Continuations of Jordan Measures*

In the present section we shall consider the general form of that process which in the case of plane figures allows one to generalise from the definition of areas for a finite union of rectangles, with sides parallel to the axes of coordinates, to areas of all those figures for which areas are defined by elementary geometry or classical analysis. This extension was given with complete precision by the French mathematician Jordan around 1880. The basic idea of Jordan goes back, incidentally, to the mathematicians of ancient Greece and consists of approximating from the inside and from the outside the "measurable" set A by sets A' and A'' to which a measure has already been prescribed, i.e., in such a way that the inclusions

$$A' \subseteq A \subseteq A''$$

are fulfilled.

* The concept of a Jordan measure has a definite historical and methodological interest but is not used in this exposition. The reader may omit this section if he wishes.

Since we can continue any measure onto a ring (Theorem 3, §3), it is natural to assume that the initial measure m be defined on a ring $\mathfrak{R} = \mathfrak{R}(S_m)$. This assumption will be used during the whole of the present section.

Definition 1. We shall call a set A *Jordan measurable* if, for any $\varepsilon > 0$, there exist in the ring $\mathfrak{R}$ sets A' and A'' which satisfy the conditions

$$A' \subseteq A \subseteq A'', \qquad m(A'' \setminus A') < \varepsilon.$$

Theorem 1. *The system $\mathfrak{R}^*$ of Jordan measurable sets is a ring.*

Indeed, let $A \in \mathfrak{R}^*$, $B \in \mathfrak{R}^*$; then, for any $\varepsilon > 0$, there exist A', A'', B', $B'' \in \mathfrak{R}$ such that

$$A' \subseteq A \subseteq A'', \qquad B' \subseteq B \subseteq B'',$$

and

$$m(A'' \setminus A') < \frac{\varepsilon}{2}, \qquad m(B'' \setminus B') < \frac{\varepsilon}{2}.$$

Hence

$$A' \cup B' \subseteq A \cup B \subseteq A'' \cup B'', \tag{1}$$

$$A' \setminus B'' \subseteq A \setminus B \subseteq A'' \setminus B'. \tag{2}$$

Since

$$(A'' \cup B'') \setminus (A' \cup B') \subseteq (A'' \setminus A') \cup (B'' \setminus B'),$$

we have

$$m[(A'' \cup B'') \setminus (A' \cup B')] \leq m[(A'' \setminus A') \cup (B'' \setminus B')]$$
$$\leq m(A'' \setminus A') + m(B'' \setminus B') < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{3}$$

Since

$$(A'' \setminus B') \setminus (A' \setminus B'') \subseteq (A'' \setminus A') \cup (B'' \setminus B'),$$

we have

$$m[(A'' \setminus B') \setminus (A' \setminus B'')] \leq m[(A'' \setminus A') \cup (B'' \setminus B')]$$
$$\leq m(A'' \setminus A') + m(B'' \setminus B') \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{4}$$

Since $\varepsilon > 0$ is arbitrary, and the sets $A' \cup B'$, $A'' \cup B''$, $A' \setminus B''$ and $A'' \setminus B'$ belong to $\mathfrak{R}$, (1), (2), (3) and (4) imply that $A \cup B$ and $A \setminus B$ belong to $\mathfrak{R}^*$.

Let $\mathfrak{M}$ be a system of those sets A for which the set $B \supseteq A$ of $\mathfrak{R}$ exists. For any A from $\mathfrak{M}$ we set, by definition,

$$\bar{\mu}(A) = \inf_{B \supseteq A} m(B),$$

$$\underline{\mu}(A) = \sup_{B \subseteq A} m(B).$$

The functions $\bar{\mu}(A)$ and $\underline{\mu}(A)$ are called, respectively, the "outer" and the "inner" measure of the set A.

Obviously, always

$$\underline{\mu}(A) \leq \bar{\mu}(A).$$

Theorem 2. *The ring $\mathfrak{R}^*$ coincides with the system of those sets $A \in \mathfrak{M}$ for which $\underline{\mu}(A) = \bar{\mu}(A)$.*

Proof. If

$$\bar{\mu}(A) \neq \underline{\mu}(A),$$

then

$$\bar{\mu}(A) - \underline{\mu}(A) = h > 0,$$

and for any A' and A'' from $\mathfrak{R}$ for which $A' \subseteq A \subseteq A''$,

$$m(A') \leq \underline{\mu}(A), \qquad m(A'') \geq \bar{\mu}(A),$$

$$m(A'' \setminus A') = m(A'') - m(A') \geq h > 0,$$

i.e., A cannot belong to $\mathfrak{R}^*$.

Conversely, if

$$\underline{\mu}(A) = \bar{\mu}(A),$$

then, for any $\varepsilon > 0$, there exist A' and A'' from $\mathfrak{R}$ for which

$$A' \subseteq A \subseteq A'',$$

$$\underline{\mu}(A) - m(A') < \frac{\varepsilon}{2},$$

$$m(A'') - \bar{\mu}(A) < \frac{\varepsilon}{2},$$

$$m(A'' \setminus A') = m(A'') - m(A') < \varepsilon,$$

i.e.,

$$A \in \mathfrak{R}^*.$$

The following theorems hold for sets from $\mathfrak{M}$.

Theorem 3. *If* $A \subseteq \bigcup_{k=1}^{n} A_k$, *then* $\bar{\mu}(A) \leq \sum_{k=1}^{n} \bar{\mu}(A_k)$.

Proof. Let us select A_k' such that

$$A_k \subseteq A_k', \qquad m(A_k') \leq \bar{\mu}(A_k) + \frac{\varepsilon}{2^k},$$

and let us form $A' = \bigcup_{k=1}^{n} A_k'$. Then,

$$m(A') \leq \sum_{k=1}^{n} m(A_k') \leq \sum_{k=1}^{n} \bar{\mu}(A_k) + \varepsilon, \qquad \bar{\mu}(A) \leq \sum_{k=1}^{n} \bar{\mu}(A_k) + \varepsilon,$$

and since ε is arbitrary, $\bar{\mu}(A) \leq \sum_{k=1}^{n} \bar{\mu}(A_k')$.

Theorem 4. *If* $A_k \subseteq A$ $(k = 1, 2, \cdots, n)$ *and* $A_i \cap A_j = \varnothing$, *then*

$$\underline{\mu}(A) \geq \sum_{k=1}^{n} \underline{\mu}(A_k).$$

Proof. Let us select $A_k' \subseteq A_k$ such that

$$m(A_k') \geq \underline{\mu}(A_k) - \frac{\varepsilon}{2^k},$$

and let us form $A' = \bigcup_{k=1}^{n} A_k'$. Then $A_i' \cap A_j' = \varnothing$ and

$$m(A') = \sum_{k=1}^{n} m(A_k') \geq \sum_{k} \underline{\mu}(A_k) + \varepsilon.$$

Since $A' \subseteq A$,

$$\underline{\mu}(A) \geq m(A') \geq \sum_{k=1}^{n} \underline{\mu}(A_k) - \varepsilon.$$

Because $\varepsilon > 9$ is arbitrary,

$$\underline{\mu}(A) \geq \sum_{k=1}^{n} \underline{\mu}(A_k).$$

Let us now define the function μ with the domain of definition

$$S_\mu = \mathfrak{R}^*$$

as the common value of the inner and outer measures:

$$\mu(A) = \underline{\mu}(A) = \bar{\mu}(A).$$

Theorems 3 and 4, and the obvious fact that for $A \in \mathfrak{R}$

$$\bar{\mu}(A) = \underline{\mu}(A) = m(A),$$

imply

Theorem 5. *The function $\mu(A)$ is a measure and a continuation of the measure m.*

The construction given can be used for any measure m defined on a ring.

The system $S_{m_2} = \mathfrak{S}$ of elementary sets in a plane is essentially connected with the coordinate system: the sets of the system $\mathfrak{S}$ consist of rectangles with sides which are parallel to the coordinate axes. In going over to the Jordan measure

$$J^{(2)} = j(m_2)$$

this dependence on the choice of a coordinate system disappears: starting from an arbitrary system of coordinates $\{\bar{x}_1, \bar{x}_2\}$ connected with the initial system $\{x_1, x_2\}$ by the orthogonal transformation

$$\bar{x}_1 = \cos\alpha \cdot x_1 + \sin\alpha \cdot x_2 + a_1,$$

$$\bar{x}_2 = -\sin\alpha \cdot x_1 + \cos\alpha \cdot x_2 + a_2,$$

we obtain the same Jordan measure

$$J = j(m_2) = j(\bar{m}_2)$$

(here $\bar{m}_2$ denotes the measure constructed with the help of rectangles with sides parallel to the axes $\bar{x}_1$, $\bar{x}_2$). This fact can be proved with the help of the following general theorem:

Theorem 6. *In order that the Jordan continuations $\mu_1 = j(m_1)$ and $\mu_2 = j(m_2)$ of the measures m_1 and m_2 defined on the rings $\mathfrak{R}_1$ and*

$\mathfrak{R}_2$ *coincide, it is necessary and sufficient that the following conditions be satisfied:*

$$\mathfrak{R}_1 \in S_{\mu_2}, \qquad m_1(A) = \mu_2(A) \text{ on } \mathfrak{R}_1,$$
$$\mathfrak{R}_2 \in S_{\mu_1}, \qquad m_2(A) = \mu_1(A) \text{ on } \mathfrak{R}_2.$$

The necessity of the condition is obvious. Let us prove their sufficiency.

Let $A \in S_{\mu_1}$. Then there exist $A', A'' \in S_{m_1}$, such that

$$A' \subseteq A \subseteq A'', \qquad m_1(A'') - m_1(A') < \frac{\varepsilon}{3},$$

and

$$m_1(A') \leq \mu_1(A) \leq m_1(A'').$$

By the conditions of the theorem, $\mu_2(A') = m_1(A')$ and $\mu_2(A'') = m_1(A'')$.

From the definition of the measure μ_2 it follows that there exist $B' \in S_{m_2}$ and $B'' \in S_{m_2}$ for which,

$$A' \supseteq B' \quad \text{and} \quad \mu_2(A') - m_2(B') < \frac{\varepsilon}{3},$$

$$B'' \supseteq A'' \quad \text{and} \quad m_2(B'') - \mu_2(A'') < \frac{\varepsilon}{3}.$$

Here

$$B' \subseteq A \subseteq B'',$$

and, obviously,

$$m_2(B'') - m_2(B') < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, $A \in S_{\mu_2}$, and from the relations

$$\mu_1(B') = m_2(B') \leq \mu_2(A) \leq m_2(B'') = \mu_1(B'')$$

it follows that

$$\mu_2(A) = \mu_1(A).$$

The theorem is proved.

To establish that the Jordan measure in the plane is independent of the choice of the system of coordinates, one need only convince oneself that a set which is obtained from an elementary set by a rotation through some angle α is Jordan measurable. It is suggested that the reader do this for himself.

If the initial measure is given, not on a ring, but on a semiring, then it is natural to consider as its Jordan continuation the measure

$$j(m) = j(r(m)),$$

obtained as a result of a continuation of m to the ring $\mathfrak{R}(S_m)$ and a subsequent continuation.

5. Countable Additivity. General Problem of Continuation of Measures

Often one must consider the union of not only a finite, but of a countable number of sets. In this connection the condition of additivity which we have imposed on measures (Definition 1, §2) turns out to be insufficient and it is natural to replace it by the stronger requirement of countable additivity.

Definition 1. The measure μ is called *countably additive* (or *σ-additive*), if, for any sets $A, A_1, \cdots, A_n, \cdots$, belonging to its domain of definition S_μ and satisfying the conditions

$$A = \bigcup_{n=1}^{\infty} A_n,$$

$$A_i \cap A_j = \varnothing \quad \text{for} \quad i \neq j,$$

the equality

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$$

holds. The plane Lebesgue measure which we constructed in §1 is σ-additive (Theorem 9). An example of a σ-additive measure of a completely different kind can be constructed in the following way. Let

$$X = \{x_1, x_2, \cdots\}$$

be an arbitrary countable set, and let the numbers $p_n > 0$ be such that

$$\sum p_n = 1.$$

The domain S_μ consists of all subsets of the set X. For each $A \subseteq X$ we set

$$\mu(A) = \sum_{x_n \in A} p_n.$$

It is easy to check that $\mu(A)$ is a σ-additive measure, where $\mu(X) = 1$. This example occurs naturally in connection with many questions of probability theory.

Let us give an example of a measure which is additive but not σ-additive. Let X be the set of all rational points of the segment $[0, 1]$, and S_μ consist of the intersections of the set X with arbitrary intervals (a, b), segments $[a, b]$ and half segments $(a, b]$, $[a, b)$. It is easy to see that S_μ is a semiring. For each such set we put

$$\mu(A_{ab}) = b - a.$$

This is an additive measure. It is not σ-additive because, for example, $\mu(X) = 1$ and at the same time X is the union of a countable number of separate points, the measure of each one of which is zero.

In this and the two following sections we shall consider σ-additive measures and their different σ-additive continuations.

Theorem 1. *If the measure m, defined on some semiring S_m, is countably additive, then the measure $\mu = r(m)$, obtained from it by continuation to the ring $\mathfrak{R}(S_m)$, is also countably additive.*

Proof. Let

$$A \in \mathfrak{R}(S_m), \qquad B_n \in \mathfrak{R}(S_m), \qquad n = 1, 2, \cdots,$$

and

$$A = \bigcup_{n=1}^{\infty} B_n,$$

where $B_s \cap B_r = \varnothing$ for $s \neq r$. Then there exist sets A_j and B_{ni} from S_m such that

$$A = \bigcup_j A_j, \qquad B_n = \bigcup_i B_{ni},$$

where the sets on the right-hand sides of each of these equations are pairwise non-intersecting and the union over i and j is finite. (Theorem 3, §2).

Let $C_{nij} = B_{ni} \cap A_j$. It is easy to see that the sets C_{nij} are pairwise non-intersecting, and hence,

$$A_j = \bigcup_n \bigcup_i C_{nij},$$

$$B_{ni} = \bigcup_j C_{nij}.$$

Therefore, and because of the additivity of the measure m on S_m, we have

$$m(A_j) = \sum_n \sum_j m(C_{nij}), \tag{1}$$

$$m(B_{ni}) = \sum_j m(C_{nij}), \tag{2}$$

and, by definition of the measure $r(m)$ on $\mathfrak{R}(S_m)$,

$$\mu(A) = \sum_j m(A_j), \tag{3}$$

$$\mu(B_n) = \sum_i m(B_{ni}). \tag{4}$$

Equations (1), (2), (3) and (4) imply $\mu(A) = \sum_n \mu(B_n)$. (The summations over i and j are finite, the series in n converge.)

One could show that a Jordan continuation of a σ-additive measure is always σ-additive; there is however no need to do this in this special case since it will follow from the theory of Lebesgue continuations which will be given in the next section.

Let us now show that, for the case of σ-additive measures, Theorem 2 of §3 may be extended to countable coverings.

Theorem 2. *If the measure μ is σ-additive, and the sets A, A_1, A_2, $\cdots$, A_n, $\cdots$ belong to S_μ, then*

$$A \subseteq \bigcup_{n=1}^{\infty} A_n$$

implies the inequality

$$\mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

Proof. By Theorem 1, it is enough to give the proof for measures defined on a ring, since from the validity of Theorem 2 for $\mu = r(m)$ it immediately follows that it can be applied also to the measure m. If S_μ is a ring, the sets

$$B_n = (A \cap A_n) \setminus \bigcup_{k=1}^{n-1} A_k$$

belong to S_μ. Since

$$A = \bigcup_{n=1}^{\infty} B_n, \qquad B_n \subseteq A_n,$$

and since the sets B_n are pairwise non-intersecting,

$$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

From now on we shall, without special mention, consider only σ-additive measures.

We have already considered above two methods of continuation of measures. In connection with the continuation of the measure m to the ring $\Re(S_m)$ in §3 we noted the uniqueness of this continuation. The case of a Jordan continuation $j(m)$ of an arbitrary measure m is analogous. If the set A is Jordan measurable with respect to the measure m (belongs to $S_{j(m)}$), then, for any measure μ continuing m and defined on A, the value $\mu(A)$ coincides with the value $J(A)$ of the Jordan continuation $J = j(m)$. One can show that the extension of the measure m beyond the boundaries of the system $S_{j(m)}$ is not unique. More precisely this means the following. Let us call the set A the *set of uniqueness* for the measure m, if:

1) there exists a measure which is a continuation of the measure m, defined for the set A;

2) for any two measures of this kind μ_1 and μ_2

$$\mu_1(A) = \mu_2(A).$$

The following theorem holds: The system of sets of uniqueness for the measure m coincides with the system of sets which are Jordan measurable with respect to the measure m, i.e., with the system of sets $S_{j(m)}$.

However, if one considers only σ-additive measures and their continuation (σ-additive), then the system of sets of uniqueness will be, generally speaking, wider.

Since it is the case of σ-additive measures that will interest us in the future let us establish

Definition 2. The set A is called the set of σ-uniqueness for a σ-additive measure μ, if:

1) there exists a σ-additive continuation λ of the measure m defined for A (i.e., such that $A \in S_\lambda$);
2) for two such σ-additive continuations λ_1 and λ_2 the equation

$$\lambda_1(A) = \lambda_2(A)$$

holds. If A is a set of σ-additivity for the σ-additive measure μ, then, by our definition, there exists only one possible $\lambda(A)$ for the σ-additive continuation of the measure μ, defined on A.

6. Lebesgue Continuation of Measure, Defined on a Semiring with a Unit

Even though the Jordan continuation allows one to generalise the concept of measure to quite a wide class of sets, it still remains insufficient in many cases. Thus, for example, if we take as the initial measure the area, and as the domain of its definition the semiring of rectangles and consider the Jordan continuation of this measure, then even such a comparatively simple set as the set of points, the coordinates of which are rational and satisfy the condition $x^2 + y^2 \leq 1$, is not Jordan measurable.

A generalisation of a σ-additive measure defined on some semiring to a class of sets which is maximal in the well known sense can be obtained with the help of the so-called Lebesgue continuation. In this section we shall consider the Lebesgue continuation of a measure defined on a semiring with a unit. The general case will be considered in §7.

The construction given below represents, to a large degree, a repetition, in abstract terms, of the construction of the Lebesgue measure for plane sets given in §1.

Let a σ-additive measure m be given on some semiring of sets S_m with unit E. We shall define on the system $\mathfrak{S}$ of all subsets of the set E the functions $\mu^*(A)$ and $\mu_*(A)$ in the following way.

Definition 1. The number

$$\mu^*(A) = \inf_{A \subseteq \bigcup_n B_n} \sum_n m(B_n),$$

where the lower bound is taken over all coverings of the set A by finite or countable systems of sets $B_n \in S_m$, is called the *outer measure* of the set $A \subseteq E$.

Definition 2. The number

$$\mu_*(A) = m(E) - \mu^*(E \setminus A)$$

is called the *inner measure* of the set $A \subseteq E$.

From Theorem 2, §3 it follows that always $\mu_*(A) \leq \mu^*(A)$.

Definition 3. The set $A \subseteq E$ is called *measurable* (Lebesgue), if

$$\mu_*(A) = \mu^*(A).$$

If A is measurable, then we shall denote the common value $\mu_*(A) = \mu^*(A)$ by $\mu(A)$ and call it the (Lebesgue) measure of the set A.

It is obvious that, if A is measurable, then its complement is also measurable.

Theorem 2, §5 immediately implies that for any σ-additive continuation μ of the measure m the inequality

$$\mu_*(A) \leq \mu(A) \leq \mu^*(A)$$

holds. Therefore, for a measurable set A, each σ-additive continuation μ of the measure m (if it exists at all) necessarily equals the common value $\mu_*(A) = \mu^*(A)$. The Lebesgue measure is nothing but the σ-additive continuation of the measure m to the set of all measurable (in the sense of Definition 3) sets. The definition of measurability can obviously also be formulated in the following way:

Definition 3′. The set $A \subseteq E$ is called measurable, if

$$\mu^*(A) + \mu^*(E \setminus A) = m(E).$$

It is expedient to use, aside of the initial measure m, its continuation $m' = r(m)$ onto the ring $\mathfrak{R}(S_m)$ which is already known to us (§3). It is clear that the following definition is equivalent to Definition 1.

Definition 1′. The number

$$\mu^*(A) = \inf_{A \subseteq \bigcup_n B_n'} \sum_n m'(B_n'), \qquad B_n' \subseteq \mathfrak{R}(S_m),$$

is called the outer measure of the set A.

Indeed, since the measure m' is σ-additive (Theorem 1, §5), any sum $\sum_n m'(B_n')$, where $B_n' \in \mathfrak{R}(S_m)$, can be replaced by the sum

$$\sum_{n,k} m(B_{nk}), \qquad B_{nk} \in S_m,$$

which is equal to it, and where $B_n' = \bigcup_k B_{nk}$, $B_{ni} \cap B_{nj} = \varnothing$ if $i \neq j$.

The following are basic facts.

Theorem 1. *If*

$$A \subseteq \bigcup_n A_n,$$

where $\{A_n\}$ is a finite or countable system of sets, then

$$\mu^*(A) \leq \sum_n \mu^*(A_n).$$

Theorem 2. *If $A \in \mathfrak{R}$, then $\mu_*(A) = m'(A) = \mu^*(A)$,* i.e., *all the sets from $\mathfrak{R}(S_m)$ are measurable, and for them the inner and outer measures coincide with m'.*

Theorem 3. *For the measurability of the set A the following condition is necessary and sufficient: for any $\varepsilon > 0$ there exists a $B \in \mathfrak{R}(S_m)$ such that*

$$\mu^*(A \,\triangle\, B) < \varepsilon.$$

In §1 these statements were proved for the plane Lebesgue measure (Theorems 3–5, §1). The proofs given there can be carried over word for word to the general case considered here, therefore we shall not repeat them.

Theorem 4. *The system $\mathfrak{M}$ of all measurable sets is a ring.*

Proof. Since it is always true that

$$A_1 \cap A_2 = A_1 \setminus (A_1 \setminus A_2)$$

and

$$A_1 \cup A_2 = E \setminus [(E \setminus A_1) \cap (E \setminus A_2)],$$

it suffices to show the following. If $A_1 \in \mathfrak{M}$ and $A_2 \in \mathfrak{M}$, then also

$$A = A_1 \setminus A_2 \in \mathfrak{M}.$$

Let A_1 and A_2 be measurable; then there exist $B_1 \in \mathfrak{R}(S_m)$ and $B_2 \in \mathfrak{R}(S_m)$ such that

$$\mu^*(A_1 \triangle B_1) < \frac{\varepsilon}{2} \quad \text{and} \quad \mu^*(A_2 \triangle B_2) < \frac{\varepsilon}{2}.$$

Setting $B = B_1 \setminus B_2 \in \mathfrak{R}(S_m)$ and using the relation

$$(A_1 \setminus A_2) \triangle (B_1 \setminus B_2) \subseteq (A_1 \triangle B_1) \cup (A_2 \triangle B_2),$$

we obtain

$$\mu^*(A \triangle B) < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this implies that the set A is measurable.

Remark. Obviously E is the unit of the ring $\mathfrak{M}$ which, therefore, is an algebra of sets.

Theorem 5. *On the system $\mathfrak{M}$ of measurable sets, the function $\mu(A)$ is additive.*

The proof of this theorem is a word for word repetition of the proof of Theorem 7, §1.

Theorem 6. *On the system $\mathfrak{M}$ of measurable sets, the function $\mu(A)$ is σ-additive.*

Proof. Let

$$A = \bigcup_{n=1}^{\infty} A_n, \qquad A \subseteq \mathfrak{M}, \qquad A_i \cap A_j = \varnothing \quad \text{for} \quad i \neq j.$$

By Theorem 1,

$$\mu^*(A) \leq \sum_n \mu(A_n), \tag{1}$$

and by Theorem 5, for any N

$$\mu^*(A) \geq \mu^*\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mu^*(A_n),$$

implying

$$\mu^*(A) \geq \sum_n \mu(A_n). \tag{2}$$

Inequalities (1) and (2) yield the assertion of the theorem.

Thus we have established that the function $\mu(A)$, defined on the system $\mathfrak{M}$, possesses all the properties of a σ-additive measure.

Hence the following definition is verified:

Definition 4. One calls the function $\mu(A)$, defined on a system $S_\mu = \mathfrak{M}$ of measurable sets, and coinciding on this system with the outer measure $\mu^*(A)$, the *Lebesgue continuation* $\mu = L(m)$ of the measure $m(A)$.

In §1, considering the plane Lebesgue measure, we have shown that not only the finite but also the countable unions and intersections of measurable sets are also measurable sets. This is true also in the general case i.e., the following theorem holds.

Theorem 7. *The system $\mathfrak{M}$ of Lebesgue measurable sets is a Borel algebra with unit E.*

Proof. Since

$$\bigcap_n A_n = E \setminus \bigcup_n (E \setminus A_n),$$

and since the complement of a measurable set is measurable, it suffices to show the following. If $A_1, A_2, \cdots, A_n, \cdots$ belong to $\mathfrak{M}$, then $A = \bigcup_n A_n$ also belongs to $\mathfrak{M}$. The proof of this statement given in Theorem 8, §1, for plane sets, is literally preserved also in the general case.

Exactly as in the case of a plane Lebesgue measure its σ-additivity implies its continuity, i.e., if μ is a σ-additive measure, defined on a B-algebra, $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$ is a decreasing chain of measurable sets and

$$A = \bigcap_n A_n,$$

then

$$\mu(A) = \lim_{n\to\infty} \mu(A_n),$$

and if $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ is an increasing chain of measurable sets and

$$A = \bigcup_n A_n,$$

then

$$\mu(A) = \lim_{n\to\infty} \mu(A_n).$$

The proof given in §1 for a plane measure (Theorem 10) can be carried over to the general case.

1) From the results of §§5 and 6 it is easy to deduce that every set A which is Jordan measurable is Lebesgue measurable; moreover its Jordan and Lebesgue measures are equal. This immediately implies that the Jordan continuation of a σ-additive measure is σ-additive.

2) Every set A which is Lebesgue measurable is a set of uniqueness for the initial measure m. Indeed, for any $\epsilon > 0$ there exists for A a $B \in \mathfrak{R}$ such that $\mu^*(A \triangle B) < \epsilon$. Whatever the extension λ of the measure m may be,

$$\lambda(B) = m'(B),$$

since the continuation of the measure m to $\mathfrak{R} = \mathfrak{R}(S_m)$ is unique. Furthermore,

$$\lambda(A \triangle B) \leq \mu^*(A \triangle B) < \varepsilon,$$

and therefore

$$|\lambda(A) - m'(B)| < \varepsilon.$$

Thus we have, for two arbitrary continuations $\lambda_1(A)$ and $\lambda_2(A)$ of the measure m,

$$|\lambda_1(A) - \lambda_2(A)| < 2\varepsilon,$$

which, because of the arbitrariness of ϵ, implies

$$\lambda_1(A) = \lambda_2(A).$$

One can show that the system of Lebesgue measurable sets exhausts the whole system of sets of uniqueness for the initial measure m.

3) Let m be some σ-additive measure with the domain of definition S and let $\mathfrak{M} = L(S)$ be the domain of definition of its Lebesgue continuation. From Theorem 3 of this section it easily follows that whatever the semiring S_1 for which

$$S \subseteq S_1 \subseteq \mathfrak{M},$$

we always have

$$L(S_1) = L(S).$$

7. Lebesgue Continuation of Measures In the General Case

If the semiring S_m on which the initial measure m is defined does not have a unit, then the exposition of §6 must be slightly changed. Definition 1 of the outer measure is preserved, but the outer measure μ^* turns out to be defined only on the system S_{μ^*} of such sets A for which the coverings $\bigcup_n B_n$ by sets from S_m with a finite sum

$$\sum_n m(B_n)$$

exists. Definition 2 loses its meaning. The inner measure may be defined (in a slightly different way) also in the general case, but we shall not go into this. For the definition of measurability of sets it is expedient to take now the property of measurable sets implied by Theorem 3.

Definition 1. The set A is called *measurable*, if for any $\varepsilon > 0$ there exists a set $B \in S_m$ such that $\mu^*(A \triangle B) < \varepsilon$.

Theorems 4, 5 and 6 and the final Definition 4 stay in force. In the proofs we used the assumption of the existence of a unit only in proving Theorem 4. To give the proof of Theorem 4 for the general case we must show again that from $A_1 \in M$, $A_2 \in M$ it follows that $A_1 \cup A_2 \in M$. This proof is carried out exactly as for $A_1 \setminus A_2$ on the basis of the inclusion

$$(A_1 \cup A_2) \triangle (B_1 \cup B_2) \subseteq (A_1 \triangle B_1) \cup (A_2 \triangle B_2).$$

In the case when S_m does not have a unit, Theorem 7, §6 is replaced by the following:

Theorem 1. *For any initial measure m, the system of sets $\mathfrak{M} = S_{L(m)}$ which are Lebesgue measurable is a δ-ring. For measurable A_n the set $A = \bigcup_{n=1}^{\infty} A_n$ is measurable if and only if the measures $\mu\left(\bigcup_{n=1}^{\infty} A_n\right)$ are bounded by some constant which does not depend on N.*

The proof of this assertion is left to the reader.

Remark. In our exposition the measures are always finite, therefore the necessity of the last condition is obvious.

Theorem 1 implies the following

Corollary. *The system $\mathfrak{M}_A$ of all sets $B \in \mathfrak{M}$ which are subsets of a fixed set $A \in \mathfrak{M}$ forms a Borel algebra.*

For example, the system of all Lebesgue measurable (in the sense of the usual Lebesgue measure on the line) subsets of any segment $[a, b]$ is a Borel algebra of sets.

In conclusion let us mention one more property of Lebesgue measures.

Definition 2. The measure μ is called *complete*, if $\mu(A) = 0$ and $A' \subseteq A$ imply $A' \in S_\mu$.

It is obvious that here $\mu(A') = 0$. Without any difficulty one can show that the Lebesgue continuation of any measure is complete. This follows from the fact that for $A' \subseteq A$ and $\mu(A) = 0$ necessarily $\mu^*(A') = 0$, and any set C for which $\mu^*(C) = 0$ is measurable, since $\varnothing \in \mathfrak{R}$ and

$$\mu^*(C \triangle \varnothing) = \mu^*(C) = 0.$$

Let us point out the connection between the process of Lebesgue continuation of measures and the process of completion of a metric space. For this let us note that $m'(A \triangle B)$ can be taken as the distance between the elements A and B of the ring $\mathfrak{R}(S_m)$. Then $\mathfrak{R}(S_m)$ becomes a metric space (generally speaking not complete), and its completion, by Theorem 3, §6, consists exactly of all the measurable sets. (Here, however, the sets A and B are indistinguishable, from the metric point of view, if $\mu(A \triangle B) = 0$.)

CHAPTER II

MEASURABLE FUNCTIONS

8. Definition and Basic Properties of Measurable Functions

Let X and Y be two arbitrary sets, and assume that two systems of subsets $\mathfrak{S}$ and $\mathfrak{S}'$, respectively, have been selected from them. The abstract function $y = f(x)$, with the domain of definition X, taking on values from Y, is called $(\mathfrak{S}, \mathfrak{S}')$—measurable if from $A \in \mathfrak{S}'$ it follows that $f^{-1}(A) \in \mathfrak{S}$.

For example, if we take as X and Y the real axis D^1 (i.e., consider real functions of a real variable), and as $\mathfrak{S}$ and $\mathfrak{S}'$ take the system of all open (or all closed) subsets of D^1, then the stated definition of measurability reduces to the definition of continuity (§12 of Volume I). Taking for $\mathfrak{S}$ and $\mathfrak{S}'$ the system of all Borel sets, we arrive at the so-called B-measurable (or Borel measurable) functions.

In what follows we shall be interested in the concept of measurability mainly from the point of view of the theory of integration. Of basic importance in this connection is the concept of μ-measurability of real functions, defined on some set X, where one takes for $\mathfrak{S}$ the system of all μ-measurable subsets of the set X and for $\mathfrak{S}'$ the set of B-sets on the straight line. For simplicity, we shall assume that X is the unit of the domain of definition S_μ of the measure μ. Since, according to the results of §6, every σ-additive measure can be continued to some Borel algebra, it is natural to assume from the beginning that S_μ is a B-algebra. Therefore we shall formulate the definition of measurability for real functions in the following way:

Definition 1. The real function $f(x)$, defined on the set X, is called μ-measurable if for any Borel set A of the real line

$$f^{-1}(A) \in S_\mu.$$

Let us denote the set of those $x \in X$ for which condition Q is satisfied by $\{x : Q\}$.

Theorem 1. *In order that the function $f(x)$ be μ-measurable, it is necessary and sufficient that for any real C the set $\{x : f(x) < c\}$ be μ-measurable (i.e., belongs to S_μ).*

Proof. The necessity of the condition is clear, since the half line $(-\infty, c)$ is a Borel set. To prove the sufficiency let us first of all note that the Borel closure $B(\Sigma)$ of the system Σ of all half lines $(-\infty, c)$ coincides with the system B^1 of all Borel sets of the real axis. By assumption, $f^{-1}(\Sigma) \subseteq S_\mu$. But then

$$f^{-1}(B(\Sigma)) = B(f^{-1}(\Sigma)) \subseteq B(S_\mu),$$

and, since by assumption S_μ is a B-algebra, $B(S_\mu) = S_\mu$. The theorem is thus proved.

Theorem 2. *The limit of a sequence of μ-measurable functions which converges for every $x \in X$ is μ-measurable.*

Proof. Let $f_n(x) \to f(x)$, then

$$\{x : f(x) < c\} = \bigcup_k \bigcup_n \bigcap_{m>n} \left\{x : f_m(x) < c - \frac{1}{k}\right\}. \tag{1}$$

Indeed, if $f(x) < c$, then there exists a k, such that $f(x) < c - 2/k$; moreover, for this k one can find an n large enough so that for $m \geq n$ the inequality

$$f_m(x) < c - \frac{1}{k}$$

is satisfied, and this means that x will enter the right-hand side of (1).

Conversely, if x belongs to the right-hand side of (1), then there exists a k, such that, for all sufficiently large m,

$$f_m(x) < c - \frac{1}{k};$$

but then $f(x) < c$; i.e., x enters the left-hand side of equation (1).

If the functions $f_n(x)$ are measurable, the sets

$$\left\{x : f_m(x) < c - \frac{1}{k}\right\}$$

belong to S_μ. Since S_μ is a Borel algebra, the set

$$\{x : f(x) < c\}$$

also belongs to S_μ, by (1), which proves that $f(x)$ is measurable.

For the further study of measurable functions it is convenient to represent each of them as a limit of a sequence of so-called simple functions.

Definition 2. The function $f(x)$ is called *simple* if it is μ-measurable and takes on not more than a countable number of values.

It is clear that the concept of a simple function depends on the choice of the measure μ.

The structure of simple functions is characterised by the following theorem:

Theorem 3. *The function $f(x)$, taking on not more than a countable number of values*

$$y_1, y_2, \cdots, y_n, \cdots,$$

is μ-measurable if and only if all the sets

$$A_n = \{x : f(x) = y_n\}$$

are μ-measurable.

Proof. The necessity of the condition is obvious, since every A_n is the inverse image of a set of one point $\{y_n\}$, and every set of one point is a Borel set. Sufficiency follows from the fact that by the conditions of the theorem the inverse image $f^{-1}(B)$ of any set $B \subseteq D^1$ is the union $\bigcup_{y_n \in B} A_n$ of not more than a countable number of measurable sets A_n, i.e., is measurable.

The further use of simple functions will be based on the following theorem.

Theorem 4. *In order that the function $f(x)$ be μ-measurable it is necessary and sufficient that it be representable as a limit of a uniformly convergent sequence of simple functions.*

Proof. The sufficiency is clear from Theorem 2. To show the necessity, let us consider an arbitrary measurable function $f(x)$, and let us set $f_n(x) = m/n$ if $m/n \leq f(x) < (m + 1)/n$ (here m are integers, and n are positive integers). It is clear that the functions $f_n(x)$ are simple; they converge uniformly to $f(x)$ as $n \to \infty$, since $|f(x) - f_n(x)| \leq 1/n$.

Theorem 5. *The sum of two μ-measurable functions is μ-measurable.*

Proof. Let us first show this assertion for simple functions. If $f(x)$ and $g(x)$ are two simple functions taking on the values

$$f_1, f_2, \cdots, f_n, \cdots$$

and

$$g_1, g_2, \cdots g_n, \cdots,$$

respectively, then their sum $h(x) = f(x) + g(x)$ can take on only the values $h = f_i + g_i$, where each of these values is taken on on the set

$$\{x : h(x) = h\} = \bigcup_{f_i + g_j = h} (\{x : f(x) = f_i\} \cap \{x : g(x) = g_j\}). \qquad (2)$$

The number of possible values h is finite or countable and the corresponding sets $\{x : h(x) = h\}$ are measurable, since the right-hand side of equation (2) is obviously a measurable set.

To prove the theorem for arbitrary measurable functions $f(x)$ and $g(x)$, let us consider sequences of simple functions $\{f_n(x)\}$ and $\{g_n(x)\}$ which converge to $f(x)$ and $g(x)$, respectively. Then the simple functions $f_n(x) + g_n(x)$ converge uniformly to the function $f(x) + g(x)$, which, by Theorem 4, is measurable.

Theorem 6. *A B-measurable function of a μ-measurable function is μ-measurable.*

Proof. Let $f(x) = \varphi[\psi(x)]$, where φ is Borel measurable and ψ is μ-measurable. If $A \subseteq D^1$ is an arbitrary μ-measurable set, then its inverse image $A' = \varphi^{-1}(A)$ is B-measurable, and the inverse image $A'' = \psi^{-1}(A')$ of the set A' is μ-measurable. Since $f^{-1}(A) = A''$, the function f is measurable.

The theorem just proved is applicable, in particular, in the case of continuous functions φ (they are always B-measurable).

Theorem 7. *The product of μ-measurable functions is μ-measurable.*

Proof. Since $fg = \frac{1}{4}[(f+g)^2 - (f-g)^2]$, the assertion follows from Theorems 5 and 6, and the fact that $\varphi(t) = t^2$ is a continuous function.

Exercise. Show that if $f(x)$ is measurable and does not take on the value zero, then $1/f(x)$ is also measurable.

In studying measurable functions one can often neglect the values of a function on a set of measure zero. In connection with this let us introduce the following

Definition. Two functions f and g, defined on one and the same measurable set E, are called *equivalent* (denoted by $f \sim g$), if

$$\mu\{x : f(x) \neq g(x)\} = 0.$$

One says that some property is satisfied *almost everywhere* on E, if it is satisfied at all points of E with the exception of points which form a set of measure zero. Thus we can say that two functions are equivalent if they coincide almost everywhere.

Theorem 8. *If two functions f and g which are continuous on some segment E are equivalent, then they coincide.*

Proof. Let us assume that, at some point $x_0, f(x_0) \neq g(x_0)$, i.e., $f(x_0) - g(x_0) \neq 0$. Since $f - g$ is a continuous function, one can find a neighbourhood of the point x_0, at all points of which the function $f - g$ is different from zero. This neighbourhood has a positive measure; thus

$$\mu\{x:f(x) \neq g(x)\} > 0,$$

i.e., the continuous functions f and g cannot be equivalent if they take on different values at at least one point.

It is obvious that the equivalence of two arbitrary measurable functions (i.e., in general, discontinuous) does not at all imply their identity; for example, the function which is equal to unity at rational points and equal to zero at irrational points is equivalent to the function which is identically zero.

Theorem 9. *The function $f(x)$ which is defined on some measurable set E and is equivalent on this set to some measurable function $g(x)$ is also measurable.*

Indeed, from the definition of equivalence it follows that

$$\{x:f(x) > a\} \quad \text{and} \quad \{x:g(x) > a\}$$

can differ from each other only by a set of measure zero; hence if the second is measurable, so is the first.

The definition of a measurable function given above is quite formal. In 1913, N. N. Luzin proved the following theorem which shows that measurable functions are functions which, in a well known sense, can be approximated by continuous functions.

Luzin's Theorem. *In order that the function $f(x)$ be measurable on the segment $[a, b]$ it is necessary and sufficient that for any $\epsilon > 0$ there exist a function $\varphi(x)$ which is continuous on $[a, b]$, and such that*

$$\mu\{x:f(x) \neq \varphi(x)\} \leq \varepsilon.$$

In other words, a measurable function can be made continuous, if one excludes from consideration its values on a set of arbitrarily small measure. This property which Luzin called the C-property can be taken as a definition of a measurable function.

9. Sequences of Measurable Functions. Different Types of Convergence.

Theorems 5 and 7 of the preceding section show that arithmetic operations on measurable functions again lead to measurable functions. According to Theorem 2 of §8, the class of measurable functions, in contrast to the set of continuous functions, is also closed with respect to the operation of going to the limit. For measurable functions it is expedient to introduce, aside from the usual convergence at every point, several other definitions of convergence. These definitions of convergence, their basic properties and the connections between them will be investigated in the present section.

Definition 1. The sequence of functions $f_n(x)$, defined on some space with measure X, is said to *converge almost everywhere* to the function $F(x)$, if

$$\lim_{n\to\infty} f_n(x) = F(x) \tag{1}$$

for almost all $x \in X$ (i.e., the set of those points at which (1) does not hold has measure zero).

Example. The sequence of functions $f_n(x) = (-x)^n$, defined on the segment $[0, 1]$, converges as $n \to \infty$ to the function $F(x) \equiv 0$ almost everywhere (precisely, everywhere with the exception of the point $x = 1$).

Theorem 2 of §8 admits of the following generalisation.

Theorem 1. *If the sequence of μ-measurable functions $f_n(x)$ converges to the function $F(x)$ almost everywhere, then $F(x)$ is also measurable.*

Proof. Let A be that set on which

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

By the condition, $\mu(E \setminus A) = 0$. The function $F(x)$ is measurable on A by Theorem 2, §8. Since on a set of measure zero every function is obviously measurable, $F(x)$ is measurable on $E\setminus A$; therefore, it is also measurable on the set E.

Exercise. Let the sequence of measurable functions $f_n(x)$ converge almost everywhere to some limit function $f(x)$. Show that the sequence $f_n(x)$ converges to $g(x)$ if and only if $g(x)$ is equivalent to $f(x)$.

The following important theorem, proved by D. F. Egorov, establishes the connection between the concept of convergence almost everywhere and uniform convergence.

Theorem 2. *Let the sequence of measurable functions $f_n(x)$ converge on E almost everywhere to $f(x)$. Then there exists for any $\delta > 0$ a measurable set $E_\delta \subset E$ such that*

1) $\mu(E_\delta) > \mu(E) - \delta$,
2) *on the set E_δ the sequence $f_n(x)$ converges uniformly to $f(x)$.*

Proof. By Theorem 1, the function $f(x)$ is measurable. Set

$$E_n^m = \bigcap_{i \geq n} \left\{x : |f_i(x) - f(x)| < \frac{1}{m}\right\}.$$

Thus E_n^m, for fixed m and n, denotes the set of those points x for which

$$|f_i(x) - f(x)| < \frac{1}{m}$$

for all $i \geq n$. Let

$$E^m = \bigcup_n E_n^m.$$

From the definition of the set E_n^m it is clear that for fixed m,

$$E_1^m \subseteq E_2^m \subseteq \cdots \subseteq E_n^m \subseteq \cdots,$$

therefore, because of the fact that the σ-additive measure is continuous, one can find for any m and any $\delta > 0$ an $n_0(m)$ such that

$$\mu(E^m \setminus E_{n_0(m)}^m) < \frac{\delta}{2^m}.$$

Let us set

$$E_\delta = \bigcap_m E_{n_0(m)}^m$$

and show that the E_δ so constructed satisfies the requirements of the theorem.

Let us first prove that on E_δ the sequence $\{f_i(x)\}$ converges uniformly to the function $f(x)$. This follows immediately from the fact that if $x \in E_\delta$, then for any m

$$|f_i(x) - f(x)| < \frac{1}{m} \quad \text{for} \quad i \geq n_0(m).$$

Let us now evaluate the measure of the set $E \setminus E_\delta$. For this let us note that for every m, $\mu(E \setminus E^m) = 0$. Indeed, if $x_0 \in E \setminus E^m$, then there exist arbitrarily large values of i for which

$$|f_i(x_0) - f(x_0)| \geq \frac{1}{m},$$

i.e., the sequence $\{f_n(x)\}$ does not converge to $f(x)$ at the point x_0. Since by assumption $\{f_n(x)\}$ converges to $f(x)$ almost everywhere,

$$\mu(E \setminus E^m) = 0.$$

This implies

$$\mu(E \setminus E^m_{n_0(m)}) = \mu(E^m \setminus E^m_{n_0(m)}) < \frac{\delta}{2^m}.$$

Hence

$$\mu(E \setminus E_\delta) = \mu(E \setminus \bigcap E^m_{n_0(m)}) = \mu\left(\bigcup_m (E \setminus E^m_{n_0(m)})\right)$$

$$\leq \sum_m \mu(E \setminus E^m_{n_0(m)}) < \sum_{m=1}^{\infty} \frac{\delta}{2^m} = \delta.$$

The theorem is thus proved.

Definition 2. One says that the sequence of measurable functions $f_n(x)$ *converges in measure* to the function $F(x)$, if for any $\sigma > 0$

$$\lim_{n \to \infty} \mu\{x : |f_n(x) - F(x)| \geq \sigma\} = 0.$$

Theorems 3 and 4 given below establish the connection between convergence almost everywhere and convergence in measure.

Theorem 3. *If a sequence of measurable functions $f_n(x)$ converges almost everywhere to some function $F(x)$, then it converges to the same limit function $F(x)$ in measure.*

Proof. Theorem 1 implies that the limit function $F(x)$ is measurable. Let A be that set (of measure zero) on which $f_n(x)$ does not converge to $F(x)$. Let, moreover,

$$E_k(\sigma) = \{x: |f_k(x) - F(x)| \geq \sigma\}, \qquad R_n(\sigma) = \bigcup_{k=n}^{\infty} E_k(\sigma),$$

$$M = \bigcap_{n=1}^{\infty} R_n(\sigma).$$

It is clear that all these sets are measurable. Since

$$R_1(\sigma) \supset R_2(\sigma) \supset \cdots,$$

we have, because the measure is continuous,

$$\mu(R_n(\sigma)) \to \mu(M) \quad \text{for} \quad n \to \infty.$$

Let us now check that

$$M \subseteq A. \tag{2}$$

Indeed, if $x_0 \notin A_1$, i.e., if

$$\lim_{n\to\infty} f_n(x_0) = F(x_0),$$

then, for a given $\sigma > 0$, we can find an n such that

$$|f_n(x_0) - F(x_0)| < \sigma,$$

i.e., $x_0 \notin E_n(\sigma)$ and hence $x_0 \notin M$.

But $\mu(A) = 0$, therefore (2) implies that $\mu(M) = 0$ and hence

$$\mu(R_n(\sigma)) \to 0 \quad \text{for} \quad n \to \infty,$$

since $E_n(\sigma) \subseteq R_n(\sigma)$. The theorem is proved.

It is not difficult to convince oneself by an example that the convergence in measure of a sequence of functions does not imply convergence almost everywhere. Indeed, let us define for each natural k on the open segment $(0, 1]$ k functions

$$f_1^{(k)}, f_2^{(k)}, \cdots f_k^{(k)}$$

by the following method:

$$f_i^{(k)}(x) = \begin{cases} 1 & \text{for } \dfrac{i-1}{k} < x \le \dfrac{i}{k}, \\ 0 & \text{for all other values of } x. \end{cases}$$

Numbering all these functions in order, we obtain a sequence which, as is easy to check, converges in measure to zero, but at the same time does not converge at a single point (prove this!).

Exercise. Let the sequence of measurable functions $f_n(x)$ converge in measure to some limit function $f(x)$. Show that the sequence will converge in measure to the function $g(x)$ if and only if $g(x)$ is equivalent to $f(x)$.

Even though the example given above shows that Theorem 3 cannot be completely reversed, the following theorem holds:

Theorem 4. *Let the sequence of measurable functions $f_n(x)$ converge in measure to $f(x)$. Then one can select from the sequence $\{f_n(x)\}$ a subsequence $\{f_{n_k}(x)\}$ which converges to $f(x)$ almost everywhere.*

Proof. Let $\varepsilon_1, \varepsilon_2, \cdots$ be some sequence of positive numbers for which

$$\lim_{n\to\infty} \varepsilon_n = 0,$$

and let the positive numbers $\eta_1, \eta_2, \cdots, \eta_n, \cdots$ be such that

$$\eta_1 + \eta_2 + \cdots$$

converges. Let us construct a sequence of indices

$$n_1 < n_2 < \cdots$$

in the following way: n_1 is a natural number such that

$$\mu\{x: |f_{n_1}(x) - f(x)| \ge \varepsilon_1\} < \eta_1$$

(such an n_1 certainly exists); further, let n_2 be a number such that

$$\mu\{x: |f_{n_2}(x) - f(x)| \ge \varepsilon_2\} < \eta_2, \qquad n_2 > n_1.$$

In general, let n_k be a number such that

$$\mu\{x: |f_{n_k}(x) - f(x)| \ge \varepsilon_k\} < \eta_k, \qquad n_k > n_{k-1}.$$

We shall show that the sequence we have constructed converges to $f(x)$ almost everywhere. Indeed, let

$$R_i = \bigcup_{k=i}^{\infty} \{x: |f_{n_k}(x) - f(x)| \ge \varepsilon_k\}, \qquad Q = \bigcap_{i=1}^{\infty} R_i.$$

Since

$$R_1 \supset R_2 \supset R_3 \supset \cdots \supset R_n \supset \cdots,$$

the continuity of the measure implies $\mu(R_i) \to \mu(Q)$.

On the other hand, it is clear that $\mu(R_i) < \sum_{k=i}^{\infty} \eta_k$, which yields $\mu(R_i) \to 0$ for $i \to \infty$. Since $\mu(R_i) \to 0$,

$$\mu(Q) = 0.$$

It remains to show that at all points of the set $E \setminus Q$, the relation

$$f_{n_k}(x) \to f(x)$$

holds. Let $x_0 \in E \setminus Q$. Then one can find an i_0 such that $x_0 \notin R_{i_0}$. This means that for all $k \geq i_0$

$$x_0 \notin \{x : |f_{n_k}(x) - f(x)| \geq \varepsilon_k\},$$

i.e.,

$$|f_{n_k}(x_0) - f(x_0)| < \varepsilon_k.$$

Since, by assumption, $\varepsilon_k \to 0$,

$$\lim_{n \to \infty} f_{n_k}(x_0) = f(x_0).$$

The theorem is proved.

CHAPTER III

THE LEBESGUE INTEGRAL

In the preceding chapter we investigated the basic properties of measurable functions, which are a quite wide generalisation of continuous functions. For measurable functions, the classical definition of an integral known from analysis and usually called the Riemann integral is generally not applicable. For example, the well known Dirichlet function which equals zero at irrational points is obviously measurable, but is not integrable in the sense of Riemann. Thus this concept of an integral turns out to be of little use with respect to measurable functions.

The reason for this is clear. Let us assume, for simplicity, that we are considering the functions on a segment. Introducing the concept of the Riemann integral, we split the segment on which the function $f(x)$ is given into small segments, and taking in each of these parts an arbitrary point ξ_k, we form the sum

$$\sum_k f(\xi_k)\,\Delta x_k.$$

Essentially, we replace here the value of the function $f(x)$ at every point of the segment $\Delta x_k = [x_k, x_{k+1}]$ by its value at some arbitrarily selected point ξ_k of this interval. However, it is only natural to do this if the values of the function $f(x)$ at neighbouring points are close to each other, i.e., if $f(x)$ is continuous or if the set of its points of discontinuity is "not too large".*

The basic idea of the Lebesgue integral consists of the fact that, as opposed to the Riemann integral, the points x are grouped not by their closeness on the x-axis, but by the closeness of the values of the functions at these points. This immediately gives rise to

* As is well known, a bounded function is integrable in the sense of Riemann if and only if the set of its points of discontinuity has measure zero.

the possibility of generalising the concept of the integral to a quite wide class of functions.

Moreover, the Lebesgue integral is defined in exactly the same way for functions which are defined in any spaces with measures, whereas the Riemann integral is first introduced for functions of one variable, and only then is it carried over with corresponding changes to the case of several variables.

Everywhere, where the contrary is not especially stated, we shall consider some σ-additive measure $\mu(A)$ which is defined on a Borel algebra of sets with a unit X. All the sets $A \subseteq X$ considered will be assumed to be σ-measurable, and the functions $f(x)$ to be defined for $x \in X$ and μ-measurable.

10. The Lebesgue Integral for Simple Functions

We shall first introduce the concept of a Lebesgue integral for functions which we called simple in the preceding section, i.e., for measurable functions which take on a finite or countable number of values.

Let $f(x)$ be some simple function which takes on the values

$$y_1, y_2, \cdots, y_n, \cdots, \quad y_i \neq y_j \quad \text{for} \quad i \neq j.$$

It is natural to define the integral of the function $f(x)$ over the set A by the equation

$$\int_A f(x)\, d\mu = \sum_n y_n \mu\{x : x \in A, f(x) = y_n\}. \tag{1}$$

Thus we arrive at the following definition.

Definition. The simple function $f(x)$ is called *integrable* (with respect to the measure μ) over the set A if the sequence (1) converges absolutely. If $f(x)$ is integrable, then the sum of the series (1) is called the integral of $f(x)$ over the set A.

In this definition we assume that all the y_n are different. One can, however, represent the value of the integral of a simple function as a sum of products of the form $c_k\mu(B_k)$ and not assume that all the c_k are different. The following lemma allows us to do this:

Lemma. *Let* $A = \bigcup_k B_k$, $B_i \cap B_j = \varnothing$ *for* $i \neq j$, *and assume that on each set* B_k *the function* $f(x)$ *takes on only one value* c_k. *Then*

$$\int_A f(x)\,d\mu = \sum_k c_k\mu(B_k); \tag{2}$$

moreover, the function $f(x)$ *is integrable over* A *if and only if the series* (2) *converges absolutely.*

Proof. It is easy to see that every set

$$A_n = \{x : x \in A, f(x) = y_n\}$$

is the union of those B_k for which $c_k = y_n$. Therefore

$$\sum_n y_n\mu(A_n) = \sum_n y_n \sum_{c_k=y_n} \mu(B_k) = \sum_k c_k\mu(B_k).$$

Since the measure is non-negative,

$$\sum_n |y_n|\,\mu(A_n) = \sum_n |y_n| \sum_{c_k=y_n} \mu(B_k) = \sum_k |c_k|\,\mu(B_k),$$

i.e., the series $\sum_n y_n\mu(A_n)$ and $\sum_k c_k\mu(B_k)$ both either converge absolutely or diverge.

Let us establish some properties of the Lebesgue integral for simple functions:

A) $$\int_A f(x)\,d\mu + \int_A g(x)\,d\mu = \int_A \{f(x) + g(x)\}\,d\mu;$$

moreover, from the existence of the integrals on the left-hand side it follows that the integrals on the right-hand side exist.

To prove this let us assume that $f(x)$ takes on the values f_i on the sets $F_i \subseteq A$, and $g(x)$ the values g_i on the sets $G_i \subseteq A$, since

$$J_1 = \int_A f(x)\,d\mu = \sum_i f_i\mu(F_i), \tag{3}$$

$$J_2 = \int_A g(x)\,d\mu = \sum_j g_j\mu(G_j). \tag{4}$$

Then, by the lemma,

$$J = \int_A \{f(x) + g(x)\}\, d\mu = \sum_i \sum_j (f_i + g_j)\mu(F_i \cap G_j); \quad (5)$$

but

$$\mu(F_i) = \sum_j \mu(F_i \cap G_j),$$

$$\mu(G_j) = \sum_i \mu(F_i \cap G_j).$$

Hence, from the absolute convergence of the series (3) and (4), there follows the absolute convergence of the series (5); here

$$J = J_1 + J_2.$$

B) For any constant k,

$$k \int_A f(x)\, d\mu = \int_A \{kf(x)\}\, d\mu;$$

moreover, the existence of the integral on the left-hand side implies the existence of the integral on the right. (This can be shown immediately.)

C) A simple function $f(x)$ which is bounded on the set A is integrable over A; moreover, if $|f(x)| \leq M$ on A, then

$$\left| \int_A f(x)\, d\mu \right| \leq M\mu(A).$$

(This can be shown immediately.)

11. General Definition and Basic Properties of the Lebesgue Integral

Definition. We shall say that the function $f(x)$ is *integrable* over the set A, if there exists a sequence of simple functions $f_n(x)$ which are integrable over A and converge uniformly to $f(x)$. We shall denote the limit

$$J = \lim_{n\to\infty} \int_A f_n(x)\, d\mu \quad (1)$$

by

$$\int_A f(x)\, d\mu$$

and call it the integral of the function $f(x)$ over the set A.

This definition is correct if the following conditions are satisfied:

1. The limit (1) for any uniformly convergent sequence of simple functions which are integrable over A exists.
2. This limit for a given function $f(x)$ does not depend on the choice of the sequence $\{f_n(x)\}$.
3. For simple functions the definitions of integrability and integral are equivalent to those given in §10.

All these conditions are indeed satisfied.

To prove the first it suffices to note that by properties A), B) and C) for integrals of simple functions,

$$\left| \int_A f_n(x)\, d\mu - \int_A f_m(x)\, d\mu \right| \leq \mu(A) \sup_{x \in A} |f_n(x) - f_m(x)|.$$

To prove the second condition, we must consider the two sequences $\{f_n(x)\}$ and $\{f_n^*(x)\}$, and use the fact that

$$\left| \int_A f_n(x)\, d\mu - \int_A f_n^*(x)\, d\mu \right|$$

$$\leq \mu(A) \left\{ \sup_{x \in A} |f_n(x) - f(x)| + \sup_{x \in A} |f_n^*(x) - f(x)| \right\}.$$

Finally, to prove the third condition it suffices to consider the sequence $f_n(x) = f(x)$.

Let us establish the basic properties of the Lebesgue integral.

Theorem 1.

$$\int_A 1 \cdot d\mu = \mu(A).$$

Proof. It follows immediately from the definition of the integral.

Theorem 2. *For any constant k,*

$$k \int_A f(x)\, d\mu = \int_A \{kf(x)\}\, d\mu,$$

where the existence of the integral on the left-hand side implies the existence of the integral on the right.

Proof. The proof is obtained from property B) by proceeding to the limit for an integral of simple functions.

Theorem 3.

$$\int_A f(x)\, d\mu + \int_A g(x)\, d\mu = \int_A \{f(x) + g(x)\}\, d\mu,$$

where the existence of the integral on the left implies the existence of the integral on the right.

Proof. The proof is obtained from property A) by proceeding to the limit for an integral of simple functions.

Theorem 4. *A function $f(x)$ which is bounded on the set A is integrable over A.*

Proof. The proof is obtained from property C) by proceeding to the limit for an integral of simple functions.

Theorem 5. *If $f(x) \geq 0$, then*

$$\int_A f(x)\, d\mu \geq 0$$

(assuming that the integral exists).

Proof. For simple functions this follows immediately from the definition; for the general case the proof is based on the possibility of approximating non-negative functions by non-negative simple functions (for example by the method, given in the proof of Theorem 4 of §9).

Corollary 1. *If $f(x) \geq g(x)$, then*

$$\int_A f(x)\, d\mu \geq \int_A g(x)\, d\mu.$$

Corollary 2. *If on A, $m \le f(x) \le M$, then*

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

Theorem 6. *If*

$$A = \bigcup_n A_n, \qquad A_i \cap A_j = \varnothing \quad \text{for} \quad i \ne j,$$

then

$$\int_A f(x)\, d\mu = \sum_n \int_{A_n} f(x)\, d\mu;$$

moreover, the existence of the integral on the left-hand side implies the existence of the integrals and the absolute convergence of the series on the right.

Proof. Let us first check the assertion of the theorem for a simple function $f(x)$, which takes on the values

$$y_1, y_2, \cdots y_k, \cdots.$$

Let

$$B_k = \{x : x \in A, f(x) = y_k\},$$

$$B_{nk} = \{x : x \in A_n, f(x) = y_k\};$$

then

$$\int_A f(x)\, d\mu = \sum_k y_k\, \mu(B_k) = \sum_k y_k \sum_n \mu(B_{nk})$$

$$= \sum_n \sum_k y_k\, \mu(B_{nk}) = \sum_n \int_{A_n} f(x)\, d\mu. \tag{1}$$

Since, under the assumption of integrability of $f(x)$ over A, the series $\sum_k y_k \mu(B_k)$ converges absolutely, and the measure is non-negative, all the other series of the chain of equations (1) also converge absolutely.

In the case of an arbitrary function $f(x)$, its integrability over A implies that for any $\varepsilon > 0$ there exists a simple function $g(x)$

which is integrable over A and satisfies the condition

$$|f(x) - g(x)| < \varepsilon. \tag{2}$$

For $g(x)$,

$$\int_A g(x)\, d\mu = \sum_n \int_{A_n} g(x)\, d\mu; \tag{3}$$

moreover, $g(x)$ is integrable over every set A_n and the series (3) converges absolutely. This last fact and the estimate (2) imply that $f(x)$ is also integrable over every A_n, and

$$\sum_n \left| \int_{A_n} f(x)\, d\mu - \int_{A_n} g(x)\, d\mu \right| \leq \sum_n \varepsilon\mu(A_n) \leq \varepsilon\mu(A),$$

$$\left| \int_A f(x)\, d\mu - \int_A g(x)\, d\mu \right| \leq \varepsilon\mu(A),$$

which together with (3) yields the absolute convergence of the series $\sum_n \int_{A_n} f(x) d\mu$ and the estimate

$$\left| \sum_n \int_{A_n} f(x)\, d\mu - \int_A f(x)\, d\mu \right| \leq 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary,

$$\sum_n \int_{A_n} f(x)\, d\mu = \int_A f(x)\, d\mu.$$

Corollary. *If $f(x)$ is integrable over A, then $f(x)$ is also integrable over any $A' \subseteq A$.*

Theorem 7. *If the function $\varphi(x)$ is integrable over A and $|f(x)| \leq \varphi(x)$, then $f(x)$ is also integrable over A.*

Proof. If $f(x)$ and $\varphi(x)$ are simple functions, then A can be represented as a union of a finite or countable number of sets, on each of which $f(x)$ and $\varphi(x)$ are constant:

$$f(x) = a_n, \qquad \varphi(x) = \alpha_n, \quad \text{where} \quad |a_n| \leq \alpha_n.$$

Since $\varphi(x)$ is integrable, we have

$$\sum_n |a_n| \mu(A_n) \leq \sum_n \alpha_n \mu(A_n) = \int_A \varphi(x)\, d\mu.$$

Therefore $f(x)$ is also integrable and

$$\left| \int_A f(x)\, d\mu \right| = \left| \sum_n a_n \mu(A_n) \right| \leq \sum_n |a_n| \mu(A_n) = \int_A |f(x)|\, d\mu.$$

For the general case the theorem is proved by proceeding to the limit.

Theorem 8. *The integrals*

$$J_1 = \int_A f(x)\, d\mu, \qquad J_2 = \int_A |f(x)|\, d\mu$$

exist or do not exist simultaneously.

Proof. From the existence of the integral J_2 there follows the existence of the integral J_1 by Theorem 7.

The converse follows, in the case of a simple function, from the definition of the integral, and in the general case it is proved by going to the limit and using the fact that always

$$\|a| - |b\| \leq |a - b|.$$

Theorem 9. (Tchebichev Inequality) *If* $\varphi(x) \geq 0$ *on* A, *then*

$$\mu\{x : x \in A, \varphi(x) \geq c\} \leq \frac{1}{c} \int_A \varphi(x)\, d\mu.$$

Proof. Setting

$$A' = \{x : x \in A, \varphi(x) \geq c\},$$

we have

$$\int_A \varphi(x)\, d\mu = \int_{A'} \varphi(x)\, d\mu + \int_{A \setminus A'} \varphi(x)\, d\mu \geq \int_{A'} \varphi(x)\, d\mu \geq c\mu(A').$$

Corollary. *If*

$$\int_A |f(x)|\,d\mu = 0,$$

then $f(x) = 0$ *almost everywhere.*

Indeed, by the Tchebichev inequality, we have

$$\mu\left\{x : x \in A,\ |f(x)| \geq \frac{1}{n}\right\} \leq n \int_A |f(x)|\,d\mu = 0$$

for all n. Therefore

$$\mu\{x : x \in A, f(x) \neq 0\} \leq \sum_{n=1}^{\infty} \mu\left\{x : x \in A,\ |f(x)| \geq \frac{1}{n}\right\} = 0.$$

12. Limiting Processes Under the Lebesgue Integral Sign

The problem of proceeding to the limit under the integral sign, or, what is the same thing, of the possibility of term by term integration of a convergent series, is often encountered in various problems.

It was established in classical analysis that a sufficient condition for the possibility of such a limiting process is the uniform convergence of the corresponding series.

In this section we shall derive some theorems concerning limiting processes under the Lebesgue integral sign, which represent quite far reaching generalisations of the corresponding theorems of classical analysis.

Theorem 1. *If the sequence* $f_n(x)$ *on* A *converges to* $f(x)$ *and if for all* n

$$|f_n(x)| \leq \varphi(x),$$

where $\varphi(x)$ *is integrable on* A, *then the limit function* $f(x)$ *is integrable on* A *and*

$$\int_A f_n(x)\,d\mu \to \int_A f(x)\,d\mu.$$

Proof. From the conditions of the theorem it easily follows that $|f(x)| \leq \varphi(x)$. Let

$$A_k = \{x : k - 1 \leq \varphi(x) < k\};$$

and also let

$$B_m = \bigcup_{k \geq m} A_k = \{x : \varphi(x) \geq m\}.$$

By Theorem 6 of the preceding section

$$\int_A \varphi(x)\, d\mu = \sum_k \int_{A_k} \varphi(x)\, d\mu, \tag{$*$}$$

and the series $(*)$ converges absolutely.

Here

$$\int_{B_m} \varphi(x)\, d\mu = \sum_{k \geq m} \int_{A_k} \varphi(x)\, d\mu.$$

From the convergence of the series $(*)$ there follows the existence of an m such that

$$\int_{B_m} \varphi(x)\, d\mu < \frac{\varepsilon}{5}.$$

The inequality $\varphi(x) < m$ holds on $A \setminus B_m$. By Egorov's theorem, $A \setminus B_m$ can be represented in the form $A \setminus B_m = C \cup D$, where $\mu(D) < \varepsilon/5m$ and on the set C the sequence $\{f_n\}$ converges uniformly to f.

Let us select an N such that, for $n > N$, on the set C

$$|f_n(x) - f(x)| < \frac{\varepsilon}{5\mu(C)}.$$

Then

$$\int_A (f_n(x) - f(x))\, d\mu = \int_{B_m} f_n(x)\, d\mu - \int_{B_m} f(x)\, d\mu$$

$$+ \int_D f_n(x)\, d\mu - \int_D f(x)\, d\mu + \int_C (f_n(x) - f(x))\, d\mu < 5 \cdot \frac{\varepsilon}{5} = \varepsilon.$$

Corollary. *If* $|f_n(x)| \leq M$ *and* $f_n(x) \to f(x)$, *then*

$$\int_A f_n(x)\, d\mu \to \int_A f(x)\, d\mu.$$

Remark. In as much as the values which the function takes on on the set of measure zero do not influence the value of the integral, it suffices to assume in Theorem 1 that $\{f_n(x)\}$ converges to $f(x)$ almost everywhere.

Theorem 2. *Let, on the set* A,

$$f_1(x) \leq f_2(x) \leq \cdots \leq f_n(x) \leq \cdots,$$

where the functions $f_n(x)$ *are integrable and their integrals are not greater than a certain constant:*

$$\int_A f_n(x)\, d\mu \leq K.$$

Then almost everywhere on A *the limit*

$$f(x) = \lim_{n\to\infty} f_n(x) \tag{1}$$

exists, the function $f(x)$ *is integrable on* A, *and*

$$\int_A f_n(x)\, d\mu \to \int_A f(x)\, d\mu.$$

On a set on which the limit (1) does not exist, the function $f(x)$ can be given arbitrarily, for example assuming on this set $f(x) = 0$.

Proof. Let us assume $f_1(x) \geq 0$, since the general case can be easily reduced to this by going over to the functions

$$f_n(x) = f_n(x) - f_1(x).$$

Let us consider the set

$$\Omega = \{x : x \in A, f_n(x) \to \infty\}.$$

It is easy to see that $\Omega = \bigcap_r \bigcup_n \Omega_n^{(r)}$, where

$$\Omega_n^{(r)} = \{x : x \in A, f_n(x) > r\}.$$

By Tchebischev's inequalities (Theorem 9, §11),

$$\mu(\Omega_n^{(r)}) \leq \frac{K}{r}.$$

Since

$$\Omega_1^{(r)} \subseteq \Omega_2^{(r)} \subseteq \cdots \subseteq \Omega_n^{(r)} \subseteq \cdots,$$

we have $\mu\left(\bigcup_n \Omega_n^{(r)}\right) \leq K/r$, but from the fact that for any r

$$\Omega \subseteq \bigcup_n \Omega_n^{(r)},$$

it follows that $\mu(\Omega) \leq K/r$. Since r is arbitrary

$$\mu(\Omega) = 0.$$

This proves that the monotone sequence $f_n(x)$ has a finite limit almost everywhere on A.

Let us now set $\varphi(x) = r$ for those x for which

$$r - 1 \leq f(x) < r, \qquad r = 1, 2, \cdots.$$

If one can show that $\varphi(x)$ is integrable on A, then the assertion of our theorem will be an immediate consequence of Theorem 1.

Let us denote by A_r the set of those points $x \in A$ for which $\varphi(x) = r$ and let us set

$$B_s = \bigcup_{r=1}^{s} A_r.$$

Since on B_s the functions $f_n(x)$ and $f(x)$ are bounded, and since, always $\varphi(x) \leq f(x) + 1$,

$$\int_{B_s} \varphi(x)\, d\mu \leq \int_{B_s} f(x)\, d\mu + \mu(A)$$

$$= \lim_{n\to\infty} \int_{B_s} f_n(x)\, d\mu + \mu(A) \leq K + \mu(A).$$

But

$$\int_{B_s} \varphi(x)\, d\mu = \sum_{r=1}^{s} r\mu(A_r).$$

However, the boundedness of these sums means that the series

$$\sum_{r=1}^{\infty} r\mu(A_r) = \int_A \varphi(x)\, d\mu$$

converges. Thus, it is proved that $\varphi(x)$ is integrable on A.

Corollary. *If $\psi_n(x) \geq 0$ and*

$$\sum_{n=1}^{\infty} \int_A \psi_n(x)\, d\mu < \infty,$$

the series $\sum_{n=1}^{\infty} \psi_n(x)$ converges almost everywhere on A and

$$\int_A \left(\sum_{n=1}^{\infty} \psi_n(x)\right) d\mu = \sum_{n=1}^{\infty} \int_A \psi_n(x)\, d\mu.$$

Theorem 3. (Fatou) *If the sequence of measurable non-negative functions $\{f_n(x)\}$ converges almost everywhere on A to $f(x)$ and*

$$\int_A f_n(x)\, d\mu \leq K,$$

then $f(x)$ is integrable on A and

$$\int_A f(x)\, d\mu \leq K.$$

Proof. Let us set

$$\varphi_n(x) = \inf_{k \geq n} f_k(x);$$

$\varphi_n(x)$ is measurable because

$$\{x : \varphi_n(x) < c\} = \bigcup_{k \geq n} \{x : f_k(x) < c\}.$$

Moreover, $0 \leq \varphi_n(x) \leq f_n(x)$, hence $\varphi_n(x)$ is integrable and

$$\int_A \varphi_n(x)\, d\mu \leq \int_A f_n(x)\, d\mu \leq K.$$

Finally,

$$\varphi_1(x) \leq \varphi_2(x) \leq \cdots \leq \varphi_n(x) \leq \cdots,$$

and

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

almost everywhere. Therefore, applying the preceding result to $\{\varphi_n(x)\}$, we obtain the required result.

Theorem 4. *If* $A = \bigcup_n A_n$, $A_i \cap A_j = \varnothing$ *for* $i \neq j$ *and the series*

$$\sum_n \int_{A_n} |f(x)|\, d\mu \tag{2}$$

converges, then the function $f(x)$ *is integrable on* A *and*

$$\int_A f(x)\, d\mu = \sum_n \int_{A_n} f(x)\, d\mu.$$

New here, as compared to Theorem 6 of §11, is the assertion that from the convergence of the series (2) it follows that $f(x)$ is integrable on A.

Let us first give the proof for the case of a simple function which takes on the values f_i on the sets B_i. Setting

$$A_{ni} = A_n \cap B_i,$$

we have

$$\int_{A_n} |f(x)|\, d\mu = \sum_i |f_i|\, \mu(A_{ni}).$$

From the convergence of the series (2) there follows the convergence of the series

$$\sum_n \sum_i |f_i|\, \mu(A_{ni}) = \sum_i |f_i|\, \mu(B_i \cap A).$$

The convergence of the last series means that the integral

$$\int_A \varphi(x)\, d\mu = \sum_i f_i\, \mu(B_i \cap A)$$

exists.

In the general case we approximate the function $f(x)$ by the function $\tilde{f}(x)$ in such a way that

$$|f(x) - \tilde{f}(x)| < \varepsilon. \tag{3}$$

Then

$$\int_{A_n} |\tilde{f}(x)| d\mu \leq \int_{A_n} |f(x)| d\mu + \varepsilon\mu(A_n),$$

and, since the series

$$\sum_n \mu(A_n) = \mu(A)$$

converges, the convergence of the series (2) implies the convergence of the series

$$\sum_n \int_{A_n} |\tilde{f}(x)| d\mu,$$

i.e., by what has been just proved, the integrability over A of the simple function $\tilde{f}(x)$. But then, by (3), the initial function $f(x)$ is also integrable over A.

13. Comparison of the Lebesgue Integral and the Riemann Integral

Let us clarify the relation between the Lebesgue integral and the usual Riemann integral. Here we shall limit ourselves to the simplest case of a linear Lebesgue measure on a line.

Theorem. *If the Riemann integral*

$$J = (R) \int_a^b f(x)\, dx,$$

exists, then $f(x)$ is Lebesgue integrable on $[a, b]$ and

$$\int_{[a,b]} f(x)\, d\mu = J.$$

Proof. Let us consider the subdivision of $[a, b]$ into 2^n parts by the points

$$x_k = a + \frac{k}{2^n}(b - a)$$

and the Darboux sums, which correspond to this subdivision:

$$\bar{S}_n = \frac{b - a}{2^n} \sum_{k=1}^{2^n} M_{nk},$$

$$\underline{S}_n = \frac{b - a}{2^n} \sum_{k=1}^{2^n} m_{nk},$$

where M_{nk} is the upper bound of $f(x)$ on the segment

$$x_{k-1} \leq x \leq x_k,$$

and m_{nk} is the lower bound of $f(x)$ on the same segment. By the definition of the Riemann integral,

$$J = \lim_{n\to\infty} \bar{S}_n = \lim_{n\to\infty} \underline{S}_n.$$

Let us set

$$\bar{f}_n(x) = M_{nk} \quad \text{for} \quad x_{k-1} \leq x < x_k,$$

$$\underline{f}_n(x) = m_{nk} \quad \text{for} \quad x_{k-1} \leq x < x_k.$$

At the point $x = b$ the functions $\bar{f}_n$ and $\underline{f}_n$ can be defined arbitrarily. It is easy to compute that

$$\int_{[a,b]} \bar{f}_n(x)\, d\mu = \bar{S}_n,$$

$$\int_{[a,b]} \underline{f}_n(x)\, d\mu = \underline{S}_n.$$

Since the sequence $\{\bar{f}_n\}$ does not increase, and the sequence $\{\underline{f}_n\}$ does not decrease, we have almost everywhere

$$\bar{f}_n(x) \to \bar{f}(x) \geq f(x),$$

$$\underline{f}_n(x) \to \underline{f}(x) \leq f(x).$$

By Theorem 2, §12,

$$\int_{[a,b]} \bar{f}(x)\, d\mu = \lim_{n\to\infty} \bar{S}_n = J = \lim_{n\to\infty} \underline{S}_n = \int_{[a,b]} \underline{f}(x)\, d\mu.$$

Therefore

$$\int_{[a,b]} |\bar{f}(x) - \underline{f}(x)|\, d\mu = \int_{[a,b]} \{\bar{f}(x) - \underline{f}(x)\}\, d\mu = 0,$$

and, hence, almost everywhere

$$\bar{f}(x) - \underline{f}(x) = 0,$$

i.e.,

$$\bar{f}(x) = \underline{f}(x) = f(x),$$

$$\int_{[a,b]} f(x)\, d\mu = \int_{[a,b]} \bar{f}(x)\, d\mu = J.$$

The theorem is proved.

It is easy to give examples of bounded functions which are Lebesgue integrable but which are not Riemann integrable (for example, the Dirichlet function, mentioned earlier, which is equal to unity for rational x and equal to zero for irrational x).

In general, unbounded functions cannot be Riemann integrable, however many of them are Lebesgue integrable. In particular, any function $f(x)$ for which the Riemann integral

$$\int_{\varepsilon}^{1} |f(x)|\, dx$$

has a finite limit J as $\varepsilon \to 0$ is Lebesgue integrable on $[0, 1]$; moreover,

$$\int_{[0,1]} f(x)\, d\mu = \lim_{\varepsilon\to 0} \int_{\varepsilon}^{1} f(x)\, dx.$$

In connection with this it is interesting to mention that the improper integrals

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

are not Lebesgue integrable in the case when

$$\lim_{\varepsilon\to 0}\int_{\varepsilon}^{1}|f(x)|\,dx = \infty.$$

Lebesgue integration has always an absolute character in the sense of Theorem 8, §11.

14. Direct Products of Systems of Sets and Measures

An important role in analysis is played by theorems which reduce double (or generally multiple) integrals to iterated integrals. The basic result in the theory of multiple Lebesgue integrals is the so-called Fubini theorem which will be proved in §16. We shall preliminarily establish some useful concepts and facts which, by the way, are also of independent interest.

The set Z of ordered pairs (x, y), where $x \in X$ and $y \in Y$, is called a *direct product* of the sets X and Y and is denoted by $Z = X \times Y$. Analogously, the set U of ordered finite sequences $(x_1, x_2, \cdots, x_n)$, where $x_k \in X_k$, is called the direct product of the sets $X_1, X_2, \cdots, X_n$, and is denoted by

$$Z = X_1 \times X_2 \times \cdots \times X_n = \overline{|\times|}\, X_k.$$

In the special case when

$$X_1 = X_2 = \cdots = X_n = X,$$

the set Z is the n-th power of the set X:

$$Z = X^n.$$

For example, the n-dimensional coordinate space D^n is the n-th power of the real axis D^1. The unit cube J^n, i.e., the set of elements of D^n with coordinates which satisfy the inequality

$$0 \le x_k \le 1, \qquad k = 1, 2, \cdots, n,$$

is the n-th power of the unit segment $J^1 = [0, 1]$.

If $\mathfrak{S}_1, \mathfrak{S}_2, \cdots, \mathfrak{S}_n$ are systems of subsets of the sets $X_1, X_2, \cdots, X_n$, then

$$\mathfrak{R} = \mathfrak{S}_1 \times \mathfrak{S}_2 \times \cdots \times \mathfrak{S}_n$$

denotes the system of subsets of the set $X = \overline{|\times|} X_k$, which are given in the form

$$A = A_1 \times A_2 \times \cdots \times A_n,$$

where $A_k \in \mathfrak{S}_k$.

If $\mathfrak{S}_1 = \mathfrak{S}_2 = \cdots = \mathfrak{S}_n = \mathfrak{S}$, then $\mathfrak{R}$ is the n-th power of the system $\mathfrak{S}$:

$$\mathfrak{R} = \mathfrak{S}^n.$$

For example, the system of parallelopipeds in D^n is the n-th power of the system of segments in D^1.

Theorem 1. *If $\mathfrak{S}_1, \mathfrak{S}_2, \cdots, \mathfrak{S}_n$ are semirings, then $\mathfrak{R} = \overline{|\times|} \mathfrak{S}_k$ is also a semiring.*

Proof. In accordance with the definition of a semiring (§2), we have to show that if $A, B \in \mathfrak{R}$, then $A \cap B \in \mathfrak{R}$, and if, moreover, $B \subseteq A$, then $A = \bigcup_{i=1}^{m} C_i$, where $C_1 = B$, $C_i \cap C_j = \varnothing$ for $i \neq j$ and $C_i \in \mathfrak{R}$ $(i = 1, 2, \cdots, m)$.

Let us give the proof for the case $n = 2$.

I) Let $A \in \mathfrak{S}_1 \times \mathfrak{S}_2$, $B \in \mathfrak{S}_1 \times \mathfrak{S}_2$; this means that

$$A = A_1 \times A_2, \qquad A_1 \in \mathfrak{S}_1, \qquad A_2 \in \mathfrak{S}_2,$$

$$B = B_1 \times B_2, \qquad B_1 \in \mathfrak{S}_1, \qquad B_2 \in \mathfrak{S}_2.$$

Then

$$A \cap B = (A_1 \cap B_1) \times (A_2 \cap B_2),$$

and since

$$A_1 \cap B_1 \in \mathfrak{S}_1, \qquad A_2 \cap B_2 \in \mathfrak{S}_2,$$

we have

$$A \cap B \in \mathfrak{S}_1 \times \mathfrak{S}_2.$$

II) Let now, in the assumptions of paragraph I), $B \subseteq A$. Then

$$B_1 \subseteq A_1, \qquad B_2 \subseteq A_2,$$

and, because $\mathfrak{S}_1$ and $\mathfrak{S}_2$ are semirings, we have the decompositions

$$A_1 = B_1 \cup B_1^{(1)} \cup \cdots \cup B_1^{(k)},$$

$$A_2 = B_2 \cup B_2^{(1)} \cup \cdots \cup B_2^{(l)},$$

$$\begin{aligned} A = A_1 \times A_2 = &(B_1 \times B_2) \cup (B_1 \times B_2^{(1)}) \cup \cdots \cup (B_1 \times B_2^{(l)}) \\ &\cup (B_1^{(1)} \times B_2) \cup (B_1^{(1)} \times B_2^{(1)}) \cup \cdots \cup (B_1^{(1)} \times B_2^{(l)}) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\cup (B_1^{(k)} \times B_2) \cup (B_1^{(k)} \times B_2) \cup \cdots \cup (B_1^{(k)} \times B_2^{(l)}). \end{aligned}$$

In the last decomposition the first term is $B_1 \times B_2 = B$ and all terms belong to the system $\mathfrak{S}_1 \times \mathfrak{S}_2$. The theorem is proved.

However, from the assumption that the systems $\mathfrak{S}_k$ are rings or Borel rings, it does not yet follow that the product $\underset{k}{\boxed{\times}}\, \mathfrak{S}_k$ is a ring or a Borel ring, respectively.

Let the measures

$$\mu_1(A_1), \mu_2(A_2), \cdots, \mu_n(A_n), \qquad A_k \in \mathfrak{S}_k,$$

be given on the semirings $\mathfrak{S}_1, \mathfrak{S}_2, \cdots, \mathfrak{S}_n$.

Let us define on

$$\mathfrak{R} = \mathfrak{S}_1 \times \mathfrak{S}_2 \times \cdots \times \mathfrak{S}_n$$

the measure

$$\mu = \mu_1 \times \mu_2 \times \cdots \times \mu_n$$

by the condition: if $A = A_1 \times A_2 \times \cdots \times A_n$, then

$$\mu(A) = \mu_1(A_1)\mu_2(A_2) \cdots \mu_n(A_n).$$

We still have to show that $\mu(A)$ is a measure, i.e., that $\mu(A)$ is additive. We shall do this for the case $n = 2$. Assume the decomposition

$$A = A_1 \times A_2 = \bigcup B^{(k)}, \qquad B^{(i)} \cap B^{(j)} = \varnothing \quad \text{for} \quad i \neq j,$$
$$B^{(k)} = B_1^{(k)} \times B_2^{(k)}$$

given. As was shown in §2, there exist decompositions

$$A_1 = \bigcup_m C_1^{(m)}, \qquad A_2 = \bigcup_n C_2^{(n)}$$

such that the sets $B_1^{(k)}$ are unions of some $C_1^{(m)}$, and the sets $B_2^{(k)}$ are unions of some $C_2^{(m)}$. Obviously,

$$\mu(A) = \mu_1(A_1)\mu_2(A_2) = \sum_m \sum_n \mu_1(C_1^{(m)})\mu_2(C_2^{(n)}), \tag{1}$$

$$\mu(B^{(k)}) = \mu_1(B_1^{(k)})\mu_2(B_2^{(k)}) = \sum_m \sum_n \mu_1(C_1^{(m)})\mu_2(C_2^{(n)}); \tag{2}$$

moreover, all the terms which appear in the right-hand side of equation (2) appear once in the right-hand side of equation (1). Therefore,

$$\mu(A) = \sum_k \mu(B^{(k)}),$$

which was to be proved.

Thus, in particular, the additivity of elementary measures in the n-dimensional Euclidean space follows from the additivity of the linear measure on a line.

Theorem 2. *If the measures $\mu_1, \mu_2, \cdots, \mu_n$ are σ-additive, then the measure $\mu = \mu_1 \times \mu_2 \times \cdots \times \mu_n$ is also σ-additive.*

Proof. Let us prove the theorem for the case $n = 2$. Let us denote by λ_1 the Lebesgue space of measure μ_1. Let $C = \bigcup_{n=1}^{\infty} C_n$, where the sets C and C_n belong to $\mathfrak{S}_1 \times \mathfrak{S}_2$, i.e.,

$$C = A \times B, \qquad A \in \mathfrak{S}_1, \qquad B \in \mathfrak{S}_2,$$

$$C_n = A_n \times B_n, \qquad A_n \in \mathfrak{S}_1, \qquad B_n \in \mathfrak{S}_2.$$

Let us set, for $x \in X$,

$$f_n(x) = \begin{cases} \mu_2(B_n) & \text{if} \quad x \in A_n, \\ 0 & \text{if} \quad x \notin A_n. \end{cases}$$

It is easy to see that for $x \in A$,

$$\sum_n f_n(x) = \mu_2(B).$$

Therefore, by the corollary to Theorem 2, §12,

$$\sum_n \int_A f_n(x)\, d\lambda_1 = \int_A \mu_2(B)\, d\mu_1(A) = \mu(C);$$

but

$$\int_A f_n(x)\, d\lambda_1 = \mu_2(B_n)\mu_1(A_n) = \mu(C_n),$$

and, consequently

$$\sum_n \mu(C_n) = \mu(C).$$

The Lebesgue continuation of the measure $\mu_1 \times \mu_2 \times \cdots \times \mu_n$ we shall call the product of the measures μ_k and denote by

$$\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n = \overline{|\otimes|}\, \mu_k.$$

In the case

$$\mu_1 = \mu_2 = \cdots = \mu_n = \mu,$$

we obtain the concept of a power of the measure μ:

$$\mu^n = \overline{\underset{k=1}{\overset{n}{|\otimes|}}}\, \mu_k, \qquad \mu_k = \mu.$$

For example, the n-dimensional Lebesgue measure μ^n is the n-th power of the linear Lebesgue measure μ^1.

15. Expressing the Plane Measure by the Integral of a Linear Measure and the Geometric Definition of the Lebesgue Integral

Let a domain G in the (x, y)-plane be bounded by the vertical lines $x = a$, $y = b$ and by the curves $y = \varphi(x)$, $y = \psi(x)$.

As is well known, the area of the domain G is equal to the integral

$$V(G) = \int_a^b \{\varphi(x) - \psi(x)\}\, dx.$$

Moreover, the difference $\varphi(x_0) - \psi(x_0)$ is equal to the length of the intersection of the domain G with the vertical line $x = x_0$. Our task is to carry over such a method of measuring areas to arbitrary measure products

$$\mu = \mu_x \otimes \mu_y.$$

Here we shall assume that the measures μ_x and μ_y are defined on Borel rings, are σ-additive and have the closure property (if $B \subseteq A$ and $\mu(A) = 0$, then B is measurable) which, as was shown earlier, is possessed by all Lebesgue continuations.

Let us introduce the notation:

$$A_x = \{y: (x, y) \in A\},$$
$$A_y = \{x: (x, y) \in A\}.$$

If X and Y are real lines (and $X \times Y$ a plane), then A_{x_0} is the projection of an intersection of the set A and a vertical line $x = x_0$ onto the Y-axis.

Theorem 1. *Under the assumptions enumerated above, for any μ-measurable set A*,*

$$\mu(A) = \int_X \mu_y(A_x)\, d\mu_x = \int_Y \mu_x(A_y)\, d\mu_y.$$

Proof. It is obvious that it suffices to show the equality

$$\mu(A) = \int_X \varphi_A(x)\, d\mu_x, \quad \text{where} \quad \varphi_A(x) = \mu_y(A_x), \tag{1}$$

since the second part of the assertion of the theorem is quite analogous to the first. Let us note that the theorem automatically

---

* Note that integration over X is in fact reduced to integration over the set $\bigcup_y A_y \subset X$, outside of which the integrand is equal to zero. Analogously, $\int_y = \int_{\bigcup_x A_x}$.

includes the statement that for almost all x (in the sense of the μ_x measure) the sets A_x are measurable with respect to the measure μ_y, and that the function $\varphi_A(x)$ is measurable with respect to the measure μ_x. Otherwise formula (1) would not make sense.

The measure μ is a Lebesgue continuation of the measure

$$m = \mu_x \times \mu_y,$$

which is defined on the system S_m of sets of the form

$$A = A_{y_0} \times A_{x_0}.$$

For such sets equation (1) is obvious, since for them

$$\varphi_A(x) = \begin{cases} \mu_y(A_{x_0}) & \text{for} \quad x \in A_{y_0}, \\ 0 & \text{for} \quad x \notin A_{y_0}. \end{cases}$$

It is easy to see that equation (1) applies also to sets from $\mathfrak{R}(S_m)$, which decompose into a finite number of pairwise non-intersecting sets from S_m.

The proof of equation (1) for the general case is based on the following lemma, which is also of independent interest for the theory of Lebesgue continuations.

Lemma. *For any μ-measurable set A there exists a set B of the form*

$$B = \bigcap_n B_n, \qquad B_1 \supseteq B_2 \supseteq \cdots \supseteq B_n \supseteq \cdots,$$

$$B_n = \bigcup_k B_{nk}, \qquad B_{n1} \subseteq B_{n2} \subseteq \cdots \subseteq B_{nk} \subseteq \cdots,$$

where the sets B_{nk} belong to $\mathfrak{R}(S_m)$; moreover, $A \subseteq B$ and

$$\mu(A) = \mu(B). \tag{2}$$

The proof of the lemma is based on the fact that, by the definition of measurability, for any n the set A can be included in the union $C_n = \bigcup_r \Delta_{nr}$ of the sets Δ_{nr} of S_m in such a way that

$$\mu(C_n) < \mu(A) + \frac{1}{n}.$$

Setting $B_n = \bigcap_{k=1}^{n} C_k$, it is easy to see that the sets B_n will have the form $B_n = \bigcup_s \delta_{ns}$, where δ_{ns} belongs to S_m. Finally, setting $B_{nk} = \bigcup_{s=1}^{k} \delta_{ns}$, we obtain the system of sets B_{nk} required by the lemma.

It is easy to carry over equation (1) from the sets $B_{nk} \in \Re(S_m)$ to the sets B_n and B with the help of Theorem 2, §12, since

$$\varphi_{B_n}(x) = \lim_{k\to\infty} \varphi_{B_{nk}}(x), \qquad \varphi_{B_{n1}} \leq \varphi_{B_{n2}} \leq \cdots,$$

$$\varphi_B(x) = \lim_{n\to\infty} \varphi_{B_n}(x), \qquad \varphi_{B_1} \geq \varphi_{B_2} \geq \cdots.$$

If $\mu(A) = 0$, then $\mu(B) = 0$ and almost everywhere

$$\varphi_B(x) = \mu_y(B_x) = 0.$$

Since $A_x \subseteq B_x$, the set A_x is measurable for almost all x, and

$$\varphi_A(x) = \mu_y(A_x) = 0,$$

$$\int \varphi_A(x)\, d\mu_x = 0 = \mu(A).$$

Consequently, for the sets A for which $\mu(A) = 0$, formula (1) is true. Let us represent A if it is arbitrary in the form $A = B \setminus C$, where, by (2),

$$\mu(C) = 0.$$

Since formula (1) holds for the sets B and C, it is easy to see that it also holds for the set A itself.

The proof of Theorem 1 is complete.

Let us now consider the special case when Y is the real axis, μ_y a linear Lebesgue measure and the set A the set of points (x, y) of the form

$$x \in M, \qquad 0 \leq y \leq f(x), \tag{3}$$

where M is some μ_x-measurable set and $f(x)$ an integrable non-negative function. In this case

$$\mu_y(A_x) = \begin{cases} f(x) & \text{for} \quad x \in M, \\ 0 & \text{for} \quad x \notin M, \end{cases}$$

and

$$\mu(A) = \int_M f(x)\, d\mu_x.$$

We have proved the following

Theorem 2. *The Lebesgue integral of a non-negative function $f(x)$ is equal to the measure $\mu = \mu_x \times \mu_y$ of the set A, defined by relation* (3).

In the case when X is also the real axis, the set M is a segment, and the function $f(x)$ is Riemann integrable, this theorem reduces to the known fact that the integral can be expressed as the area under the graph of the function.

16. Fubini's Theorem

Let us consider a triple product of the form

$$U = X \times Y \times Z. \tag{1}$$

We shall identify the point

$$(x, y, z) \in U$$

with the points

$$((x, y), z),$$

$$(x, (y, z))$$

of the products

$$(X \times Y) \times Z, \tag{2}$$

$$X \times (Y \times Z). \tag{3}$$

With this understanding one can consider the products (1), (2) and (3) as one and the same thing.

If the measures μ_x, μ_y, μ_z are given on X, Y, Z, then the measure

$$\mu_u = \mu_x \otimes \mu_y \otimes \mu_z$$

can be defined either as

$$\mu_u = (\mu_x \otimes \mu_y) \otimes \mu_z,$$

or as

$$\mu_u = \mu_x \otimes (\mu_y \otimes \mu_z).$$

The rigorous proof for the equivalence of these definitions is omitted here, although it is not difficult.

We shall give the application of these general constructions to the proof of the basic theorem of the theory of multiple integrals.

Fubini's Theorem. *Let the measures μ_x and μ_y be defined on Borel rings, σ-additive and complete; let, moreover,*

$$\mu = \mu_x \otimes \mu_y,$$

and let the function $f(x,y)$ be integrable with respect to the measure μ on the sets

$$A = A_{y_0} \times A_{x_0}.$$

*Then**

$$\int_A f(x, y)\, d\mu = \int_X \left(\int_{A_x} f(x, y)\, d\mu_y\right) d\mu_x = \int_Y \left(\int_{A_y} f(x, y) d\mu_x\right) d\mu_y. \quad (4)$$

Proof. The assertion of the theorem includes the existence of the inner integrals in parentheses for almost all values of the variable, with respect to which the integration in the parentheses is taken.

Let us first give the proof for the case $f(x,y) \geq 0$. For this purpose let us consider the triple product

$$U = X \times Y \times D^1,$$

where the third factor is the rèal axis, and the product of measures

$$\lambda = \mu_x \otimes \mu_y \otimes \mu^1 = \mu \otimes \mu^1,$$

where μ^1 is a linear Lebesgue measure.

---

* See footnote on page 83.

In U we define the subset W by the condition

$$(x, y, z) \in W,$$

if

$$x \in A_y, \qquad y \in A_x,$$
$$0 \leq z \leq f(x, y).$$

By Theorem 2 of §15,

$$\lambda(W) = \int_A f(x, y)\, d\mu. \tag{5}$$

On the other hand, by Theorem 1 of §15,

$$\lambda(W) = \int_X \xi(W_x)\, d\mu_x, \tag{6}$$

where $\xi = \mu_y \times \mu^1$ and W_x denotes the set of pairs (y, z) for which $(x, y, z) \in W$. Here, by Theorem 1 of §14,

$$\xi(W_x) = \int_{A_x} f(x, y)\, d\mu_y. \tag{7}$$

Combining (5), (6) and (7) we obtain

$$\int_A f(x, y)\, d\mu = \int_X \left(\int_{A_x} f(x, y)\, d\mu_y\right) d\mu_x,$$

which was to be proved.

The general case is reduced to the case $f(x, y) \geq 0$ with the help of the equations

$$f(x, y) = f^+(x, y) - f^-(x, y),$$

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

Remark 1. One can show that if the function $f(x, y)$ is μ-measurable the integral

$$\int_A f(x, y)\, d\mu$$

exists, provided

$$\int_X \left(\int_{A_x} |f(x, y)| \, d\mu_y \right) d\mu_x$$

exists.

Examples showing when equation (4) does not hold:

1) Let

$$A = [-1, 1]^2$$

and

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

Then

$$\int_{-1}^{1} f(x, y) \, dx = 0$$

for $y \neq 0$, and

$$\int_{-1}^{1} f(x, y) \, dy = 0$$

for $x \neq 0$. Therefore

$$\int_{-1}^{1} \left(\int_{-1}^{1} f(x, y) \, dx \right) dy = \int_{-1}^{1} \left(\int_{-1}^{1} f(x, y) \, dy \right) dx = 0;$$

however, the integral does not exist in the sense of a Lebesgue integral over the square, since

$$\int_{-1}^{1} \int_{-1}^{1} |f(x, y)| \, dx \, dy \geq \int_0^1 dr \int_0^{2\pi} \frac{\sin \varphi \cos \varphi}{r} \, d\varphi = 2 \int_0^1 \frac{dr}{r} = \infty.$$

2) $A = [0, 1]^2$,

$$f(x, y) = \begin{cases} 2^{2n} & \text{for } \dfrac{1}{2^n} \leq x < \dfrac{1}{2^{n-1}}, \quad \dfrac{1}{2^n} \leq y < \dfrac{1}{2^{n-1}}, \\[2ex] -2^{2n+1} & \text{for } \dfrac{1}{2^{n+1}} \leq x < \dfrac{1}{2^n}, \quad \dfrac{1}{2^n} \leq y < \dfrac{1}{2^{n-1}}, \\[2ex] 0 & \text{for all other cases.} \end{cases}$$

One can compute that

$$\int_0^1 \left(\int_0^1 f(x, y)\, dx \right) dy = 0, \qquad \int_0^1 \left(\int_0^1 f(x, y)\, dy \right) dx = \frac{1}{4}.$$

17. The Integral as a Set Function

Let us consider the integral $F(A) = \int_A f(x)dx$ as a set function, assuming that S_μ is a Borel algebra with unit X and that $\int_X f(x)dx$ exists. Then, as was shown earlier:

1. $F(A)$ is defined on a Borel algebra S_μ;
2. $F(A)$ is real;
3. $F(A)$ is additive, i.e., for every decomposition

$$A = \bigcup_n A_n$$

of the set $A \in S_\mu$ into the sets $A_n \in S_\mu$,

$$F(A) = \sum_n F(A_n);$$

4. $F(A)$ is absolutely continuous, i.e., $\mu(A) = 0$ implies $F(A) = 0$.

Radon's Theorem. *If the function* $F(A)$ *possesses the properties* 1, 2, 3 *and* 4, *then it may be represented in the form*

$$F(A) = \int_A f(x)\, d\mu.$$

We shall show that the function $f = dF/d\mu$ can be uniquely determined up to values on a set of measure zero. Indeed, if for all $A \in S_\mu$,

$$F(A) = \int_A f_1(x)\, d\mu = \int_A f_2(x)\, d\mu,$$

then for any n for the sets

$$A_n = \left\{x : f_2 - f_1 > \frac{1}{n}\right\}$$

we have

$$\mu(A_n) \leq n \int_{A_n} (f_1 - f_2)\, d\mu = 0.$$

Analogously, for

$$B_n = \left\{x : f_2 - f_1 > \frac{1}{n}\right\}$$

we have

$$\mu\,(B_n) = 0.$$

Since

$$\{x : f_1 \neq f_2\} = \bigcup_n A_n \cup \bigcup_m B_m,$$

$$\mu\{x : f_1 \neq f_2\} = 0.$$

Our assertion is proved.

CHAPTER IV

FUNCTIONS WHICH ARE SQUARE INTEGRABLE

One of the most important spaces among the various linear normed spaces which are encountered in functional analysis is the Hilbert Space. Its name is due to the German mathematician D. Hilbert who introduced this space in connection with investigations in the theory of integral equations. It is a natural infinite analogue of the n-dimensional Euclidean space. We have already made acquaintance with one of the possible generalisations of Hilbert space in Chapter III of Volume I, it is the space l_2 whose elements are sequences of numbers

$$x = (x_1, x_2, \cdots, x_n, \cdots),$$

which satisfy the relation

$$\sum_{n=1}^{\infty} x_i^2 < \infty.$$

The concept of the Lebesgue integral allows us to introduce another, in some cases more convenient, realisation of the same space—the space of square integrable functions. In this chapter we shall consider the definition and the basic properties of the space of square integrable functions and establish that it is isometric (with corresponding assumptions about the measure with respect to which the integration is performed) to the space l_2.

In the next chapter we shall give an axiomatic definition of Hilbert space.

18. The L_2 Space

Below we shall consider functions $f(x)$, defined on some set R for which a measure $\mu(E)$ is given satisfying the condition

$\mu(R) < \infty$. The functions $f(x)$ are assumed to be measurable and defined almost everywhere on R. We shall not distinguish between functions which are equivalent on R. Instead of $\int_R$ we shall, for the sake of conciseness, write $\int$.

Definition 1. One says that the function $f(x)$ is *square integrable (or summable) over* R if the integral

$$\int f^2(x)\, d\mu$$

exists (i.e., is finite). The set of all functions which are square integrable over R we shall denote by L_2.

We shall now establish the basic properties of such functions.

Theorem 1. *The product of two square integrable functions is an integrable function.*

The proof follows immediately from the inequality

$$|f(x)g(x)| \leq \frac{f^2(x) + g^2(x)}{2}$$

and the properties of the Lebesgue integral.

Corollary. *Every square integrable function* $f(x)$ *is integrable.*

Indeed, it suffices to set $g(x) \equiv 1$ in Theorem 1.

Theorem 2. *The sum of two* L_2 *functions is also in* L_2.

Proof. Indeed,

$$(f(x) + g(x))^2 \leq f^2(x) + 2|f(x)g(x)| + g^2(x),$$

and, by Theorem 1, each of the three functions on the right is summable.

Theorem 3. *If* $f(x) \in L_2$ *and* α *is an arbitrary number, then* $\alpha f(x) \in L_2$.

Proof. If $f \in L_2$, then

$$\int (\alpha f(x))^2 \, d\mu = \alpha^2 \int f^2(x) \, d\mu < \infty.$$

Theorems 2 and 3 show that linear combinations of L_2 functions again belong to L_2; moreover, it is clear that sums of L_2 functions and their products with numbers satisfy all the conditions 1–8, enumerated in the definition of a linear space (Chapter III, §24 of Volume I), in other words, the class L_2 of square integrable functions is a linear space.

Let us now define the scalar product of L_2 functions by setting

$$(f, g) = \int_R f(x) g(x) \, d\mu. \tag{1}$$

As is well known, one understands by a scalar product any real function of a pair of vectors of the linear space which satisfies the following conditions:

1) $(f, g) = (g, f)$,
2) $(f_1 + f_2, g) = (f_1, g) + (f_2, g)$,
3) $(\lambda f, g) = \lambda(f, g)$,
4) $(f, f) > 0$ if $f \neq 0$.

From the basic properties of the integral it immediately follows that expression (1) does indeed satisfy conditions 1–3. Moreover, since we have agreed not to distinguish between functions which are equivalent (and in particular, to take as a unit element the set of all functions in R which are equivalent to $f(x) \equiv 0$), condition 4 is also satisfied (see the corollary to Theorem 9, §11). This leads us to the following definition.

Definition 2. By L_2 space one understands a Euclidean space,* the elements of which are classes of equivalent square integrable functions; addition of elements and their multiplication by numbers are defined as the usual addition and multiplication of functions, and the scalar product by the formula

$$(f, g) = \int f(x) g(x) \, d\mu. \tag{1}$$

* A Euclidean space is a linear space in which a scalar product has been introduced.

In L_2, as in every Euclidean space, the Cauchy-Bunjakovskii* inequality, having in this case the form

$$\left(\int f(x)g(x)\,d\mu\right)^2 \leq \int f^2(x)\,d\mu \cdot \int g^2(x)\,d\mu, \tag{2}$$

and the triangle inequality, having the form

$$\sqrt{\int (f(x)+g(x))^2\,d\mu} \leq \sqrt{\int f^2(x)\,d\mu} + \sqrt{\int g^2(x)\,d\mu}, \tag{3}$$

are satisfied.

In particular, for $g(x) \equiv 1$, the Cauchy-Bunjakovskii inequality implies the following useful inequality:

$$\left(\int f(x)\,d\mu\right)^2 \leq \mu(R) \int f^2(x)\,d\mu. \tag{4}$$

Let us introduce a norm in L_2, setting

$$\|f\| = \sqrt{(f,f)} = \sqrt{\int f^2(x)\,d\mu}, \qquad f \in L_2. \tag{5}$$

Exercise. Assuming properties 1–4 of a scalar product, show that the norm, defined by equation (5), satisfies conditions 1–3 of the definition of the norm (§21 of Volume I).

The following theorem plays a very important role in many problems of analysis:

Theorem 4. *The L_2 space is complete.*

Proof. a) Let $\{f_n(x)\}$ be a fundamental sequence in L_2, i.e., let

$$\|f_n - f_m\| \to 0 \quad \text{for} \quad n, m \to \infty.$$

Then we can select a subsequence of indices $\{n_k\}$ in such a way that

$$\|f_{n_k} - f_{n_{k+1}}\| \leq \frac{1}{2^k}.$$

* Translator's Note: This is usually known as the "Schwarz" inequality.

This implies, by inequality (4), that

$$\int |f_{n_k}(x) - f_{n_{k+1}}(x)|\, d\mu \leq [\mu(R)]^{1/2} \left\{ \int (f_{n_k}(x) - f_{n_{k+1}}(x))^2\, d\mu \right\}^{1/2}$$

$$\leq \frac{1}{2^k} [\mu(R)]^{1/2}.$$

This inequality and Theorem 2, §12 (Corollary) yield the result that the series

$$|f_{n_1}(x)| + |f_{n_2}(x) - f_{n_1}(x)| + \cdots$$

converges almost everywhere on R. Then also the series

$$f_{n_1}(x) + f_{n_2}(x) - f_{n_1}(x)) + \cdots$$

converges almost everywhere on R to some function

$$f(x) = \lim_{k \to \infty} f_{n_k}(x). \tag{6}$$

Thus we have shown that if the sequence $\{f_n(x)\}$ is fundamental in L_2, then one can always find a subsequence which converges almost everywhere.

b) Let us now show that the function $f(x)$ given by equation (6) belongs to L_2 and that

$$\|f_n(x) - f(x)\| \to 0 \quad \text{for} \quad n \to \infty. \tag{7}$$

For any sufficiently large k and l we have

$$\int (f_{n_k}(x) - f_{n_l}(x))^2\, d\mu < \varepsilon.$$

Accordingly, by Theorem 3, §12, we can proceed to the limit under the integral sign for $l \to \infty$ in this inequality. We obtain,

$$\int (f_{nk}(x) - f(x))^2\, d\mu \leq \varepsilon,$$

which yields $f \in L_2$ and $f_{n_k} \to f$. But from the fact that the fundamental sequence contains a subsequence, converging to some limit, it follows that it converges to the same limit.* The theorem is proved.

* We understand here by convergence the fact that equation (7) is satisfied; concerning this see the beginning of §19.

19. Mean Convergence. Sets in L_2 which are Everywhere Complete.

Introducing a norm in L_2 we have also introduced a certain kind of new convergence for the square integrable functions, namely:

$$f_n \to f \qquad (\text{in } L_2)$$

means that

$$\lim_{n\to\infty} \int [f_n(x) - f(x)]^2 \, d\mu = 0.$$

Such a convergence of functions is called *convergence in the mean* or, more exactly, convergence in the mean square.

Let us find the connection between the concept of convergence in the mean and uniform convergence, as well as the concept of convergence almost everywhere which we introduced in Chapter II.

Theorem 1. *If the sequence $\{f_n(x)\}$ of functions in L_2 converges uniformly to $f(x)$, then $f(x) \in L_2$ and $\{f_n(x)\}$ converges to $f(x)$ in the mean.*

Proof. Let $\varepsilon > 0$. If n is sufficiently large, then

$$|f_n(x) - f(x)| < \varepsilon,$$

whence

$$\int (f_n(x) - f(x))^2 \, d\mu < \varepsilon^2 \mu(R).$$

This inequality immediately yields the assertion of the theorem.

From Theorem 1 it follows that if one can approximate any function $f \in L_2$ to an arbitrary degree of accuracy by functions $f_n \in M \subseteq L_2$ in the sense of uniform convergence, then one can use them to approximate any function from L_2 also in the sense of convergence in the mean.

Hence one can approximate any function $f \in L_2$ to any degree of accuracy by simple functions belonging to L_2.

We shall show that any simple function $f \in L_2$, and therefore, also in general any function from L_2, can be approximated as closely as desired by simple functions which take on only a finite number of values.

Let $f(x)$ take on the values $y_1, \cdots, y_n, \cdots$ on the sets $E_1, \cdots, E_n, \cdots$. Since f^2 is integrable, the series

$$\sum_n y_n^2 \mu(E_n) = \int f^2(x)\, d\mu$$

converges. Let us select a number N in such a way that

$$\sum_{n>N} y_n^2 \mu(E_n) < \varepsilon,$$

and set

$$f_N(x) = \begin{cases} f(x) & \text{for} \quad x \in E_i, \quad i \leq N, \\ 0 & \text{for} \quad x \in E_i, \quad i > N. \end{cases}$$

Then we have

$$\int (f(x) - f_N(x))^2\, d\mu = \sum_{n>N} y_n^2 \mu(E_n) < \varepsilon,$$

i.e., the functions f_N, which take on a finite number of values, approximate the function f as closely as required.

Let R be a metric space having a measure which satisfies the following condition (satisfied in all cases of practical interest): all open and all closed sets in R are measurable and for any $M \subseteq R$,

$$\mu^*(M) = \inf_{M \subseteq G} \mu(G), \tag{$*$}$$

where the lower bound is taken over all open sets G containing M. Then the following theorem holds.

Theorem 2. *The set of all continuous functions is complete in L_2.*

Proof. By what has been said above it suffices to show that every simple function which takes on a finite number of values is a limit, in the sense of mean convergence, of continuous functions. Furthermore, since every simple function which takes on a finite number of values is a linear combination of characteristic functions $\chi_M(x)$ of measurable sets, it suffices to give the proof for these latter ones. Let M be a measurable set in the metric space R. Then, from condition ($*$) it immediately follows that for any $\varepsilon > 0$ one can find a closed set F_M and an open set G_M for which

$$F_M \subset M \subset G_M \quad \text{and} \quad \mu(G_M) - \mu(F_M) < \varepsilon.$$

Let us now define the function $\varphi_\varepsilon(x)$ by setting

$$\varphi_\varepsilon^k(x) = \frac{\rho(x, R\setminus G_M)}{\rho(x, R\setminus G_M) + \rho(x, F_M)}.$$

This function equals 0 for $x \in R\setminus G_M$ and equals 1 for $x \in F_M$. It is continuous since each of the functions $\rho(x, F_M)$ and $\rho(x, R\setminus G_M)$ is continuous and since their sum is equal to zero. The function $\chi_M(x) - \varphi_\varepsilon(x)$ is not greater that unity on $G_M\setminus F$ and equals zero outside this set. Therefore

$$\int (\chi_M(x) - \varphi_\varepsilon(x))^2\, d\mu < \varepsilon,$$

which yields the assertion of the theorem.

Theorem 3. *If the sequence $\{f_n(x)\}$ converges to $f(x)$ in the mean, then one can select from it a subsequence $\{f_{n_k}(x)\}$ which converges to $f(x)$ almost everywhere.*

Proof. If the sequence $\{f_n(x)\}$ converges in the mean, then it is fundamental in L_2; hence, repeating the arguments of part a) of the proof of Theorem 4, §18, we see that one can select from $\{f_n(x)\}$ a subsequence $\{f_{n_k}(x)\}$ which converges almost everywhere to some function $\varphi(x)$. Furthermore, the arguments of part b) of the same proof show that $\{f_{n_k}(x)\}$ converges to $\varphi(x)$ also in the mean, which yields $\varphi(x) = f(x)$ almost everywhere.

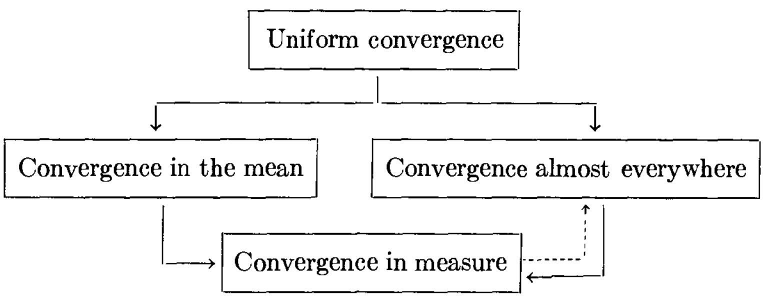
It is not difficult to convince oneself by examples that the convergence in the mean of some sequence does not imply that this sequence itself converges almost everywhere. Indeed, the sequence of functions f_{n_k} constructed on page 57 obviously converges in the mean to the function $f \equiv 0$ but still, as shown there, does not converge everywhere. We shall now show that convergence almost everywhere (and even everywhere) does not imply convergence in the mean. Let

$$f_n(x) = \begin{cases} n & \text{for} \quad x \in \left(0, \dfrac{1}{n}\right), \\ 0 & \text{for all other values of } x. \end{cases}$$

It is clear that the sequence $\{f_n(x)\}$ converges to 0 everywhere on $[0, 1]$, however at the same time

$$\int_0^1 f_n^2(x)\, dx = n \to \infty.$$

Tchebishev's inequality (§11, Theorem 9) implies that if the sequence converges in the mean, it converges in measure. Therefore, Theorem 3 which we proved here independently follows from Theorem 4, §9. The relations between the different types of convergence of functions can be represented by the following scheme:



where the dotted arrow means that one can select from the sequence which converges in measure a subsequence which converges almost everywhere.

20. L_2 Spaces with a Countable Basis

Generally speaking, the L_2 space of functions which are square integrable depends on the choice of the space R and the measure μ. Its full notation should be $L_2(R, \mu)$. Only in exceptional cases is the space $L_2(R, \mu)$ finite dimensional. Of more importance in analysis are those spaces $L_2(R, \mu)$ which, in a certain sense to be defined later, are of *countable dimension*.

In order to characterise these spaces we shall need yet another concept of measure theory.

In the set $\mathfrak{M}$ of measurable subsets of the space R (we assume its measure to be finite) we can introduce distance by setting

$$\rho(A, B) = \mu(A \triangle B).$$

If we identify those sets A and B for which $\mu(A \triangle B) = 0$ (i.e., in other words we consider not single sets but classes of sets which coincide up to a set of measure zero), then the set $\mathfrak{M}$ with the distance function $\rho(A, B)$ satisfies all the axioms of a metric space.

Definition. One says that the measure μ has *countable basis* if the metric space $\mathfrak{M}$ contains an everywhere countable dense set.

In other words, the measure μ has a countable basis if there exists a countable system

$$\Delta = \{A_n\}, \qquad n = 1, 2, \cdots,$$

of measurable subsets of the space R (countable basis of the measure μ) such that, for any measurable $M \subseteq R$ and any $\varepsilon > 0$, one can find an $A_k \in \Delta$ for which

$$\mu(M \triangle A_k) < \varepsilon.$$

In particular, the measure μ obviously has a countable basis if it can be represented as a continuation of a measure, defined on a countable system of sets S_m. Indeed, in this case the ring $\mathfrak{R}(S_m)$ (it is obviously countable) is, by Theorem 3, §6, the required basis.

In particular, the Lebesgue measure on a segment of the real axis is generated by a system of intervals, with rational end points, as elementary sets. Since the set of such intervals is countable, the Lebesgue measure has a countable basis.

The product $\mu = \mu_1 \times \mu_2$ of two measures with countable basis also has a countable basis, since the finite sums of pairwise products of elements of the basis of the measure μ_1 with elements of the basis of the measure μ_2 form, as is easily checked, the basis of the measure $\mu = \mu_1 \times \mu_2$. Therefore the Lebesgue measure of a plane (and also an n-dimensional space) has a countable basis.

Let

$$A_1^*, \quad A_2^*, \quad \cdots, \quad A_n^*, \quad \cdots \tag{1}$$

be the countable basis of the measure μ. It is easy to see, that, enlarging the system of sets (1), one can form a countable basis for the measure μ:

$$A_1, \quad A_2, \quad \cdots, \quad A_n, \quad \cdots, \tag{2}$$

which satisfies the following conditions:

1) the system of sets (2) is closed with respect to subtraction;

2) the system of sets (2) contains R.

From conditions 1) and 2) it follows that the system (2) is closed with respect to a finite number of unions and intersections of sets.

This follows from the obvious equalities

$$A_1 \cap A_2 = A_1 \setminus (A_1 \setminus A_2),$$

$$A_1 \cup A_2 = R \setminus \{(R \setminus A_1) \cap (R \setminus A_2)\}.$$

Theorem. *If the measure μ has a countable basis, then there exists in $L_2(R, \mu)$ an everywhere countable dense set of functions*

$$f_1, \quad \cdots, \quad f_n, \quad \cdots$$

Proof. As such a basis in $L_2(R, \mu)$ we can select the finite sums

$$\sum_{k=1}^{n} c_k f_k(x), \tag{3}$$

where c_k are rational numbers and $f_k(x)$ are the characteristic functions of the elements of the countable basis of the measure μ.

Indeed, as we have already shown in the preceding section, the set of step functions which take on only a finite number of different values is everywhere dense in L_2. Since it is clear that any function of this set can be approximated as closely as required by a function of the same kind but taking on only rational values, and since the set of functions of the form (3) is countable, it suffices, in order to prove our assertion, to show that any step function taking on the values

$$y_1, y_2, \cdots, y_n, \cdots \quad \text{(all } y_i \text{ are rational)},$$

on the sets

$$E_1, E_2, \cdots, E_n, \cdots, \qquad \bigcup_i E_i = R,\ E_i \cap E_j = \varnothing \quad \text{for } i \neq j,$$

can be approximated as closely as desired in the sense of the L_2 metric by a function of the form (3). By this remark, one can, without loss of generality, assume that our basis of the measure μ satisfies conditions 1) and 2).

By the definition of a countable basis of the measure μ, there exist, for any $\varepsilon > 0$, sets $A_1, A_2, \cdots, A_n, \cdots$ from our basis of measure μ, such that $\rho(E_k, A_k) < \varepsilon$, i.e.,

$$\mu[(E_k \setminus A_k) \cup (A_k \setminus E_k)] < \varepsilon.$$

Let us set

$$A_k' = A_k \setminus \bigcup_{i<k} A_i, \qquad k = 1, 2, \cdots, n,$$

and define

$$f^*(x) = \begin{cases} y_k & \text{for} \quad x \in A_k', \\ 0 & \text{for} \quad x \in R \setminus \bigcup_{i=1}^{n} A_i'. \end{cases}$$

It is easy to see that for sufficiently small ε the quantity

$$\mu\{x : f(x) \neq f^*(x)\}$$

is as small as required and, therefore, the integral

$$\int (f(x) - f^*(x))^2 \, d\mu \leq (2 \max |y_n|)^2 \mu\{x : f(x) \neq f^*(x)\}$$

is arbitrarily small for sufficiently small ε.

By virtue of our assumption concerning the basis of the measure μ, the function $f^*(x)$ is a function of the form (3). The theorem is proved.

For the special case when R is a segment of the real line and μ a Lebesgue measure, the countable basis in $L_2(R)$ can also be obtained by a more classical method: as such a basis we could take, for example, the set of all polynomials with rational coefficients. It is everywhere dense (even in the sense of uniform convergence) in the set of continuous functions, and these latter ones form an everywhere dense set in $L_2(R, \mu)$.

In the following we shall limit ourselves to the consideration of spaces $L_2(R, \mu)$ which have a countable everywhere dense set (in other words, are separable—see §9 of Volume I).

21. Orthogonal Systems of Functions. Orthogonalisation.

In this section we shall investigate functions $f \in L_2$ which are given on some measurable set R with measure μ; we shall assume that the measure has a countable basis and satisfies the condition $\mu(R) < \infty$. As before, we shall not distinguish between equivalent functions.

Definition 1. The system of functions

$$\varphi_1(x), \varphi_2(x), \cdots, \varphi_n(x) \tag{1}$$

is called *linearly dependent,* if there exist constants $c_1, c_2, \cdots, c_n$, not all equal to zero, such that

$$c_1\varphi_1(x) + c_2\varphi_2(x) + \cdots + c_n\varphi_n(x) = 0 \tag{2}$$

almost everywhere on R. If, however, (2) implies that

$$c_1 = c_2 = \cdots = c_n = 0, \tag{3}$$

then the system (1) is called *linearly independent.*

It is clear that a linearly independent system cannot contain a function which is equivalent to $\Psi(x) \equiv 0$.

Definition 2. The infinite system of functions

$$\varphi_1(x), \varphi_2(x), \cdots, \varphi_n(x), \cdots \tag{4}$$

is called linearly independent, if any finite part of it is linearly independent.

Let us denote by

$$M = M(\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots) = M\{\varphi_k\}$$

the set of all finite linear combinations of functions of the system (4). This set is called the *linear closure* of the system (4). By

$$\bar{M} = \bar{M}(\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots) = \bar{M}\{\varphi_k\}$$

we shall denote the closure of the set M in the space L_2. $\bar{M}$ is called the *closed linear closure* of the system (4).

It is easy to see that the set $\bar{M}$ consists of those and only those functions $f \in L_2$ which can be approximated by finite linear combinations of functions of the system (4) with a prescribed accuracy.

Definition 3. The system of functions (4) is called *complete*, if for it,

$$\bar{M} = L_2.$$

Let there exist in the space L_2 a countable, everywhere dense set of functions

$$f_1, f_2, \cdots, f_n, \cdots.$$

Discarding from this system those functions which are linearly dependent on the preceding ones, we arrive at a linearly independent system of functions

$$g_1, g_2, \cdots, g_n, \cdots$$

which, as is easily seen, is complete.

In the case that there exists in L_2 a finite system (1) of linearly independent functions,

$$L_2 = \bar{M}(\varphi_1, \varphi_2, \cdots, \varphi_n) = M(\varphi_1, \varphi_2, \cdots, \varphi_n)$$

is an n-dimensional Euclidean space.

In all cases which are interesting for analysis the space L_2 is infinite dimensional.

It is obvious that the system (4) is complete if one can approximate each of the functions, belonging to some set which is everywhere complete in L_2, by linear combinations of functions belonging to the system (4) with any desired degree of accuracy.

Let $R = [a, b]$ be a segment of the real axis with the usual Lebesgue measure. Then the system

$$1, x, x^2, \cdots, x^n, \cdots \tag{5}$$

is complete in the space $L_2(R, \mu)$.

Indeed, by Weierstrass' theorem, the linear combinations of functions (5) are complete in the set of all continuous functions. The fact that the system (5) is complete now follows from our remark and Theorem 2, §19.

The functions $f(x)$ and $g(x)$ are called *mutually orthogonal*, if

$$(f, g) = \int f(x)g(x)\, d\mu = 0.$$

Every system $\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots$ of functions from L_2 which are different from zero and are pairwise orthogonal we shall call an *orthogonal system.* An orthogonal system is called *normalised* if $\| \varphi_n \| = 1$ for all n; in other words,

$$\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots$$

is an orthonormal system of functions if

$$(\varphi_i, \varphi_k) = \int \varphi_i(x)\varphi_k(x)\, d\mu = \begin{cases} 0 & \text{for } i \neq k, \\ 1 & \text{for } i = k. \end{cases}$$

Examples. 1. A classical example of an orthonormal system of functions on the segment $[-\pi, \pi]$ is the trigometric system:

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{\cos x}{\sqrt{\pi}}, \quad \frac{\cos 2x}{\sqrt{\pi}}, \quad \cdots, \quad \frac{\sin x}{\sqrt{\pi}}, \quad \frac{\sin 2x}{\sqrt{\pi}}, \quad \cdots.$$

2. The polynomials

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n[(x^2 - 1)^n]}{dx^n}, \quad n = 0, 1, 2, \cdots,$$

which are called *Legendre polynomials* form an orthogonal system of functions on the segment $[-1, 1]$. The functions

$$\sqrt{\frac{2n + 1}{2}}\, P_n(x)$$

form an orthonormal system.

It is easy to see that an *orthogonal system of functions is linearly independent.* Indeed, multiplying the relation

$$c_1\varphi_1 + c_2\varphi_2 + \cdots + c_n\varphi_n = 0$$

by φ_i and integrating, we obtain

$$c_i(\varphi_i, \varphi_i) = 0,$$

and since $(\varphi_i, \varphi_i) > 0$, we get $c_i = 0$.

Let us further note that, *if in the space L_2 there exists an everywhere countable dense set $f_1, f_2, \cdots, f_n, \cdots$, then any orthonormal system of functions $\{\varphi_\alpha\}$ is at most countable.*

Indeed, let $\alpha \neq \beta$, then

$$||\varphi_\alpha - \varphi_\beta|| = \sqrt{2}.$$

For each α let us select from our everywhere dense set an f_α in such a way that

$$||\varphi_\alpha - f_\alpha|| < \frac{1}{\sqrt{2}}.$$

It is clear that $f_\alpha \neq f_\beta$ if $\alpha \neq \beta$, and, since the set of all the f_α is countable, the set of the φ_α themselves is also not more than countable.

In studying finite-dimensional spaces an important role is played by the concept of the orthogonal normalised basis, i.e., the orthogonal system of unit vectors, the linear closure of which coincides with the whole space. In the infinite case the analogue of such a basis is the complete orthonormal system of functions, i.e., a system

$$\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots,$$

such that

1) $$(\varphi_i, \varphi_k) = \delta_{ik},$$

2) $$\bar{M}(\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots) = L_2.$$

Above we have given examples of complete orthonormal systems of functions on the segments $[-\pi, \pi]$ and $[-1, 1]$. The existence of a complete orthonormal system of functions in any separable space L_2 follows from the following theorem:

Theorem. *Let the system of functions*

$$f_1, f_2, \cdots, f_n, \cdots \tag{6}$$

be linearly independent. Then there exists a system of functions

$$\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots \tag{7}$$

satisfying the following conditions:

1) *the system* (7) *is orthonormal;*

2) *every function* φ_n *is a linear combination of functions* $f_1, f_2, \cdots, f_n$:

$$\varphi_n = a_{n1} f_1 + a_{n2} f_2 + \cdots + a_{nn} f_n,$$

where $a_{nn} \neq 0$;

3) *every function* f_n *is a linear combination of functions* $\varphi_1, \varphi_2, \cdots, \varphi_n$:

$$f_n = b_{n1}\varphi_1 + b_{n2}\varphi + \cdots + b_{nn}\varphi_n,$$

where $b_{nn} \neq 0$.

Every function of the system (2) *is uniquely determined (up to the sign) by conditions* 1)–3).

Proof. The function $\varphi_1(x)$ is uniquely determined (up to sign) by the conditions of the theorem. Indeed,

$$\varphi_1 = a_{11} f_1,$$

$$(\varphi_1, \varphi_1) = a_{11}^2 (f_1, f_1) = 1,$$

which yields

$$b_{11} = \frac{1}{a_{11}} = \sqrt{(f_1, f_1)} \quad \text{and} \quad \varphi_1 = \pm \frac{f_1}{\sqrt{(f_1, f_1)}}.$$

Let the functions $\varphi_k (k < n)$ satisfying conditions 1)–3) be already found. Then f_n can be represented in the form

$$f_n = b_{n1}\varphi_1 + \cdots + b_{n,n-1}\varphi_{n-1} + h_n,$$

where $(h_n, \varphi_k) = 0$ for $k < n$.

Obviously $(h_n, h_n) > 0$ (the assumption $(h_n, h_n) = 0$ would have resulted in a contradiction to the linear independence of the system (6)).

Let us set

$$\varphi_n = \frac{h_n}{\sqrt{(h_n, h_n)}}.$$

Then we have

$$(\varphi_n, \varphi_i) = 0, \qquad i < n,$$

$$(\varphi_n, \varphi_n) = 1,$$

$$f_n = b_n\varphi_1 + \cdots + b_{nn}\varphi_n, \qquad b_{nn} = \sqrt{(h_n, h_n)} \neq 0,$$

i.e., the function $\varphi_n(x)$ satisfies the conditions of the theorem.

The process of going over from the system (6) to the system (7) which satisfies conditions 1)–3) is called the *orthogonalisation process*.

Obviously,

$$M(f_1, f_2, \cdots, f_n, \cdots) = M(\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots)$$

and, therefore, the systems (6) and (7) are either both complete or both incomplete.

Thus, in any problem connected with the approximation of the function f by linear combinations of the functions (6), one can replace the system (6) by an orthonormal system (7) which has been obtained from (6) by an orthogonalisation process.

As we have already said earlier, from the existence of a countable everywhere dense set in L_2 there follows the existence of a countable dense system of linearly independent functions. Orthogonalising this system we obtain a complete dense countable orthonormal system.

22. Fourier Series on Orthogonal Systems. Riesz–Fischer Theorem.

Introducing in the n-dimensional Euclidean space $R^{(n)}$ an orthogonal normalised basis $e_1, e_2, \cdots, e_n$, we can write each vector $x \in R^{(n)}$ in the form

$$x = \sum_{k=1}^{n} c_k e_k, \tag{1}$$

where

$$c_k = (x, e_k).$$

The content of this section is, in a well known sense, a generalisation to the case of an infinite dimensional space of the decomposition (1). Let

$$\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots \tag{2}$$

be an orthonormal system, and $f \in L_2$.

We shall pose the following problem: for a given n select the coefficients α_k $(k = 1, 2, \cdots, n)$ in such a way that the distance, in the sense of the metric of the space L_2, between f and the sum

$$S_n = \sum_{k=1}^{n} \alpha_k \varphi_k \tag{3}$$

be as small as possible. Set $c_k = (f, \varphi_k)$. Since the system (2) is orthonormal,

$$\|f - S_n\|^2 = \left(f - \sum_{k=1}^{n} \alpha_k \varphi_k, f - \sum_{k=1}^{n} \alpha_k \varphi_k\right)$$

$$= (f, f) - 2\left(f, \sum_{k=1}^{n} \alpha_k \varphi_k\right) + \left(\sum_{k=1}^{n} \alpha_k \varphi_k, \sum_{j=1}^{n} \alpha_j \varphi_j\right)$$

$$= -\|f\|^2 - 2\sum_{k=1}^{n} \alpha_k c_k + \sum_{k=1}^{n} \alpha_k^2 = \|f\|^2 - \sum_{k=1}^{n} c_k^2 + \sum_{k=1}^{n} (\alpha_k - c_k)^2. \tag{4}$$

It is clear that this expression takes on its minimum when the last term is equal to zero, i.e., for

$$\alpha_k = c_k, \qquad k = 1, 2, \cdots, n. \tag{5}$$

In this case

$$\|f - S_n\|^2 = (f, f) - \sum_{k=1}^{n} c_k^2. \tag{6}$$

Definition. The numbers

$$c_k = (f, \varphi_k)$$

are called the *Fourier coefficients* of the function $f \in L_2$ on the orthogonal system (2) and the series

$$\sum_{k=1}^{\infty} c_k \varphi_k$$

(it may not converge) is called the *Fourier series* of the function on the system (2).

We have shown that among all the sums (3) for given n the one which differs least from f (in the sense of the L_2 metric) is the partial sum of the Fourier series of this function. Geometrically this result can be explained in the following way. The function

$$f - \sum_{k=1}^{n} \alpha_k \varphi_k$$

is orthogonal to all linear combinations of the form

$$\sum_{k=1}^{n} \beta_k \varphi_k,$$

i.e., it is orthogonal to the subspace generated by the functions $\varphi_1, \varphi_2, \cdots, \varphi_n$, in that and only in that case when condition (5) is satisfied. (Check this!) Thus, the result just obtained is a generalisation of a well known theorem of elementary geometry: the length of the perpendicular from a point to a line on a plane is smaller than the length of any of the inclined lines drawn from the same point.

Since always $\| f - S_n \|^2 \geq 0$, equation (4) implies

$$\sum_{k=1}^{n} c_k^2 \leq \| f \|^2.$$

Here n is arbitrary, and the right side does not depend on n; hence the series $\sum_{k=1}^{\infty} c_k^2$ converges and

$$\sum_{k=1}^{\infty} c_k^2 \leq \| f \|^2. \tag{7}$$

This inequality is called *Bessel's inequality*.

Let us introduce the following important concept.

Definition. An orthonormal system (2) is called *closed* if for every function $f \in L_2$ the equation

$$\sum_{k=1}^{\infty} c_k^2 = \|f^2\| \tag{8}$$

holds, which one calls *Parseval's equation.*

One can see from (6) that the fact that the system (2) is closed is equivalent to the fact that the partial sums of the Fourier series of each function $f \in L_2$ converge to f in the sense of the L_2 metric (i.e., in the mean).

The concept of closure of an orthonormal system is closely connected with the concept of completion of systems of functions introduced in §21.

Theorem 1. *In L_2 space every complete orthornomal system is closed, and conversely.*

Proof. Let $\{\varphi_n(x)\}$ be closed; then, whatever the function $f \in L_2$ may be, the sequence of partial sums of its Fourier series converges to it in the mean. This means that the linear combinations of elements of the system $\{\varphi_n(x)\}$ are everywhere complete in L_2, i.e., $\{\varphi_n\}$ is complete. Conversely, let $\{\varphi_n\}$ be complete, i.e., one can approximate any function $f \in L_2$ as closely as desired, in the sense of the L_2 metric, by a linear combination

$$\sum_{k=1}^{N} a_k\varphi_k$$

of elements $\{\varphi_k\}$; then the partial sum

$$\sum_{k=1}^{n} c_k\varphi_k$$

of the Fourier series for f gives, generally speaking, an even closer approximation of the function f and, therefore, the series

$$\sum_{k=1}^{\infty} c_k\varphi_k$$

converges to f in the mean, and Parseval's equality holds.

In §21 we have shown the existence of complete orthonormal systems in L_2. In as much as the concepts of closure and completion coincide for orthonormal systems of functions in L_2, the existence of closed orthogonal systems in L_2 does not require a separate proof, and the examples of complete orthonormal systems given in §21 are at the same time examples of closed systems.

From the Bessel inequality (7) it follows that in order that the numbers $c_1, c_2, \cdots$ be Fourier coefficients of some function $f \in L_2$ on some orthonormal system it is necessary that the series

$$\sum_{k=1}^{\infty} c_k^2$$

converge. It turns out that this condition is not only necessary but also sufficient. Indeed, the following theorem holds.

Theorem 2. (Riesz-Fischer). *Let $\{\varphi_n\}$ be an arbitrary orthonormal system in L_2 and let the numbers*

$$c_1, c_2, \cdots, c_n, \cdots$$

be such that the series

$$\sum_{k=1}^{\infty} c_k^2 \tag{9}$$

converges. Then there exists a function $f \in L_2$ such that

$$c_k = (f, \varphi_k),$$

and

$$\sum_{k=1}^{\infty} c_k^2 = (f, f).$$

Proof. Set

$$f_n = \sum_{k=1}^{n} c_k \varphi_k.$$

Then

$$\|f_{n+p} - f_n\|^2 = \|c_{n+1}\varphi_{n+1} + \cdots + c_{n+p}\varphi_{n+p}\|^2 = \sum_{k=n+1}^{n+p} c_k^2.$$

Since the series (9) converges, this implies, on the strength of the completeness of L_2, that the sequence $\{f_n\}$ converges in the mean to some function $f \in L_2$. Moreover,

$$(f, \varphi_i) = (f_n, \varphi_i) + (f - f_n, \varphi_i), \tag{10}$$

where the first term on the right is equal to c_i for $n \geq i$ and the second term tends to zero an $n \to \infty$ since

$$|(f - f_n, \varphi_i)| \leq \|f - f_n\| \cdot \|\varphi_i\|.$$

The left side of equation (10) does not depend on n; therefore, taking the limit as $n \to \infty$, we obtain

$$(f, \varphi_i) = c_i.$$

Since, by the definition of $f(x)$,

$$\|f - f_n\| \to 0 \quad \text{for} \quad n \to \infty,$$

the equation

$$\sum_{k=1}^{\infty} c_k^2 = (f, f)$$

is satisfied by f. The theorem is proved.

Let us in conclusion establish the following useful theorem.

Theorem 3. *In order that the orthonormal system of functions* (2) *be complete, it is necessary and sufficient that there does not exist a function in L_2 which is orthogonal to all functions of the system* (2) *and which is not equivalent to $\Psi(x) \equiv 0$.*

Proof. Let the system (2) be complete and therefore closed. If $f \in L_2$ is orthogonal to all functions of the system (2), than all its Fourier coefficients are equal to zero. Then, from Parseval's equation we obtain

$$(f, f) = \sum c_k^2 = 0,$$

i.e., $f(x)$ is equivalent to $\Psi(x) \equiv 0$.

Conversely, let $\{f_n\}$ be not complete, i.e., let there exist a function $g \in L_2$ such that

$$(g, g) > \sum_{k=1}^{\infty} c_k^2, \qquad \text{where} \qquad c_k = (g, \varphi_k);$$

then, by the Riesz-Fischer theorem, there exists a function $f \in L_2$, such that

$$(f, \varphi_k) = c_k, \qquad (f, f) = \sum_{k=1}^{\infty} c_k^2.$$

The function $f - g$ is orthogonal to all φ_i. By the inequality

$$(f, f) = \sum_{k=1}^{\infty} c_k^2 < (g, g),$$

it cannot be equivalent to $\Psi(x) \equiv 0$. The theorem is proved.

23. The Isomorphism of the Spaces L_2 and l_2

From the Riesz-Fischer theorem immediately follows the following important

Theorem. *The space L_2 is isomorphic* to the space l_2.*

Proof. Let us select in L_2 an arbitrary complete orthonormal system $\{\varphi_n\}$ and let us put into correspondence with every function $f \in L_2$ a sequence $c_1, c_2, \cdots, c_n, \cdots$ of its Fourier coefficients on this system. Since $\sum c_k^2 < \infty$, $(c_1, c_2, \cdots, c_n, \cdots)$ is therefore some element of l_2. Conversely, by virtue of the Riesz-Fischer theorem, there corresponds to every element $(c_1, c_2, \cdots, c_n, \cdots)$ of l_2 some function f of L_2, which has the numbers $c_1, c_2, \cdots, c_n, \cdots$ as its Fourier coefficients. The correspondence between the elements of l_2 and L_2 which we have established is one

* Two Euclidean spaces R and R' are said to be isomorphic if one can establish a one to one correspondence between their elements in such a way that if

$$x \leftrightarrow x', \qquad y \leftrightarrow y',$$

then

1) $$x + y \leftrightarrow x' + y',$$

2) $$\alpha x \leftrightarrow \alpha x',$$

3) $$(x, y) = (x', y').$$

It is clear that two isomorphic Euclidean spaces, considered just as metric spaces, are measurable.

to one. Moreover, if

$$f^{(1)} \leftrightarrow (c_1^{(1)}, c_2^{(1)}, \cdots, c_n^{(1)}, \cdots)$$

and

$$f^{(2)} \leftrightarrow (c_1^{(2)}, c_2^{(2)}, \cdots, c_n^{(2)}, \cdots),$$

then

$$f^{(1)} + f^{(2)} \leftrightarrow (c_1^{(1)} + c_1^{(2)}, c_2^{(1)} + c_2^{(2)}, \cdots, c_n^{(1)} + c_n^{(2)}, \cdots)$$

and

$$kf^{(1)} \leftrightarrow (kc_1^{(1)}, kc_2^{(1)}, \cdots, kc_n^{(1)}, \cdots),$$

i.e., the sum goes over into a sum, and the product by a number goes over into the product of a corresponding element by the same number. Finally, the Parseval equation yields that

$$(f^{(1)}, f^{(2)}) = \sum_{n=1}^{\infty} c_n^{(1)} c_n^{(2)}. \tag{1}$$

Indeed, from the fact that

$$(f^{(1)}, f^{(1)}) = \sum (c_n^{(1)})^2, \qquad (f^{(2)}, f^{(2)}) = \sum (c_n^{(2)})^2,$$

and

$$\begin{aligned}(f^{(1)} + f^{(2)}, f^{(1)} + f^{(2)}) &= (f^{(1)}, f^{(1)}) + 2(f^{(1)}, f^{(2)}) + (f^{(2)}, f^{(2)}) \\ &= \sum (c_n^{(1)} + c_n^{(2)})^2 + \sum (c_n^{(1)})^2 + 2 \sum c_n^{(1)} c_n^{(2)} + \sum (c_n^{(2)})^2\end{aligned}$$

(1) follows. Thus the correspondence which we have established between the elements of the space L_2 and l_2 is indeed an isomorphism. The theorem is proved.

This theorem means that one can consider l_2 to be a "coordinate notation" of the space L_2. It allows us to carry over facts we have established earlier for l_2 to L_2. For example, we have shown in Chapter III of Volume I that every linear functional in l_2 has the form

$$\varphi(x) = (x, y),$$

where y is some element of l_2, uniquely defined by the functional φ. From this, and from the isomorphism we have established between L_2 and l_2, it follows that every linear functional in l_2 has the form

$$\varphi(f) = (f, g) = \int f(x)g(x)\, d\mu,$$

where $g(x)$ is some fixed function from L_2. In §24, Volume I it was established that $\bar{l}_2 = l_2$. This, and the theorem just obtained, yield $\bar{L}_2 = L_2$.

The isomorphism established between L_2 and l_2 is closely related to certain problems in quantum mechanics. At first, quantum mechanics appeared in the form of two externally different theories: Heisenberg's "matrix mechanics" and Schrödinger's "wave mechanics". Later, Schrödinger established the equivalence of the two theories. From the purely mathematical point of view the difference between the two theories is reduced to the fact that in constructing the corresponding mathematical apparatus Heisenberg used the space l_2 and Schrödinger the space L_2.

CHAPTER V

THE ABSTRACT HILBERT SPACE. INTEGRAL EQUATIONS WITH A SYMMETRIC KERNEL.

In the preceding chapter we have established that the spaces L_2 (in the case of a separable measure) and l_2 are isomorphic, i.e., in essence represent two different realisations of one and the same space. This space, usually called Hilbert space, plays an important part in analysis and its applications. Often it is useful not to tie oneself down ahead of time by one or another realisation of this space, but to define it axiomatically, as is done for example in linear algebra with respect to the n-dimensional Euclidean space.

24. Abstract Hilbert Space

Definition 1. The set H of elements f, g, h, $\cdots$ of an arbitrary nature is called (abstract) Hilbert space, if the following conditions are satisfied:

I. H is a linear space.

II. A scalar product of elements is defined in H, i.e., to each pair of elements f and g there corresponds a number (f, g) such that,

1) $$(f, g) = (g, f),$$

2) $$(\alpha f, g) = \alpha(f, g),$$

3) $$(f_1 + f_2, g) = (f_1, g) + (f_2, g),$$

4) $$(f, f) > 0 \quad \text{if} \quad f \neq 0.$$

In other words, conditions I and II mean that H is a Euclidean space. The number $\|f\| = \sqrt{(f, f)}$ is called the norm of the element f.

III. The space H is complete in the sense of the $\rho(f, g) = \|f - g\|$ metric.

IV. The space H is infinite-dimensional, i.e., for any natural n, one can find in H, n linearly independent vectors.

V. The space H is separable*, i.e., there exists in it a countable everywhere dense set.

It is easy to give examples of spaces which satisfy all the enumerated axioms. Such is the space l_2 which we considered in Chapter II of Volume I. Indeed, it is a Euclidean space, which is infinitely dimensional since, for example, its elements

$$e_1 = (1, 0, 0, \cdots, 0, \cdots),$$

$$e_2 = (0, 1, 0, \cdots, 0, \cdots),$$

$$e_3 = (0, 0, 1, \cdots, 0, \cdots),$$

are linearly independent; the fact that it is complete and separable was proved in Chapter II, §9 and §13 of Volume I. The space L_2 of functions which are square integrable with respect to some separable measure which is equivalent to l_2 also satisfies the same axioms.

The following statement holds: *all Hilbert spaces are isomorphic.*

To prove this fact it obviously suffices to establish that every Hilbert space is isomorphic to the coordinate space l_2. This last assertion is proved in essence by the same considerations as the isomorphism of the spaces L_2 and l_2, namely:

1) One can carry over to Hilbert space without any changes those definitions of orthogonality, closure and completeness which were introduced for the elements of the space L_2 in §21.

2) Selecting in the Hilbert space H a countable everywhere dense set and applying to it the process of orthogonalisation described (for L_2) in §21, we construct in H a complete orthonormal system of elements, i.e., the system

$$h_1, h_2, \cdots, h_n, \cdots, \tag{1}$$

* Condition V is often omitted, i.e., one considers nonseparable Hilbert spaces.

which satisfies the conditions

a) $$(h_i, h_k) = \begin{cases} 0 & \text{for } i \neq k, \\ 1 & \text{for } i = k, \end{cases}$$

b) the linear combinations of elements of the system (1) are everywhere dense in H.

3) Let F be an arbitrary element in H. Set $c_k = (f, h_k)$. Then, the series $\sum c_k^2$ converges and $\sum c_k^2 = (f, f)$ for any complete orthonormal system $\{h_k\}$ and any element $f \in H$.

4) Let, again, $\{h_k\}$ be some complete orthonormal system of elements in H. Whatever the sequence of numbers

$$c_1, c_2, \cdots, c_n, \cdots$$

may be which satisfy the condition

$$\sum c_k^2 < \infty,$$

there exists in H an element $f \in H$ such that

$$c_k = (f, h_k),$$

and

$$\sum c_k^2 = (f, f).$$

5) From our statements we can see that one can realise an isomorphic mapping of H onto l_2, by setting

$$f \leftrightarrow (c_1, c_2, \cdots, c_n, \cdots),$$

where

$$c_k = (f, h_k),$$

and

$$h_1, h_2, \cdots, h_n, \cdots$$

is an arbitrary complete orthonormal system in H.

It is recommended that the details of the proof be completed by the reader himself using the methods of §§21–23.

25. Subspaces. Orthogonal Complements. Direct Sum.

In correspondence with the general definition of §21, Chapter III, of Volume I we shall call a set L of elements in H having the property: if $f, g \in L$, then $\alpha f + \beta g \in L$ for any numbers α and β, *a linear manifold.* A closed linear manifold is called a *subspace.*

Let us give a few examples of subspaces of Hilbert space.

1. Let h be an arbitrary element in H. The set of all elements $f \in H$ which are orthogonal to h forms a subspace in H.

2. Let H be realised as l_2, i.e., let its elements be sequences $(x_1, x_2, \cdots, x_n, \cdots)$ of numbers such that $\sum x_k^2 < \infty$. The elements which satisfy the condition $x_1 = x_2$ form a subspace.

3. Let H be realised as the space L_2 of all square integrable functions on some segment $[a, b]$ and let $a < c < b$. Let us denote the set of all functions from H which are identically equal to zero on the segment $[a, c]$ by H_c. H_c is a subspace of the space H. If $c_1 < c_2$, then $H_{c_1} \supset H_{c_2}$; moreover, $H_a = H$, $H_b = (0)$. Thus we obtain a continuum of subsets of H which are contained in each other. Each of these subspaces (of course with the exception of H_b) is infinite-dimensional and isomorphic to the whole space H.

It is recommended that the reader check that the sets of vectors given in examples 1–3 are indeed subspaces.

Every subspace of a Hilbert space is either a finite-dimensional Euclidean space or itself a Hilbert space. Indeed, it is obvious that axioms I–III are fulfilled for each such subspace, and the following lemma shows that axiom IV also holds.

Lemma. *From the existence of a countable everywhere dense set in a metric space R there follows the existence of a countable everywhere dense set in an arbitrary subspace R' of R.*

Proof. Let

$$\xi_1, \xi_2, \cdots, \xi_n, \cdots$$

be a countable everywhere dense set in R. Let, furthermore,

$$a_n = \inf_{\eta \in R'} \rho(\xi_n, \eta),$$

and $\eta_n \in R'$ be such that

$$\rho(\xi_n, \eta_n) < 2a_n.$$

For any $\eta \in R'$, one can find an n such that

$$\rho(\xi_n, \eta) < \frac{\varepsilon}{3},$$

and, therefore,

$$\rho(\xi_n, \eta_n) < \frac{2\varepsilon}{3};$$

then $\rho(\eta, \eta_n) < \varepsilon$ i.e., the countable set

$$\eta_1, \eta_2, \cdots, \eta_n, \cdots$$

is everywhere dense in R'.

The fact that a scalar product exists in Hilbert space and the concept of orthogonality makes it possible to supplement essentially the results given in the first part of the course* in connection with closed linear subspaces of arbitrary Banach spaces.

Applying the orthogonality process to an arbitrary countable everywhere dense sequence of elements of an arbitrary subspace of Hilbert space we obtain

Theorem 1. *In every subspace M of the space H there exists an orthogonal system $\{\varphi_n\}$ the linear closure of which coincides with M:*

$$M = \bar{M}(\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots).$$

Let M be a subspace of the Hilbert space H. Let us denote by

$$M' = H \ominus M$$

the set of elements $g \in H$ which are orthogonal to all $f \in M$, and let us show that M' is also a subspace of the space H.

Since $(g_1, f) = (g_2, f) = 0$ implies $(\alpha_1 g_1 + \alpha_2 g_2, f) = 0$, it is clear that M' is linear. To prove that it is closed assume that g_n be-

* Translater's Note: The reference is to *Functional Analysis*, Volume I, [*A*].

longs to M' and converges to g. Then, for any $f \in M$

$$(g, f) = \lim_{n\to\infty} (g_n, f) = 0,$$

and therefore g also belongs to M'.

M' is called the *orthogonal complement* of the subspace M.

Theorem 1 obviously implies

Theorem 2. *If M is a closed linear subspace of the space H, then any element $f \in H$ can be uniquely represented in the form $f = h + h'$, where $h \in M$, $h' \in M'$.*

Proof. Let us first show that such a decomposition exists. To do this let us find a complete orthonormal system $\{\varphi_n\}$ in M (so that $M = \bar{M}\ \{\varphi_n\}$) and set

$$h = \sum_{n=1}^{\infty} c_n\varphi_n, \qquad c_n = (f, \varphi_n).$$

Since (by Bessel's inequality), the series of c_n^2 converges, the elements h exist and belong to M. Set

$$h' = f - h.$$

Obviously, for any n

$$(h', \varphi_n) = 0,$$

and, since any element of M can be represented in the form

$$\zeta = \sum a_n\varphi_n,$$

we have for $\zeta \in M$

$$(h', \zeta) = \sum_{n=1}^{\infty} (h', \varphi_n) = 0.$$

Let us now assume that there exists, aside from the decomposition $f = h + h'$ constructed by us, another decomposition:

$$f = h_1 + h_1', \qquad h_1 \in M, \qquad h_1' \in M'.$$

Then, for all n,

$$(h_1, \varphi_n) = (f, \varphi_n) = c_n,$$

which implies that

$$h_1 = h, \qquad h_1' = h'.$$

Theorem 2 yields

Corollary 1. *The orthogonal complement to the orthogonal complement of a closed linear subspace M coincides with M itself.*

Thus, we can speak of mutually complementary subspaces of the space H. If M and M' are two such closed linear subspaces which complement one another, and $\{\varphi_n\}$, $\{\varphi_m'\}$ are complete orthogonal systems in M and M', respectively, then the union of the systems $\{\varphi_n\}$ and $\{\varphi_m'\}$ gives a complete orthogonal system in the whole space H. Therefore we have

Corollary 2. *Every orthonormal system $\{\varphi_n\}$ can be extended to a system which is complete in H.*

If the system $\{\varphi_n\}$ is finite, then the number of functions it contains is the dimension of the space M and at the same time the index of the space M'. Thus we obtain

Corollary 3. *The orthogonal complement to a space of finite dimension n has index n, and conversely.*

If every vector $f \in M$ is represented in the form $f = h + h'$, $h \in M$, $h' \in M'$ (M' is the orthogonal complement of M), then one says that H is the direct sum of mutually orthogonal subspaces M and M', and one writes

$$H = M \oplus M'.$$

It is clear that the concept of a direct sum can be immediately generalised to any finite, or even countable, number of subspaces: one says that H is the direct sum of its subspaces

$$H = M_1 \oplus M_2 \oplus \cdots \oplus M_n \oplus \cdots,$$

if:

1) the subspaces M_i are pairwise orthogonal i.e., any vector from M_i is orthogonal to any vector from M_k for $i \neq k$;

2) every element $f \in M$ can be represented in the form

$$f = h_1 + h_2 + \cdots + h_n + \cdots, \qquad h_n \in M_n. \tag{1}$$

Moreover, if the number of subspaces M_n is infinite, then $\sum \| h_n \|^2$ is a convergent series. One can easily see that there is only one possible such representation for any f and that

$$\|f\|^2 = \sum_n \|h_n\|^2.$$

Exactly as one speaks of a direct sum of subspaces so one can also speak of a direct sum of a finite or countable number of arbitrary Hilbert spaces. Namely, if H_1 and H_2 are two Hilbert spaces, their direct sum H is defined in the following way: the elements of the space H are all possible pairs (h_1, h_2), where $h_1 \in H_1$, $h_2 \in H_2$, and the scalar product of two such pairs equals

$$((h_1, h_2), (h_1', h_2')) = (h_1, h_1') + (h_2, h_2').$$

The space H obviously contains the mutually orthogonal subspaces which consist of pairs of the form $(h_1, 0)$ and $(0, h_2)$, respectively; the first of them can be identified, in a natural way, with the space H_1 and the second with the space H_2.

Analogously one defines the sum of an arbitrary finite number of spaces. The sum $H = \sum_n \oplus H_n$ of a countable number of spaces $H_1, H_2, \cdots, H_n, \cdots$ is defined in the following way: the elements of the space H are all possible sequences of the form

$$h = (h_1, h_2, \cdots, h_n, \cdots)$$

for which

$$\sum_n \|h_n\|^2 < \infty.$$

The scalar product (h, g) of the elements h and g of H equals

$$\sum_n (h_n, g_n).$$

26. Linear and Bilinear Functionals in Hilbert Space

The fact that every Hilbert space is isomorphic to the space l_2 allows us to carry over a series of facts, established in Chapter III of Volume I for l_2, to the abstract Hilbert space.

Since in l_2 every linear functional has the form

$$\varphi(x) = (x, a),$$

where a is an element from l_2,

any linear functional $F(h)$ in H can be represented in the form

$$F(h) = (h, g), \tag{1}$$

where g depends only on F.

This implies that the definition of weak convergence, introduced in Chapter III of Volume I for an arbitrary linear space, applied to the space H, can be formulated in the following way.

The sequence of elements h_n of the space H converges weakly to $h \in H$, if:

1) the norms $\| h_n \|$ are bounded*,

2) for every $g \in H$

$$(h_n, g) \to (h, g).$$

Any orthonormal sequence

$$\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots$$

in H converges weakly to zero, since for any $h \in M$

$$c_n = (h_n, \varphi_n) \to 0 \quad \text{for} \quad n \to \infty,$$

because

$$\sum c_n^2 \leq (h, h) < \infty.$$

Moreover, such a sequence does not, of course, converge in norm to any limit.

In particular, applying this remark to the case when H is the space of square integrable functions on the segment $[a, b]$ of the real axis with the usual Lebesgue measure, we obtain the following interesting fact.

---

* See the note on page 77 of Volume 1.

Let

$$\varphi_1(t),\ \varphi_2(t),\ \cdots,\ \varphi_n(t),\ \cdots$$

be some orthonormal system of functions in this space, and let

$$f(t) = \begin{cases} 1 & \text{on the segment } [t, t_2] \subset [a, b], \\ 0 & \text{outside } [t_1, t_2]. \end{cases}$$

Then,

$$(f, \varphi_n) = \int_{t_1}^{t_2} \varphi_n(t)\, dt.$$

Thus, for any orthonormal system of functions $\varphi_n(t)$ and any t_1 and t_2 from $[a, b]$,

$$\int_{t_1}^{t_2} \varphi_n(t)\, dt \to 0.$$

If the $\varphi_n(t)$ are bounded in the set, we see from this fact, taking into account the condition

$$\int_a^b \varphi_n^2(t)\, dt = 1,$$

that $\varphi_n(t)$ with a large index inevitably changes its sign many times (which can be observed, for example, in the case of a trigonometric system).

In Chapter III of Volume I we introduced, together with weak convergence of elements of a linear normalised space, also the concept of weak convergence of a sequence of functionals. In as much as the Hilbert space coincides with its conjugate, the two concepts are identical for it; therefore, Theorem 1 of §28 of Volume I gives the following result for Hilbert space:

The unit sphere in H is weakly compact, i.e., one can select from each sequence of elements $\varphi_n \in H$ for which $\|\varphi_n\| = 1$ a weakly converging subsequence.

In what follows we shall also need the following

Theorem 1. *If ξ_n converges weakly to ξ in H, then*

$$\|\xi\| < \sup \|\xi_n\|.$$

Proof. Whatever the complete orthonormal system $\{\varphi_k\}$ in H may be, we have

$$c_k = (\xi, \varphi_k) = \lim_{n\to\infty} (\xi_n, \varphi_k) = \lim_{n\to\infty} c_{nk},$$

and

$$\sum_{m=1}^{k} c_m^2 = \lim_{n\to\infty} \sum_{m=1}^{k} c_{nm}^2 \leq \sup \sum_{m=1}^{\infty} c_{nm}^2;$$

therefore

$$\sum_{m=1}^{\infty} c_m^2 \leq \sup_n \sum_{m=1}^{\infty} c_{nm}^2,$$

which proves the theorem.

Let $B(f, g)$ be a real function of a pair of vectors from H which satisfies the following condition: for fixed g, $B(f, g)$ is a linear functional in f, and for fixed f it is a linear functional in g. $B(f, g)$ is called the *bilinear functional*. The bilinear functional $B(f, g)$ is called *symmetric*, if

$$B(f, g) = B(g, f) \quad \text{for any } f, g \in H.$$

From the theorem concerning the general form of a linear functional in H it follows that every bilinear functional in H can be written in the form

$$B(f, g) = (\zeta, g),$$

where ζ depends on f. It is easy to see that the correspondence

$$f \to \zeta$$

is a continuous linear operator in H; let us denote it by A. Thus,

$$B(f, g) = (Af, g). \tag{2}$$

Analogously, we can obtain another expression:

$$B(f, g) = (f, A^*g),$$

where A^* is another linear operator, which one calls adjoint* to A. If the functional $B(f, g)$ is symmetric, then

$$(Af, g) = B(f, g) = B(g, f) = (Ag, f) = (f, Ag),$$

i.e.,

$$A = A^*. \tag{3}$$

A linear operator satisfying condition (3) is called *self-adjoint.*

Formula (2) establishes a one to one correspondence between the bilinear functionals and continuous linear operators in H. Moreover, the symmetric bilinear functionals correspond to the self-adjoint operators and vice versa.

Setting $f = g$ in a symmetric bilinear functional, we obtain the so called quadratic functional

$$Q(f) = B(f, f);$$

by (2)

$$Q(f) = (Af, f),$$

where A is a self-adjoint linear operator.

Since the correspondence between the symmetric bilinear functionals and the quadratic functionals is one to one,† the correspondence between quadratic functionals and self-adjoint linear operators is also one to one.

27. Completely Continuous Self-Adjoint Operators in H

In Chapter IV, Volume I, we introduced the concept of a completely continuous linear operator which acts in some Banach space E. In this section we shall supplement the facts which were

* In Chapter III of Volume I, considering linear operators in an arbitrary Banach Space E, we defined the operator A^*, adjoint to some operator A, with the help of equation

$$(Ax, \varphi) = (x, A^*\varphi)$$

for all $x \in E$, $\varphi \in \overline{E}$. If E is the Hilbert space, $\overline{E} = E$ and the definition of the operator A^*, introduced in Chapter III of Volume I, goes over into the definition formulated here.

† $Q(f) = B(f, f)$, and, conversely:

$$B(f, g) = \frac{1}{4}[Q(f + g) - Q(f - g)].$$

established for arbitrary completely continuous operators, limiting ourselves only to self-adjoint completely continuous operators, acting in a Hilbert space.

Let us remember that we called an operator A completely continuous if it transformed every bounded set into a compact set. In as much as $H = \overline{H}$, i.e., H is conjugate to a separable space, all the bounded sets (and only those) are weakly compact in it; therefore, in the case of Hilbert space, the definition of complete continuity can be formulated in the following way:

The operator A acting in the Hilbert space H, is called completely continuous if it transforms every weakly compact set into a compact (in norm) set.

In Hilbert space this is equivalent to the fact that the operator A transforms every weakly converging sequence into a strongly converging one.

In this section we shall establish the following basic theorem which is a generalisation, to self-adjoint completely continuous operators, of a theorem stating how matrices of a self-adjoint linear transformation in n-dimensional space may be brought to diagonal form.

Theorem 1. *For any completely continuous self-adjoint linear operator A in Hilbert space H there exists an orthonormal system $\{\varphi_n\}$ of eigenvectors corresponding to the eigenvalues $\{\lambda_n\}$ such that every element $\xi \in H$ may be written in a unique way as $\xi \equiv \sum_k c_k\varphi_k + \xi'$, where ξ' satisfies the condition $A\xi' = 0$; moreover,*

$$A\xi = \sum_k \lambda_k c_k \varphi_k,$$

and $\lim_{n=\infty} \lambda_n = 0$.

For the proof of this basic theorem we shall need the following auxiliary assertion:

Lemma 1. *If $\{\xi_n\}$ converges weakly to ξ and the self-adjoint linear operator A is completely continuous, then*

$$Q(\xi_n) = (A\xi_n, \xi_n) \to (A\xi, \xi) = Q(\xi).$$

Proof.

$|(A\xi_n, \xi_n) - (A\xi, \xi)| \leq |(A\xi_n, \xi_n) - (A\xi_n, \xi)| + |(A\xi_n, \xi) - (A\xi, \xi)|.$

But

$$|(A\xi_n, \xi_n) - (A\xi_n, \xi)| = |(\xi_n, A(\xi_n - \xi))| \leq \|\xi_n\| \cdot \|A(\xi_n - \xi)\|,$$

and

$$|(A\xi_n, \xi) - (A\xi, \xi)| = |(\xi, A(\xi_n - \xi))| \leq \|\xi\| \cdot \|A(\xi_n - \xi)\|,$$

and since the quantities $\|\xi_n\|$ are bounded and $\|A(\xi_n - \xi)\| \to 0$,

$$|(A\xi_n, \xi_n) - (A\xi, \xi)| \to 0,$$

which was to be proved.

Lemma 2. *If the functional*

$$|Q(\xi)| = |(A\xi, \xi)|,$$

where A is a bounded self-adjoint linear operator, attains its maximum on the unit sphere at the point ξ_0, then

$$(\xi_0, \eta) = 0$$

implies that

$$(A\xi_0, \eta) = (\xi_0, A\eta) = 0.$$

Proof. Obviously $\|\xi_0\| = 1$. Let us set

$$\xi = \frac{\xi_0 + a\eta}{\sqrt{1 + a^2\|\eta\|^2}},$$

where a is an arbitrary number. From $\|\xi_0\| = 1$ it follows that

$$\|\xi\| \leq 1.$$

Since

$$Q(\xi) = \frac{1}{1 + a^2\|\eta\|^2}[Q(\xi_0) + 2a(A\xi_0, \eta) + a^2Q(\eta)],$$

we have for small a

$$Q(\xi) = Q(\xi_0) + 2a(A\xi_0, \eta) + O(a^2).$$

The last equation makes it clear that if $(A\xi_0, \eta) \neq 0$, then a can be selected in such a way that $|Q(\xi)| > |Q(\xi_0)|$, which contradicts the condition of the theorem.

Lemma 2 immediately implies that if $|Q(\xi)|$ attains its maximum for $\xi = \xi_0$, then ξ_0 is an eigenvector of the operator A.

Proof of the Theorem. We shall construct the elements φ_k by induction, in order of decreasing absolute values of the corresponding eigenvalues:

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| \geq \cdots.$$

To construct the element φ_1 let us consider the expression $Q(\xi) = |(A\xi, \xi)|$ and let us prove that it attains a maximum on the unit sphere. Let

$$S = \sup_{\|\xi\| \leq 1} |(A\xi, \xi)|$$

and $\xi_1, \xi_2, \cdots$ be a sequence for which $\|\xi_n\| \leq 1$ and

$$|(A\xi_n, \xi_n)| \to S \quad \text{for} \quad n \to \infty.$$

Since the unit sphere is weakly compact in H, one can select from $\{\xi_n\}$ a subsequence which converges weakly to some element η. Here moreover, by Theorem 1 of §26, $\|\eta\| \leq 1$ and, by Lemma 1,

$$|(A\eta, \eta)| = S.$$

Let us take the element η for φ_1. It is clear that $\|\eta\|$ is exactly equal to unity. Hence

$$A\varphi_1 = \lambda_1\varphi_1,$$

which yields

$$|\lambda_1| = \frac{|(A\varphi_1, \varphi_1)|}{(\varphi_1, \varphi_1)} = |(A\varphi_1, \varphi_1)| = S.$$

Assume now that the eigenvectors

$$\varphi_1, \varphi_2, \cdots, \varphi_n,$$

corresponding to the eigenvalues

$$\lambda_1, \lambda_2, \cdots, \lambda_n,$$

are already constructed. Let us consider the functional

$$|(A\xi, \xi)|$$

on the set of elements which belong to

$$M_n' = H \ominus M(\varphi_1, \varphi_2, \cdots, \varphi_n)$$

(i.e., orthogonal to $\varphi_1, \varphi_2, \cdots, \varphi_n$) and which satisfy the condition $\|\xi\| < 1$. M_n' is a subspace which is invariant with respect to A (since $M(\varphi_1, \varphi_2, \cdots, \varphi_n)$ is invariant and A is self-adjoint). Applying the above considerations we see that one can find a vector in M_n' which is an eigenvector of the operator A; let us call it φ_{n+1}.

Two cases are possible: 1) After a finite number of steps we obtain a subspace M_{n_0}' in which $(A\xi, \xi) \equiv 0$; 2) $(A\xi, \xi) \not\equiv 0$ on M_n' for all n.

In the first case, we see from Lemma 2 that the operator A transforms M_{n_0}' into zero, i.e., M_{n_0}' consists wholly of eigenvectors which correspond to $\lambda = 0$. The system of vectors $\{\varphi_n\}$ which we have constructed consists of a finite number of elements.

In the second case we obtain the sequence $\{\varphi_n\}$ of eigenvectors for each of which $\lambda_n \neq 0$. We shall show that $\lambda_n \to 0$. The sequence $\{\varphi_n\}$ (as every orthonormal sequence) converges weakly to zero, therefore the elements $A\varphi_n = \lambda_n\varphi_n$ must converge to zero with respect to the norm; hence $\lambda_n = \|A\varphi_n\| \to 0$.
Let

$$M' = H \ominus M\{\varphi_n\} = \bigcap_n M_n' \neq 0.$$

If $\xi \in M'$ and $\xi \neq 0$, then

$$(A\xi, \xi) \leq \lambda_n\|\xi\|^2 \quad \text{for all } n,$$

i.e.,

$$(A\xi, \xi) = 0,$$

from which, by Lemma 2 (for $\max_{\|\xi\| \leq 1} |(A\xi, \xi)| = 0$) applied to M', we see that $A\xi = 0$, i.e., the operator A transforms the subspace M' into zero.

From the construction of the system $\{\varphi_n\}$ it is clear that every vector can be represented in the form

$$\xi = \sum_k c_k\varphi_k + \xi', \quad \text{where} \quad A\xi' = 0,$$

which implies that

$$A\xi = \sum_k \lambda_k c_k \varphi_k.$$

28. Linear Equations with Completely Continuous Operators

Let us consider the equation

$$\xi = cA\xi + \eta, \tag{1}$$

where A is a completely continuous self-adjoint operator, the element $\eta \in H$ is given, and $\xi \in H$ is sought for.

Let

$$\varphi_1, \varphi_2, \cdots, \varphi_n, \cdots$$

be eigenvectors of the operator A corresponding to the eigenvalues which are different from zero. Then η can be written in the form

$$\eta = \sum_n a_n\varphi_n + \eta', \tag{2}$$

where $A\eta' = 0$. Let us look for the solution of equation (1) in the form

$$\xi = \sum_n x_n\varphi_n + \xi', \tag{3}$$

where $A\xi' = 0$. Inserting (2) and (3) into (1), we obtain

$$\sum_n x_n(1 - \lambda_n c)\varphi_n + \xi' = \sum_n a_n\varphi_n + \eta'.$$

This equation is satisfied if and only if

$$\xi' = \eta',$$

$$x_n(1 - \lambda_n c) = a_n,$$

i.e., if

$$\xi' = \eta',$$

$$x_n = \frac{a_n}{1 - \lambda_n c} \quad \text{for} \quad \lambda_n \neq \frac{1}{c}, \tag{4}$$

$$a_n = 0 \qquad \text{for} \quad \lambda_n = \frac{1}{c}.$$

The last equality gives the necessary and sufficient condition for equation (1) to be solvable and (4) determines the solution. The values x_n, corresponding to the n for which $\lambda_n = 1/c$, remain arbitrary.

29. Integral Equations with a Symmetric Kernel

The results described in the preceding section can be applied to integral equations with symmetric kernels, i.e., to equations of the form

$$f(t) = \varphi(t) + \int_a^b K(t, s) f(s)\, ds, \tag{1}$$

where $K(t, s)$ satisfies the conditions

1) $$K(t, s) = K(s, t),$$

2) $$\int_a^b \int_a^b K^2(t, s)\, dt\, ds < \infty.$$

The application of the results of §28 to equations of the form (1) is based on the following theorem.

Theorem. *Let R be some space with a measure μ prescribed on it. If the function $K(t, s)$ defined on $R^2 = R \times R$ satisfies the conditions*

$$K(t, s) = K(s, t) \tag{2}$$

and

$$\int_{R^2} K^2(t, s)\, d\mu^2 < \infty, \qquad \mu^2 = \mu \times \mu, \tag{3}$$

then the operator

$$g = Af,$$

given in the space $L_2(R, \mu)$ by the formula

$$g(t) = \int_R K(t, s) f(s)\, ds,$$

is completely continuous and self-adjoint.

Proof. Let us denote the space $L_2(R, \mu)$ by L_2. Let $\{\psi_n(t)\}$ be a complete orthonormal system in L_2. The set of all possible products $\psi_n(t)\ \psi_m(s)$ is a complete system of functions in R^2, and

$$K(t, s) = \sum_m \sum_n a_{mn} \psi_n(t) \psi_m(s),$$

where

$$a_{mn} = a_{nm}$$

(by virtue of (2)), and

$$\sum_m \sum_n a_{mn}^2 = \int_{R^2} K^2(t, s)\, d\mu^2 < \infty.$$

Set

$$f(s) = \sum_n b_n \psi_n(s),$$

then

$$g(x) = Af = \sum_m \Big(\sum_n a_{mn} b_n\Big) \psi_m(x) = \sum_m c_m \psi_m(x).$$

Moreover,

$$c_m^2 = \left(\sum_{n=1}^{\infty} a_{mn} b_n\right)^2 \leq \sum_{n=1}^{\infty} a_{mn}^2 \cdot \sum_{n=1}^{\infty} b_n^2 = \|f\|^2 \cdot a_m^2,$$

where

$$a_m^2 = \sum_n a_{mn}^2.$$

Since the series

$$\sum_{m=1}^{\infty} a_m^2 = \sum_m \sum_n a_{mn}^2$$

converges, one can find, for any $\varepsilon > 0$, an m_0 such that

$$\sum_{m=m_0+1}^{\infty} a_m^2 < \varepsilon,$$

$$\left\| g(x) - \sum_{m=1}^{m_0} c_m \psi_m(x) \right\|^2 = \sum_{m=m_0+1}^{\infty} c_m^2 < \varepsilon \, \|f\|^2.$$

Now let $\{f^{(k)}\}$ converge weakly to f. Then the set $c_m^{(k)}$ converges to c_m for every m, hence the sum

$$\sum_{m=1}^{m_0} c_m \, \psi_m(x)$$

converges in the mean for any fixed m_0 to the sum

$$\sum_{m=1}^{m_0} c_m \, \psi_m(x).$$

By inequality (4) and the boundedness of the norm $\|f^{(k)}\|$ this implies that $\{g^{(k)}(x)\}$ (where $g^{(k)} = Af^{(k)}$) converges in the mean to $g(x)$, which proves the complete continuity of the operator A. Furthermore, from condition (1) and Fubini's theorem it follows that

$$(Af, g) = \int_R \left(\int_R K(s, t) f(t) \, d\mu_t \cdot g(s) \right) d\mu_s$$

$$= \int_R f(t) \left(\int_R K(s, t) g(s) \, d\mu_s \right) d\mu_t = (f, Ag),$$

i.e., the operator A is self-adjoint. The theorem is proved. In this way the problem of solving an integral equation with a kernel satisfying conditions (2) and (3) is reduced to finding eigenfunctions and eigenvalues of the corresponding integral operator. The practical solution of this last problem usually requires the use of some limiting methods, the description of which is beyond the scope of this book.

ADDITIONS AND CORRECTIONS TO VOLUME I*

After our first book appeared, a certain number of misprints, some of which were fundamental, and various inadequacies of presentation were discovered. Below we list corrections of the errors found.

(1) Page 28, line 23 from above. Instead of G_α read $G_\alpha(x)$.

(2) Page 46, lines 23–24 from above. Instead of "the method of successive approximations is not applicable" it should read "the method of successive approximations is, generally speaking, not applicable".

(3) Page 49, after line 22 insert, "where for $f_0(x)$ one can take any continuous function".

(4) Page 51, in line 2 from the top for M write M^2, in line 3 from the top in place of M write M^n (twice), and in line 5 from the top for λ^n write $\lambda^n M^n$.

(5) Page 55, line 15 from above. For $\varepsilon/5$ write "less than $\varepsilon/5$.

(6) Page 56, line 13 from above. Instead of "of the region G" it should read "of the bounded region G".

(7) Page 57, line 28 from above. For $\varphi(x)$ write $\varphi^{(k)}(x)$.

(8) Page 61, before Theorem 7, one should insert: "We shall call the mapping $y = f(x)$ uniformly continuous if, for any $\varepsilon > 0$, one can find a $\delta > 0$ such that $\rho(f(x_1), f(x_2)) < \varepsilon$ for all x_1, x_2 for which $\rho(x_1, x_2) < \delta$. The following assertion holds: *Every continuous mapping of a compactum into a compactum is uniformly convergent.* It is proved by the same method as the uniform continuity of a function which is continuous on a segment."

---

* The additions and corrections listed here are in terms of the page numbering of the English translation of Volume I, published by Graylock, Press.

(9) Page 61. After line 17 from above, instead of the word "Proof" it should read:

"Let us first prove the necessity. If D is compact, then there exists in D a finite $\varepsilon/3$ net $f_1, f_2, \cdots, f_N$. Since each of the mappings f_i is continuous, it is uniformly continuous, therefore one can find $\delta > 0$ such that

$$\rho(f_i(x_1), f_i(x_2)) < \frac{\varepsilon}{3}, \qquad i = 1, 2, \ldots, n,$$

only if

$$\rho(x_1, x_2) < \delta.$$

For every mapping $f \in D$ one can find an f_i for which

$$\rho(f, f_i) < \frac{\varepsilon}{3}.$$

Then,

$$\rho(f(x_1), f(x_1)) \leq \rho(f(x_1), f_i(x_1)) + \rho(f_i(x_1), f_i(x_2)) + \rho(f_i(x_2), f(x_2)) < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

only if $\rho(x_1, x_2) < \delta_1$, and this means that the mappings $f \in D$ are uniformly continuous. Let us now prove the sufficiency."

(10) Page 72, line 3 from the bottom. Instead of "$\max\limits_{1\leq n\leq\infty}$" read "$\sup\limits_{1\leq n\leq\infty}$".

(11) Page 77, line 9 from above. Instead of "continuous" it should read "continuous at the point x_0". In line 11 from above instead of "$|f(x_1) - f(x_2)|$" read "$|f(x) - f(x_0)|$" and in line 13 instead of "$\| x_1 - x_2 \|$" read "$\| x - x_0 \|$."

(12) Page 80, line 4 of Theorem 3. Instead of "$x_0 \notin L_f$" read, "x_0 is a fixed element, not belonging to L_f".

(13) Page 84, line 12 from above. Instead of "$\sup_n x_n$" read "$\sup_n |x_n|$".

(14) Page 92, lines 26–27 from above. The statement that "these functionals are everywhere dense in $\bar{C}_{[a,b]}$" is not true. Instead of "satisfies the conditions of Theorem 1, i.e., linear combinations of these functionals are everywhere dense in $\bar{C}_{[a,b]}$" read: "has the property that if the sequence $\{x_n(t)\}$ is bounded and $\varphi(x_n) \to \varphi(x)$ for all $\varphi \in \Delta$, then $\{x_n(t)\}$ converges weakly to $x(t)$".

(15) Page 94. The metric introduced here leads to weak convergence of functionals in every bounded subset of the space $\bar{R}$ (but not in all of $\bar{R}$). In lines 32–33, after the words "so that", insert, "in every bounded subset of the space $\bar{R}$".

(16) The proof of Theorem 5 given here contains a mistake. It should be replaced by the following:

"Proof. 1) Let us note first of all that every eigenvalue of a continuous operator which is different from zero has finite multiplicity. Indeed, the set E_λ of all eigenvectors which correspond to a given eigenvalue λ is a linear space, the dimension of which is equal to the multiplicity of this eigenvalue. If, for some $\lambda \neq 0$, this subspace were infinite, the operator A would not be completely continuous in E_λ, and, therefore, also in the whole space.

2) Now to complete the proof of the theorem it remains to show that whatever the sequence $\{\lambda_n\}$ of pairwise different eigenvalues of a completely continuous operator A, $\lambda_n \to 0$ for $n \to \infty$. Let x_n be an eigenvector of the operator A, corresponding to the eigenvalue λ_n. The vectors x_n are linearly independent. Let $E_n (n = 1, 2, \cdots)$ be a subspace consisting of all elements of the form $y = \sum_{i=1}^{n} \alpha_i x_i$. For every $y \in E_n$ we have

$$y - \frac{1}{\lambda_n} Ay = \sum_{i=1}^{n} \alpha_i x_i - \sum_{i=1}^{n} \frac{\alpha_i \lambda_i}{\lambda_n} x_i = \sum_{i=1}^{n-1} \left(1 - \frac{\lambda_i}{\lambda_n}\right) \alpha_i x_i,$$

which shows that $y - Ay/\lambda_n \in E_{n-1}$.

Let us select a sequence of elements $\{y_n\}$ in such a way that

$$y_n \in E_n, \qquad \|y_n\| = 1, \qquad \rho(y_n, E_{n-1}) > \frac{1}{2}.$$

(The possibility of selecting such a sequence is shown on page 118).

Let us now assume that the sequence $1/\lambda_n$ is bounded. Then the set $\{A(y_n/\lambda_n)\}$ must be compact, but this is impossible since, for $p > q$,

$$\left\|A\left(\frac{y_p}{\lambda_p}\right) - A\left(\frac{y_q}{\lambda_q}\right)\right\| = \left\|y_p - \left[y_p - \frac{1}{\lambda_p} Ay_p + A\left(\frac{y_q}{\lambda_q}\right)\right]\right\| > \frac{1}{2},$$

because

$$y_p - \frac{1}{\lambda_p} Ay_p + A\left(\frac{y_q}{\lambda_q}\right) \in E_{p-1}.$$

The contradiction we have obtained proves our assertion."

(17) Page 119, line 12 from above. The statement that G_0 is a subspace is correct but is not evident. Therefore one should change the sentence: "Let G_0 be the subspace consisting of all the elements of the form $x - Ax$." into: "Let G_0 be a linear manifold consisting of all elements of the form $x - Ax$. Let us show that G_0 is closed. Let $\tilde{T}$ be a one to one mapping of the factor space E/N (where N is the subspace of elements satisfying the condition $x - Ax = 0$) on G_0. We must show that the inverse mapping $\tilde{T}^{-1}$ is continuous. It suffices to show that it is continuous at the point $y = 0$. Assume that this is not the case; then there exists a sequence $y_n \to 0$ such that, for $\xi_n = \tilde{T}^{-1}y_n$, the inequality $\| \xi_n \| \geq \rho > 0$ is satisfied. Setting $\eta_n = \xi_n/\| \xi_n \|$, $z_n = y_n/\| \xi_n \|$, we obtain the sequence $\{\eta_n\}$ which satisfies the conditions $\| \eta_n \| = 1$, $\tilde{T}\eta_n = z_n \to 0$. Selecting in each class η_n a representative x_n, in such a way that $\| x_n \| \leq 2$, we obtain a bounded sequence which satisfies the condition: $z_n = Tx_n = x_n - Ax_n \to 0$. But, since the operator

A is completely continuous, $\{Ax_n\}$ contains a continuous fundamental subsequence $\{Ax_{n_k}\}$; the sequence $x_p = z_p + Ax_p$ is also fundamental and, therefore, it converges to some element x_0. Hence $z_p = Tx_p \to Tx_0$, which implies $Tx_0 = 0$. i.e., $x_0 \in N$. But then, $\| \eta_p \| \leq \| x_p - x_0 \| \to 0$, which contradicts the condition $\| \eta_p \| = 1$. The contradiction obtained shows that the mapping $\tilde{T}^{-1}$ is continuous, and, therefore also that G_0 is closed. Hence G_0 is a subspace."

SUGGESTED REFERENCES

[A] Kolmogorov, A. N., and Fomin, S. V. *Elements of the Theory of Functions and Functional Analysis*, Volume I. New York: Graylock, 1957.

1. Banach, S. *Théorie des Operations Linéaires.* (Monografje Matematyczne, Tom I). Warszawa, 1932.
2. Dieudonné, J. *Foundations of Modern Analysis.* New York: Academic Press, 1960.
3. Goffman, C. *Real Functions.* New York: Rinehart, 1953.
4. Hall, D. W., and Spencer, G. L. *Elementary Topology.* New York: Wiley, 1955.
5. Hahn, H., and Rosenthal, A. *Set Functions.* Albuquerque: The University of New Mexico Press, 1948.
6. Halmos, P. R. *Introduction to Hilbert Space.* New York: Chelsea, 1951.
7. Halmos, P. R. *Measure Theory.* New York: Van Nostrand, 1950.
8. Hille, E. *Functional Analysis and Semi-Groups.* New York: American Mathematical Society, 1948.
9. Kestelman, H. *Modern Theories of Integration.* Oxford: Oxford University Press, 1941.
10. Munroe, M. E. *Introduction to Measure and Integration.* New York: Addison-Wesley, 1953.
11. Nathanson, I. P. *Theory of Functions of a Real Variable.* New York: Ungar, 1955.
12. Petrovskii, I. G. *Lectures on the Theory of Integral Equations.* New York: Graylock, 1957.
13. Riesz, F., and Sz-Nagy, B. *Functional Analysis.* New York: Ungar, 1955.
14. Saks, S. *Theory of the Integral.* (Monografje Matematyczne, Tom VII) Warszaw, Lwow, 1937.
15. Stone, M. H. *Linear Transformations in Hilbert Space.* New York: American Mathematical Society, 1932.
16. Taylor, A. E. *Introduction to Functional Analysis.* New York: Wiley, 1958.
17. Zaanen, A. C. *Linear Analysis.* New York: Interscience, 1953.

INDEX