UNIVERSITY MATHEMATICAL TEXTS

GENERAL EDITORS

ALEXANDER C. AITKEN, D.Sc., F.R.S.
DANIEL E. RUTHERFORD, D.Sc., Dr. Math.

Determinants and Matrices	Prof. A. C. Aitken, D.Sc., F.R.S.
Statistical Mathematics	Prof. A. C. Aitken, D.Sc., F.R.S.
Waves	Prof. C. A. Coulson, D.Sc., F.R.S.E.
Electricity	Prof. C. A. Coulson, D.Sc., F.R.S.E.
Projective Geometry	T. E. Faulkner, Ph.D.
Integration	R. P. Gillespie, Ph.D.
Partial Differentiation	R. P. Gillespie, Ph.D.
Infinite Series	Prof. J. M. Hyslop, D.Sc.
Integration of Ordinary Differential Equations	E. L. Ince, D.Sc.
Introduction to the Theory of Finite Groups	W. Ledermann, Ph.D., D.Sc.
Analytical Geometry of Three Dimensions	Prof. W. H. M'Crea, Ph.D., F.R.S.
Functions of a Complex Variable	E. G. Phillips, M.A., M.Sc.
Classical Mechanics	D. E. Rutherford, D.Sc., Dr. Math.
Vector Methods	D. E. Rutherford, D.Sc., Dr. Math.
Volume and Integral	Prof. W. W. Rogosinski, Dr. Phil.
Theory of Equations	Prof. H. W. Turnbull, F.R.S.

Other volumes in preparation

Tensor Calculus	Barry Spain, B.A., M.Sc., Ph.D.

VOLUME AND INTEGRAL

BY

WERNER W. ROGOSINSKI
DR. PHIL. (Göttingen)

PROFESSOR OF PURE MATHEMATICS
IN THE UNIVERSITY OF DURHAM
KING'S COLLEGE, NEWCASTLE UPON TYNE

With 11 Figures

OLIVER AND BOYD
EDINBURGH AND LONDON
NEW YORK: INTERSCIENCE PUBLISHERS, INC.
1962

FIRST PUBLISHED 1952

PRINTED IN GREAT BRITAIN
BY R. AND R. CLARK, LTD., EDINBURGH

PREFACE

It is now half a century since *H. Lebesgue* created his theory of the integral which has widely superseded in modern analysis the classical conception due to *B. Riemann.* It is, I think, regrettable that knowledge of the Lebesgue integral seems to be still largely confined to the research worker. There is nothing unduly abstract or unnatural in this theory, nor anything in the proofs which would be too difficult for a good honours student to grasp. If the aim of university education be the teaching of general ideas rather than that of technicalities, then the modern notion of the integral should not be omitted from the mathematical honours syllabus.

The main object of this book is to provide an introduction to the theory of the so-called *absolute* integral. It is not an introduction to the "calculus": it is assumed that the reader is familiar with this. But the book should give the student a deeper understanding of the ideas underlying the calculus. It is also hoped that he will appreciate the aesthetic side of a purely mathematical theory, quite apart from its practical implications. Such an appreciation is quite as essential as technical skill.

As the title indicates, I have tried to bring out consistently the *geometrical* aspect of integration: the integral of a (positive) function is the volume of the ordinate set of the function. This seems to me, both historically and intrinsically, the natural approach and that which is likely to suit the student best.

The first part of the book deals with the problem of volume in a space of n dimensions. First the older definition of *content* (*Peano, Jordan*) is discussed. It is followed by the theory of the modern and more satisfactory definition of *measure* (*Lebesgue*).

The second part begins with the theory of *Riemann*'s integral of a function of n variables. This is defined geometrically, using content as the underlying notion of volume. *Lebesgue*'s integral is then obtained in a similar way on replacing content by measure. The relation between the two definitions, and the striking advantages of the new integral, are thus clearly set out. The book ends with the theory of the *indefinite* integral of a function of one variable: the discussion of the familiar feature of the calculus, that differentiation and integration are inverse operations.

It has been necessary, for reasons of space economy, to restrict this account to the essentials of the theory: the properties of the spaces L^p and such important applications as length of arc and surface area had to be omitted. Nor are the notions of the *Stieltjes* integral and of a non-absolute integral (*Denjoy-Perron*) included. A list of books, suitable for comparison or further study, is given at the end. Of these, the book by *H. Kestelman* and the recent Cambridge Tract by *J. C. Burkill* proceed on lines similar to ours. In particular, it is hoped, that our geometrical account of the absolute integral may serve as a stimulating introduction to the standard work on integration, the book by *St. Saks*, which, in the two different editions, presents the modern more abstract approach to the subject.

It is somewhat difficult to provide exercises in a subject which is essentially theoretical. I have given a few: the solutions are at the end of each chapter.

Many friends have helped in preparing this text by suggestions, criticism, and proof reading, and all deserve my thanks: my Newcastle-Durham colleagues *F. F. Bonsall*, *Professor A. C. Offord* (now in London), and *Dr. J. V. Whitworth* must be specially mentioned. My main thanks, however, are due to *Dr. D. E. Rutherford* who as editor suggested the book and helped it along in many ways. Finally, I wish to express my gratitude to the Publishers and Printers for their patient and fine work under somewhat difficult circumstances.

W. W. ROGOSINSKI

DURHAM UNIVERSITY, KING'S COLLEGE
NEWCASTLE UPON TYNE
January 1952

CONTENTS

CHAPTER III

MEASURE

PART II—THE INTEGRAL

CHAPTER IV

RIEMANN'S INTEGRAL

CHAPTER V

LEBESGUE'S INTEGRAL

CHAPTER VI

INTEGRATION AND DIFFERENTIATION

PART I

VOLUME

CHAPTER I

SETS OF POINTS

1.1. The Euclidean space. The aim of the theory of measure is to give a precise meaning, in as general form as possible, to the intuitive but vague geometrical concepts of length, area, and volume. The Euclidean space of three dimensions, an ideal image of the "intuitive" space of primitive sense experience, is a logical system of abstract entities called points, lines, and planes which are inter-related to form a "geometrical" pattern according to certain rules called axioms

It is shown in co-ordinate geometry how to establish an arithmetical model of this space. On introducing a system of Cartesian co-ordinates a one-one correspondence between all points P of the space and all ordered triplets (x, y, z) of real numbers is obtained. To the planes correspond linear equations between the co-ordinates, and to the lines pairs of simultaneous linear equations. The Euclidean axioms are the equivalent of the ordinary arithmetical axioms of this co-ordinate algebra.

In a similar way the system of all real numbers x can be interpreted as an arithmetical equivalent of the line (a space of one dimension), and the system of all ordered pairs (x, y) of real numbers as an arithmetical equivalent of the plane (a space of two dimensions). All this is familiar.

1.2. The space of n dimensions. More generally, we consider complexes of n real numbers,

$$P = (x_1, x_2, \ldots x_n) = (x_i), \quad 1 \leq i \leq n, \quad . \quad (1.2.1)$$

where n is a fixed positive integer. These complexes are ordered: that is, $P=(x_i)$ and $Q=(y_i)$ are the same complex if, and only if, $x_i=y_i$ for all i. Thus, when $n=2$, the two complexes (1, 2) and (2, 1) are different.

We use geometrical language and call each complex P a *point in the space of* n *dimensions*; the numbers x_i are the *co-ordinates* of P. It should be noted that these are mere names so far, at least when $n>3$.† We shall have to attach some "geometrical" significance to them: the analogy with the true geometrical cases $n\leq 3$ will serve as a guide.

Thus the obvious definition of the *distance* between two points $P=(x_i)$ and $Q=(y_i)$ will be

$$PQ=QP=[(x_1-y_1)^2+(x_2-y_2)^2+\ldots(x_n-y_n)^2]^{\frac{1}{2}} \quad . \quad (1.2.2)$$

We shall then have, for any three points P, Q, R, the triangle relation

$$PQ\leq PR+RQ; \quad . \quad . \quad . \quad (1.2.3)$$

that is, *one side of a "triangle" is at most equal to the sum of the two other sides.*

The proof is elementary. We use the inequality

$$(a-b)^2\leq a^2+b^2+2|ab|$$

and Cauchy's inequality

$$\left(\sum_1^n a_ib_i\right)^2\leq\sum_1^n a_i^2\,.\,\sum_1^n b_i^2 \quad . \quad . \quad . \quad (1.2.4)$$

First, if R is the origin $O=(0, 0, \ldots 0)$, then

$$\begin{aligned}PQ^2=\Sigma(x_i-y_i)^2 &\leq \Sigma x_i^2+\Sigma y_i^2+2\Sigma|x_iy_i|\\ &\leq \Sigma x_i^2+\Sigma y_i^2+2(\Sigma x_i^2\,.\,\Sigma y_i^2)^{\frac{1}{2}}\\ &=[(\Sigma x_i^2)^{\frac{1}{2}}+(\Sigma y_i^2)^{\frac{1}{2}}]^2=[PO+OQ]^2.\end{aligned}$$

In the general case, when $R=(z_i)$, consider the points $P'=(x_i-z_i)$ and $Q'=(y_i-z_i)$. Clearly, $PQ=P'Q'$, $PR=P'O$, and $RQ=OQ'$, so that the general case is reduced to the previous one.

1.3. Sets. Any assemblage of points P in a given space of n dimensions is called a *set* (of points) in this space: any prescription will define a set. We denote a set, usually,

† When $n=4$, a complex (x, y, z, t) can be interpreted as an "event", that is as a point (x, y, z) at the time t.

by E †; and we shall often write $P \varepsilon E$ for "P belongs to E".

The most comprehensive set is the given space itself. We denote it by E, or by E_n, if we wish to place in evidence the number of dimensions. Thus E_1 is the set of all real numbers x (the line), E_2 is the set of all ordered pairs (x, y) (the plane), and E_3 is the set of all ordered triplets (x, y, z) of real numbers (the "ordinary" space).

The following simple sets will frequently occur.

If C is a fixed point and ρ is a given positive number, then the set of all points P in E_n for which $PC < \rho$ is called an *open sphere*, of *centre* C and *radius* ρ. It is denoted by $K_\rho(C)$, or simply by K.‡ Thus, if $n=1$, a linear "sphere" is an open interval of centre C and length 2ρ; if $n=2$, a "sphere" is a circle. In these cases we shall, of course, retain the usual words. The set of points for which $PC \leq \rho$ is called a *closed sphere*.

A *closed interval* $I = \langle a_i, b_i \rangle$ is defined as the set of all points $P = (x_i)$ for which

$$a_i \leq x_i \leq b_i, \quad 1 \leq i \leq n. \qquad (1.3.1)$$

Thus, if $n=2$, a closed interval is a rectangle; if $n=3$, it is a cuboid. Note that the sides, or edges, are, by definition, parallel to the co-ordinate axes. In these cases we shall retain the usual words.

The *edges* of an interval I (more precisely, the lengths of the edges) are the numbers $b_i - a_i$; and the *volume* of I is defined as

$$|I| = \prod_1^n (b_i - a_i). \qquad (1.3.2)$$

If all edges are equal we speak of a *cube* (or a square, if $n=2$). We allow some, or all, edges to be zero: in the extreme case, an interval may reduce to a point. In this chapter, however, the edges will usually be positive.

An *open interval*, denoted by $(I) = (a_i, b_i)$, is the set of

† E indicates the French word *ensemble*; we reserve S for another use.

‡ K indicates the German word *Kugel*.

all points for which $a_i < x_i < b_i$. Its edges and volume are defined as above. The point $(\frac{1}{2}(a_i + b_i))$ is called the *centre* of the interval I, or (I).

A set E is said to be *finite* if it contains only a finite number of points. It will be convenient to admit as finite also a "set" which contains *no* point. This set is called the *null set* (or *empty set*) and is denoted by O. A non-finite set is said to be *infinite*.

1.4. Subsets. We consider, throughout this book, a given space $\mathsf{E} = \mathsf{E}_n$.

A set E_1 is said to be a *subset* of the set E_2, if every point of E_1 also belongs to E_2: $P \varepsilon E_1$ implies $P \varepsilon E_2$. We then say that E_1 is contained in E_2, or that E_2 contains E_1, and write this as

$$E_1 \subset E_2, \quad \text{or} \quad E_2 \supset E_1 \qquad . \quad . \quad (1.4.1)$$

Clearly, this relation is transitive: if $E_1 \subset E_2$ and $E_2 \subset E_3$, then $E_1 \subset E_3$. Every set is a subset of itself: $E \subset E$.

If $P \varepsilon E$, then $\{P\} \subset E$ where $\{P\}$ is the set consisting of the point P only.

Clearly, $\mathsf{E} \supset E$ whatever E may be. At the other extreme we regard the null set as a subset of every set: $O \subset E$.

If $E_1 \subset E_2$ but $E_1 \neq E_2$, then E_1 is called a *proper subset* of E_2. Thus O is a proper subset of any non-empty set. An open sphere is a proper subset of the corresponding closed sphere; and similarly for intervals.

The set of all points (x, y, c), where c is fixed, is a two-dimensional subset (a plane) of E_3; the set of all points (a, y, c), where a and c are fixed, represents a line in E_3. Similarly, fixing k $(<n)$ of the co-ordinates, we obtain a $(n-k)$-dimensional *subspace* of E_n. More generally, we can define such a subspace by k independent linear equations between the co-ordinates.

A set E is called *bounded* if there is a closed interval I such that $E \subset I$. Any finite set, interval, or sphere is bounded. Neither the space itself nor any of its subspaces is bounded.

1.5. Complements. If $E_1 \subset E_2$, then the set of all points belonging to E_2 but not to E_1 is called the *complement of E_1 with respect to E_2*, and is denoted by $E_2 - E_1$.

If K is the open sphere $PC < \rho$, and $\bar{K}$ is the closed sphere $PC \leq \rho$, then $\bar{K} - K$ is the *circumference* $PC = \rho$ of these spheres ; and similarly for intervals. Also $E - E = \mathsf{O}$ and $E - \mathsf{O} = E$.

It should be noted that the *difference* $E_2 - E_1$ is not defined unless $E_1 \subset E_2$.

The difference $\mathsf{E} - E$ can always be formed : it consists of all points not belonging to E. In this case we speak, simply, of the *complement of E* and denote it by ${}_cE$. Clearly, ${}_c({}_cE) = E$. Also ${}_c\mathsf{E} = \mathsf{O}$ and ${}_c\mathsf{O} = \mathsf{E}$.

1.6. Products. We denote by $E_1 \,.\, E_2$, or $E_2 \,.\, E_1$, the set of all points which belong both to E_1 and E_2. The sets E_1 and E_2 are the *factors* of this *product*.

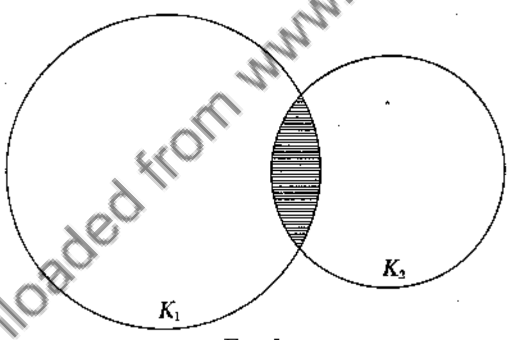

FIG. 1

Clearly, $E_1 \,.\, E_2 \subset E_1$ and $E_1 \,.\, E_2 \subset E_2$. Also $E_1 \,.\, E_2 = E_1$ if $E_1 \subset E_2$. In particular, $E \,.\, E = E \,.\, \mathsf{E} = E$, while $\mathsf{O} \,.\, E = \mathsf{O}$. Further

$$E_2 - E_1 = E_2 \,.\, {}_cE_1 \qquad . \quad . \quad . \quad (1.6.1)$$

Some authors take this as definition of the difference : it has the advantage that E_1 need not be a subset of E_2.

B

However, in the theory of volume our restriction is desirable.

If $E_1 . E_2$ is not empty, then the two sets E_1 and E_2 are said to *overlap*. In Fig. 1 the shaded part shows the product of two overlapping closed circles K_1 and K_2 in the plane.

If $E_1 . E_2 = O$ then E_1 and E_2 are *exclusive sets*. Thus an open sphere and its circumference are *exclusive*; so are E and ${}_cE$. More generally, the product

$$E_1 . E_2 \dots E_m = \prod_1^m E_k$$

is the set of all points which are common to all the factor sets E_k. If $\Pi_k = E_1 . E_2 \dots E_k$ then $\Pi_1 \supset \Pi_2 \supset \dots \supset \Pi_m$.

1.7. Sums. We denote by $E_1 + E_2$, or $E_2 + E_1$, the set of all points which belong either to E_1 or to E_2. This *sum*, or *union*, contains both its *terms* as subsets. In Fig. 1, $K_1 + K_2$ is the set of all points covered by the two circles.

Clearly, $E + {}_cE = \mathsf{E}$. Also $E_1 + E_2 = E_2$ if $E_1 \subset E_2$. In particular, $E + E = E + O = E$, and $E + \mathsf{E} = \mathsf{E}$. Also, if $E_1 \subset E_2$,

$$E_2 = E_1 + (E_2 - E_1), \quad . \quad . \quad . \quad (1.7.1)$$

the two terms being exclusive.

More generally, the sum

$$E_1 + E_2 + \dots + E_m = \sum_1^m E_k$$

is the set of all points which belong to at least one of the terms E_k.

If $\Sigma_k = E_1 + E_2 + \dots + E_k$, then $\Sigma_1 \subset \Sigma_2 \subset \dots \subset \Sigma_m$.

Next, we have the following important pair of formulae

$$E - \Sigma E_k = \Pi (E - E_k) ; \quad E - \Pi E_k = \Sigma (E - E_k) ; \quad (1.7.2)$$

that is: *the complement (with respect to E) of a sum is the product of the complements of its terms; the complement of a product is the sum of the complements of its factors.*

To prove the first formula, we observe that the set on the left-hand side contains all the points of E which do not belong to any of the E_k. These are exactly the points which belong to all the complements $E - E_k$, that is, to the product on the right-hand side. The proof for the second formula is similar.

In particular,

$$_c\Sigma E_k = \Pi_c E_k \, ; \quad _c\Pi E_k = \Sigma_c E_k \quad . \quad (1.7.3)$$

In Fig. 1, the complement of $K_1 . K_2$ is the non-shaded part of the plane, and this can be obtained by adding the " exteriors " of the two circles.

1.8. Sequences of sets. We next consider a sequence

$$E_1, E_2, \ldots E_k, \ldots$$

of sets in E, in short, a sequence (E_k). It is clear what we understand by $\sum_1^\infty E_k$ and by $\prod_1^\infty E_k$; these definitions involve no problem of convergence. Clearly, $\sum_1^\infty E_k$ is the " smallest " set containing all the E_k, and $\prod_1^\infty E_k$ is the "largest" set contained in all the E_k.

Let K_ρ be the open sphere with fixed centre C and radius ρ. Then $\sum_1^\infty K_k = \mathsf{E}$ and $\prod_1^\infty K_k = K_1$. Also $\sum_1^\infty K_{1/k} = K_1$ while $\prod_1^\infty K_{1/k}$ consists of the single point C.

The formulae (1.7.2) and (1.7.3) extend at once, with the same proofs, to infinite sums and products.

A sequence (E_k) is said to be *ascending* if $E_k \subset E_{k+1}$, and then $\prod_1^\infty E_k = E_1$. It is *descending* if $E_k \supset E_{k+1}$, and then $\sum_1^\infty E_k = E_1$. If (E_k) ascends then $(_cE_k)$ descends, and vice versa.

Next we define the two *limiting sets* of a sequence (E_k). The $\underline{\lim}\, E_k$ (*limit inferior* of the E_k) is the set of all points which belong *eventually*, that is, from a certain suffix on, to all the E_k. The $\overline{\lim}\, E_k$ (*limit superior* of the E_k) is the set of all points which belong to an infinity of the E_k. Clearly, $\underline{\lim}\, E_k \subset \overline{\lim}\, E_k$.

Exercise 1. Let C_k be the open circle with centre $((-1)^k/k, 0)$ and radius 1, in the (x, y)-plane. Find $\underline{\lim}\, C_k$ and $\overline{\lim}\, C_k$.

Exercise 2. Let (a_k) be a sequence of positive numbers, and let I_k be the closed linear interval $\langle 0, a_k \rangle$. If $l = \underline{\lim}\, a_k$ and $L = \overline{\lim}\, a_k$, find $\underline{\lim}\, I_k$ and $\overline{\lim}\, I_k$.

Consider the sets

$$\Sigma_k = \sum_k^{\infty} E_j, \quad \Pi_k = \prod_k^{\infty} E_j \quad . \quad . \quad (1.8.1)$$

Clearly, $\Pi_k \subset \Sigma_k$. The sequence (Σ_k) descends and the sequence (Π_k) ascends as k increases. We conclude that

$$\underline{\lim}\, E_k = \sum_1^{\infty} \Pi_k, \quad \overline{\lim}\, E_k = \prod_1^{\infty} \Sigma_k \quad . \quad . \quad (1.8.2)$$

For, a point of $\Sigma \Pi_k$ belongs to at least one of the Π_k, to Π_{k_0}, say. Hence it belongs to all E_k with $k \geq k_0$, that is, to $\underline{\lim}\, E_k$. Conversely, a point of $\underline{\lim}\, E_k$ belongs eventually to all E_j and hence to some Π_{k_0}; and so it belongs to $\Sigma \Pi_k$. This proves the first formula.

Similarly, a point of $\Pi \Sigma_k$ belongs to all Σ_k. Hence it belongs to an infinity of the E_k, that is, to $\overline{\lim}\, E_k$. Conversely, a point of $\overline{\lim}\, E_k$ belongs to an infinity of the E_k and hence to all Σ_k; and so it belongs to $\Pi \Sigma_k$.

Similar formulae hold for the $\underline{\lim}\, a_k$ and $\overline{\lim}\, a_k$ of a sequence of real numbers a_k. Replace the E_k by the a_k, and the symbol $\subset$ by $\leq$. Then $b = \inf a_k$,† the greatest number less than or equal to all the a_k, corresponds to ΠE_k, the largest set contained

† Inf a_k is the greatest lower bound of the a_k; sup a_k is the least upper bound. b may be $-\infty$, and B may be $+\infty$.

in all E_k. Similarly, $B = \sup a_k$, the smallest number greater than or equal to all a_k, corresponds to ΣE_k. We then have

$$\underline{\lim}\ a_k = \sup b_k, \quad \overline{\lim}\ a_k = \inf B_k, \quad . \quad . \quad (1.8.3)$$

where $b_k = \inf\limits_{i \geq k} a_i$ and $B_k = \sup\limits_{i \geq k} a_i$.

Exercise 3. Use (1.7.3) to prove that

$$_c\underline{\lim}\ E_k = \overline{\lim}\ _cE_k \quad . \quad . \quad . \quad (1.8.4)$$

We say that the sequence (E_k) *converges* to the *limit* E, and write $E_k \to E$, or $\lim E_k = E$, if

$$\underline{\lim}\ E_k = \overline{\lim}\ E_k\ (= E). \quad . \quad . \quad (1.8.5)$$

Any ascending sequence converges. In fact,

$$E_k \uparrow \sum_1^\infty E_j \quad \text{if} \quad E_k \subset E_{k+1}, \quad . \quad . \quad (1.8.6)$$

the symbol $\uparrow$ indicating ascending convergence. For, in this case, $\Pi_k = E_k$ and every $\Sigma_k = \sum\limits_1^\infty E_j = \Sigma_1$, so that

$$\underline{\lim}\ E_k = \sum_1^\infty \Pi_k = \sum_1^\infty E_k = \prod_1^\infty \Sigma_k = \overline{\lim}\ E_k.$$

Similarly, *any descending sequence converges:*

$$E_k \downarrow \prod_1^\infty E_j \quad \text{if} \quad E_k \supset E_{k+1} \quad . \quad . \quad (1.8.7)$$

Note that, for *any* sequence,

$$\Pi_k \uparrow \underline{\lim}\ E_k, \quad \Sigma_k \downarrow \overline{\lim}\ E_k. \quad . \quad . \quad (1.8.8)$$

The following result is important enough to be stated as a theorem.

THEOREM 1. *The product of a descending sequence of closed intervals is not empty.*

PROOF. Let $I_k = \langle a_i^{(k)},\ b_i^{(k)} \rangle$ and $I_k \supset I_{k+1}$. Then, for fixed i, $a_i^{(k)} \uparrow a_i$ and $b_i^{(k)} \downarrow b_i$, say, where $a_i \leq b_i$. Hence ΠI_k

is the interval $\langle a_i, b_i \rangle$ which is not empty, though it may consist of one point only.

Given a sequence of points P_k, we shall say that *the points P_k tend to the point P*, and we shall write

$$P_k \to P \quad . \quad . \quad . \quad . \quad (1.8.9)$$

if $P_kP \to 0$.

This definition does not imply that $\{P_k\} \to \{P\}$, where $\{P_k\}$ denotes the set consisting of the one point P_k. For instance, when all the P_k are different, then both the $\underline{\lim}$ and the $\overline{\lim}$ of the sets $\{P_k\}$ are empty, so that $\{P_k\} \to O$.

1.9. Enumerable sets. A set E is said to be *enumerable* if the totality of its points can be arranged as a finite or infinite sequence

$$P_1, \quad P_2, \quad \ldots \quad P_k, \quad \ldots$$

It is convenient to include the null set as an enumerable set.

Any finite set is enumerable. The set of all integers p, that is, the set of all points (p) in E_1, is enumerable. They may be arranged as

$$0, 1, -1, 2, -2, \ldots .$$

Before giving further examples we collect the following properties of enumerable sets.

(i) *Any subset of an enumerable set is enumerable.*

This is obvious. Note that the null set is enumerable, by definition. In particular :

(ii) *The difference of two enumerable sets, and the product of an enumerable set by any set, is enumerable.*

(iii) *The sum of two enumerable sets is enumerable.*

For, if $P_1, P_2, \ldots P_k, \ldots$ are the points of E_1, and $Q_1, Q_2, \ldots Q_k, \ldots$ are the points of E_2, then the points of $E_1 + E_2$ are obtained from the sequence $P_1, Q_1, P_2, Q_2, \ldots$ by omitting any point which has occurred before in that order.

Clearly, this result extends to a finite sum of enumerable sets. More generally, the following theorem holds.

THEOREM 2. *The sum of a sequence of enumerable sets is enumerable.*

PROOF. Let E_k, for $k \geq 1$, consist of the sequence of points $P_1^{(k)}, P_2^{(k)}, \ldots P_m^{(k)}, \ldots$. Then the points of ΣE_k can be arranged as a sequence in a manner indicated by the arrows in the following scheme:

$$\begin{array}{lccccccc}
E_1: & P_1^{(1)} & \longrightarrow & P_2^{(1)} & & P_3^{(1)} & \longrightarrow & P_4^{(1)} \quad \ldots \\
 & & \swarrow & & \nearrow & & \swarrow & \\
E_2: & P_1^{(2)} & & P_2^{(2)} & & P_3^{(2)} & & \ldots \\
 & \downarrow & \nearrow & & \swarrow & & & \\
E_3: & P_1^{(3)} & & P_2^{(3)} & & P_3^{(3)} & & \ldots \\
 & & \swarrow & & & & & \\
\ldots & & & & & & &
\end{array}$$

Points which have occurred before in this order are to be omitted.

1.10. The rational set. An important example of an enumerable set is the *rational set* $R = R_n$ of all points $(r_1, r_2, \ldots r_n)$ in E_n with rational co-ordinates r_i.

THEOREM 3. *The rational set R is enumerable.*

PROOF. (i) Let $n=1$. The set $R(q)$ of all rational numbers $\frac{p}{q}$, where $q(\geq 1)$ is fixed and p runs through all integers, is, clearly, enumerable. Hence $R_1 = \sum_{q=1}^{\infty} R(q)$ is enumerable, by Theorem 2.

(ii) We complete the proof by induction. Suppose that R_n is enumerable. The points of R_{n+1} are of the form $P = (r_1, r_2, \ldots r_n, r_{n+1})$, where the r_i are rationals. Let r_{n+1} be fixed, $r_{n+1} = s$, say. Then the points $P^{(s)} = (r_1, r_2, \ldots r_n, s)$ form a subset $R_{n+1}^{(s)}$ of R_{n+1}; and there is a one-one correspondence between the $P^{(s)}$ and the points $(r_1, r_2, \ldots r_n)$ of R_n. Hence $R_{n+1}^{(s)}$ is enumerable for every fixed rational s. By (i), the s can be arranged as a sequence (s_i). Also $R_{n+1} = \sum_i R_{n+1}^{(s_i)}$. Hence R_{n+1} is enumerable, by Theorem 2.

COROLLARY. *The set of all "lattice points" $(l_1, l_2, \ldots l_n)$ in E_n, where the l_i are integers, is enumerable.*

For, it is a subset of R_n.

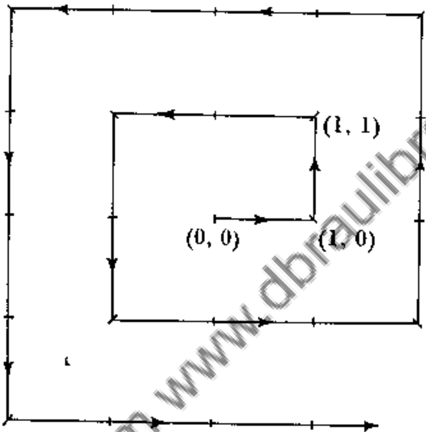

FIG. 2

Fig. 2 indicates an enumeration of the lattice points in the plane.

Exercise 4. A number α is called *algebraic* if it is a root of an equation

$$a_0x^n + a_1x^{n-1} + \ldots + a_n = 0$$

with integral coefficients a_k.

Show that *the set of all algebraic numbers is enumerable.*

1.11. Non-enumerable sets. Not every set is enumerable; and if a set is non-enumerable, then no set containing it can be enumerable.

THEOREM 4. *An interval (with at least one positive edge) is non-enumerable.*

Proof. It will, clearly, be enough to prove that the interval $\langle 0, 1\rangle$ in E_1 is not enumerable. Consider the subset of all real numbers ξ with a decimal expansion

$$\xi = .\, \epsilon_1\, \epsilon_2 \ldots \epsilon_k \ldots, \qquad (1.11.1)$$

where the ϵ_k are either 0 or 1. It is sufficient to show that they cannot be enumerated. Let

$$\begin{aligned} \xi_1 &= .\, \epsilon_1^{(1)}\, \epsilon_2^{(1)}\, \epsilon_3^{(1)} \ldots \\ \xi_2 &= .\, \epsilon_1^{(2)}\, \epsilon_2^{(2)}\, \epsilon_3^{(2)} \ldots \\ \xi_3 &= .\, \epsilon_1^{(3)}\, \epsilon_2^{(3)}\, \epsilon_3^{(3)} \ldots \\ &\ldots \end{aligned}$$

be any sequence of such numbers ξ. We have to show that this sequence cannot contain all numbers ξ.

In fact, let, for all $k \geq 1$, $\eta_k = 1$ when $\epsilon_k^{(k)} = 0$, and $\eta_k = 0$ when $\epsilon_k^{(k)} = 1$. Then the number

$$\xi = .\, \eta_1\, \eta_2 \ldots \eta_k \ldots$$

is of the considered type but is, plainly, different from every ξ_k of the above sequence.

Corollary. *The space E_n is not enumerable.*

1.12. Interior, exterior, and frontier of a set.

It will be convenient to call any *open* sphere K with centre P a *neighbourhood* of P.

A point P is called an *interior point* of a set E, if there exists a neighbourhood K of P such that $K \subset E$. The point P, of course, must belong to E. The set of all interior points of E is called *the interior* E_i of E. Clearly, $E_i \subset E$.

A point P is called an *exterior point* of E if it is an interior point of the complement ${}_cE$. The set of all exterior points is *the exterior* E_e of E. No exterior point of E belongs to E, and $E_e \subset {}_cE$. In fact, $E_e = ({}_cE)_i$ and $E_i = ({}_cE)_e$.

Any point of the space E which is neither an interior nor an exterior point of E is called a *frontier point* of E.

The frontier points form *the frontier* E_f of E. Both E and ${}_cE$ have the same frontier.

Any neighbourhood K of a frontier point P must contain points both of E and of ${}_cE$. The frontier point P itself may belong either to E or to ${}_cE$.

If $P \varepsilon E$, and if there is a neighbourhood K of P which, except for P, belongs to ${}_cE$, then P is called an *isolated point* of E. An isolated point is always a frontier point.

A given set E divides the space into three mutually exclusive parts

$$\mathsf{E} = E_i + E_f + E_e. \qquad (1.12.1)$$

An open interval, or an open sphere, is identical with its interior. The interior of a closed interval, or sphere, is the corresponding open interval, or sphere. For a finite set the interior is empty: all points are isolated, $E = E_f$, and $E_e = {}_cE$. The same holds for the set of all lattice points.

For the rational set R we have

$$R_i = O, \quad R_f = \mathsf{E}, \quad R_e = O. \qquad (1.12.2)$$

For, any neighbourhood of any point in E contains points both of R and of ${}_cR$.

Also

$$\mathsf{E}_i = \mathsf{E}, \quad \mathsf{E}_f = O, \quad \mathsf{E}_e = O, \qquad (1.12.3)$$

while

$$O_i = O, \quad O_f = O, \quad O_e = \mathsf{E}. \qquad (1.12.4)$$

1.13. Open and closed sets. A set E is said to be *open* if $E = E_i$. An open set will usually be denoted by O.

Both open spheres and open intervals are open (hence the notation), and so are the space E and the null set O.

The interior and the exterior of any set are open. Clearly, the interior is the largest open set contained in E.

The characteristic feature of an open set O *is that any point* P *of* O *has a neighbourhood* K *contained in* O.

THEOREM 5. (i) *A finite product of open sets is open.*
(ii) *A finite or infinite sum of open sets is open.*

PROOF. (i) Any point P of $\prod_1^m O_k$ has a neighbourhood $K_k \subset O_k$, for each of the k. The smallest of these neighbourhoods is contained in the product.

(ii) Any point P of ΣO_k belongs to at least one term O_{k_0}, say. Hence it has a neighbourhood K such that $K \subset O_{k_0} \subset \Sigma O_k$.

It should be noted that *an infinite product of open sets need not be open.* Thus the product of the linear open intervals $(-1/k, 1/k)$ is the set $\{0\}$ consisting of the single point 0 : a non-open set.

A set E is said to be *closed* if its complement ${}_cE$ is open; that is, if ${}_cE = E_e$ or $E = E_i + E_f$. We usually denote closed sets by F.†

Both closed spheres and closed intervals are closed. Any finite set is closed, and so is the set of the lattice points.

Both the space E and the null set O are closed; for, their respective complements O and E are open. They are the only sets which are both open and closed. The rational set R is neither open nor closed.

The frontier E_f of any set is closed. For, its complement $E_i + E_e$ is, as sum of two open sets, open.

Similarly, the set

$$\bar{E} = E_i + E_f. \qquad . \quad . \quad . \quad (1.13.1)$$

is closed. It is called the *closure* of E.

The closure of E is the "smallest" closed set containing E. For, its complement is E_e, the largest open set contained in ${}_cE$.

A set E *is closed if, and only if,* $E = \bar{E}$*:* an obvious consequence of the above.

The closure $\bar{R}$ of the rational set R is E.

† F indicates the French word *fermé*.

THEOREM 6. (i) *A finite sum of closed sets is closed.*
(ii) *A finite or infinite product of closed sets is closed.*

PROOF. By (1.7.3), we have ${}_c\Pi F_k = \Sigma {}_c F_k$, which is open, by Theorem 5, since ${}_c F_k$ is open. This proves (ii). The proof of (i) is similar.

An infinite sum of closed sets need not be closed. Thus the sum of the linear closed intervals $\langle 1/k,\ 1-1/k\rangle$, where $k \geq 2$, is the open interval $(0, 1)$.

THEOREM 7. *The complement of an open set with respect to a closed set is closed. The complement of a closed set with respect to an open set is open.*

PROOF. The difference $F - O$ is, by (1.6.1), the product of the two closed sets F and ${}_c O$; and $O - F$ is the product of the two open sets O and ${}_c F$.

1.14. Limiting points. A point P is called a *limiting point* of a set E, if *every* neighbourhood K of P contains at least one point Q of E different from P. It should be noted that P itself is not necessarily a point of E.

If $K = K_\rho$ is given, and if Q $(\neq P)$ is a point of E in K_ρ, then, considering a smaller neighbourhood K_{ρ_1} with $\rho_1 < \frac{1}{2}PQ$, we can find another point Q_1 $(\neq P)$ of E in K_{ρ_1}. Continuing in this way we see that every neighbourhood K of a limiting point P of E contains, in fact, an infinity of points of E, and that we can find a sequence of points Q_k $(\neq P)$ of E such that $Q_k \to P$. Conversely, it is clear that, if such a sequence (Q_k) exists, then P is a limiting point of E.

The set of the limiting points of E is called the *derivative* E' of E. The derivative E' consists of the interior E_i and of the non-isolated frontier points of E. Hence $E' \subset E_i + E_f = \bar{E}$.

The derivative of a finite set is empty. The derivative R' of R is E.

THEOREM 8. *The derivative* E′ *of any set* E *is closed.*

PROOF. The complement ${}_cE'$ consists of E_e and of the isolated points of E. Hence, whenever $P \,\varepsilon\, {}_cE'$, there exists a neighbourhood K of P which contains no point of E except perhaps (the isolated) P. It follows that $K \subset {}_cE'$ so that ${}_cE'$ is open.

THEOREM 9. *A set* E *is closed if, and only if,* E′ ⊂ E; *that is, if* E *contains all its limiting points.*

PROOF. If E is closed then $E = \bar{E}$. Hence $E' \subset E$.

Conversely, if $E' \subset E$, then E contains all non-isolated frontier points. Since, in any case, it contains E_i and all isolated (frontier) points, it contains $\bar{E}$. On the other hand, $E \subset \bar{E}$, so that $E = \bar{E}$: E is closed.

THEOREM 10. *Suppose that* E *is infinite and bounded. Then its derivative* E′ *is not empty.*

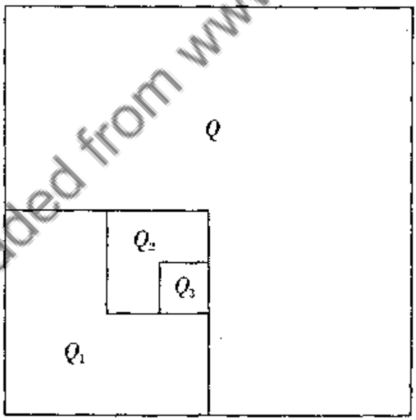

FIG. 3

PROOF. We assume E to be a plane set ; the generalisation to a general space is obvious. Since E is bounded there exists a closed square $Q \supset E$ of side a, say. We divide it into four equal closed sub-squares, each of side $\frac{1}{2}a$. Since E is infinite, at least one of them, say Q_1, contains an infinity of points of E (see Fig. 3). Dividing Q_1 again into four equal parts, we find one square $Q_2 \subset Q_1$, of side $\frac{1}{4}a$, which contains an infinity of points of E. Continuing in the same way we obtain a descending sequence of closed squares Q_k, of sides $2^{-k}a$, each of which contains an infinity of points of E. It follows from (1.8.7) and (the proof of) Theorem 1 that $Q_k \downarrow \Pi Q_k = \{P^*\}$, where P^* is a point of Q. This P^* is a limiting point of E. For, if K is any neighbourhood of P^* then, clearly, $Q_k \subset K$ if k is large enough; and hence K will contain an infinity of points of E.

As an application we prove the following generalisation of Theorem 1.

THEOREM 11. *The product of a descending sequence of bounded, closed, non-empty sets F_k is not empty.*

PROOF. We choose one point P_k of each set F_k.† Let E be the set of these points. Since the F_k descend, every P_k belongs to F_1. Hence $E \subset F_1$ and is, therefore, bounded.

If E is a finite set then at least one of its points P^*, say, will be a P_k for an infinity of suffixes k. Since the F_k descend, P^* will belong to all F_k, so that their product is not empty.

If E is infinite then there exists, by Theorem 10, at least one limiting point P^* of E. Every neighbourhood K of P^* will contain an infinity of points P_k. Since the F_k descend each such P_k belongs to all F_j with $j \leq k$. Hence K will contain an infinity of points of each F_k, so that P^* is a limiting point for each F_k. By Theorem 9, P^* belongs to each F_k since the F_k are closed. Hence $P^* \in \Pi F_k$, and this product is not empty.

† The possibility of such a simultaneous selection of points, out of an infinity of sets, is based on the so-called *axiom of choice*. The legitimacy of this axiom has been subject to logical controversies. Its use can be avoided here (*Exercise* 5).

The theorem is not true for unbounded sets. As an example take the descending sequence of closed half lines $\{x \geq k\}$ in E_1. Their product is empty. Again, the theorem is not true for open sets, as the sequence of linear open intervals $(0, 1/k)$ shows.

Exercise 6. The *diameter* of a bounded non-empty set E is defined as

$$d(E) = \sup PQ, \qquad (1.14.1)$$

where the sup is taken with respect to all points P and Q of E. Show that there exist frontier points P_0 and Q_0 of E such that $d(E) = P_0Q_0$.

Exercise 7. The *distance* between two exclusive non-empty sets E_1 and E_2 is defined as

$$\delta(E_1, E_2) = \inf PQ, \qquad (1.14.2)$$

where the inf is taken with respect to all points P of E_1 and all points Q of E_2. If E_1 is bounded, show that there exist frontier points P_0 of E_1 and Q_0 of E_2 such that $\delta(E_1, E_2) = P_0Q_0$.

1.15. Perfect sets. A set E is said to be *dense in a set* E_1 if $E_1 \subset E'$; that is, if every point of E_1 is a limiting point of E. Clearly, E is then dense in every subset of E_1. Every set E is dense in its derivative E'.

In particular, a set E is *dense in itself*, if $E \subset E'$; that is, if every point of E is a limiting point. Any sphere or interval (other than a point) is dense in itself.

A set E is called *everywhere dense* if it is dense in E. Thus the rational set R is everywhere dense. Its complement is also everywhere dense.

By Theorem 9, a set E is closed if, and only if, $E' \subset E$. A set E is said to be *perfect* if it is both closed and dense in itself, that is, if $E = E'$.

Closed spheres and intervals (other than points) are perfect sets. So are their circumferences, except in the linear space. Both E and O are perfect; for $\mathsf{E} = \mathsf{E}'$ and $\mathsf{O} = \mathsf{O}'$.

A set E is said to be *nowhere dense* if *every* open sphere

K contains an open subsphere $K^* \subset {}_cE$; that is, K^* is free of points of E.

A set may be both dense in itself and nowhere dense. Thus the circumference of a sphere in E_n is both perfect and nowhere dense, provided that $n > 1$. It is less obvious how to construct a linear perfect and nowhere dense set. The following method is due to *G. Cantor* (1845–1918), the creator of the theory of sets.

Consider the closed interval $I = \langle 0, 1 \rangle$. Let $0 < \eta \leq 1$ be given, and remove from I an open interval i_1 of centre $\frac{1}{2}$ and of length $3^{-1}\eta$. From each of the two remaining parts of I remove again a central open interval, i_2 and i_3, say, of length $3^{-2}\eta$. There remain four mutually exclusive parts of I, and again from each, central open intervals i_4, i_5, i_6, i_7 are removed, each of length $3^{-3}\eta$. Continuing in this way, a sequence of mutually exclusive open intervals $i_k = (a_k, b_k)$, say, is removed. Their total length is η.

The set $C = C_\eta$ of the remaining points of I is not empty. Thus the end-points 0 and 1 of I, and all the end-points a_k and b_k of the i_k belong to C. We shall prove that C is a perfect and nowhere dense set. It will follow that all the limiting points of the end-points a_k and b_k must also belong to C.

FIG. 4. Cantor's Set.

In Fig. 4, Cantor's original construction is sketched. Here always the "central third" is removed, $\eta = 1$, and the total length of the i_k equals 1, the length of I.

It is easy to see that C is nowhere dense. For, after the pth step of our construction there remain 2^p mutually exclusive congruent parts of I, each of length less than 2^{-p}. Hence every open interval i (a linear open "sphere"), which has points in common with C, must, clearly, contain some removed interval i_k (and even an infinity of such i_k).

Next, it follows that, if ξ ($\neq 1$) is a point of C and not itself an end-point a_k, then every interval $(\xi, \xi+h)$, where $h>0$, will contain some end-point a_k $(>\xi)$. Since the a_k belong to C and ξ is a limiting point of the a_k, ξ is a limiting point of C. Similarly, any point ξ ($\neq 0$) of C, which is not itself an end-point b_k, is a limiting point of the b_k: the set C is dense in itself.

Finally, $C = I - \Sigma i_k$, where Σi_k is open, by Theorem 5. Hence C is closed, by Theorem 7: C is perfect.

Note that the set C consists exactly of the points a_k and b_k, and of their limiting points.

Exercise 8. Show that the points of Cantor's set $(\eta=1)$ are exactly the points which admit of a ternary representation

$$\xi = \sum_1^\infty \epsilon_k 3^{-k}, \quad \epsilon_k = 0 \quad \text{or} \quad \epsilon_k = 2; \quad . \quad . \tag{1.15.1}$$

for instance, $\xi = \frac{1}{4}$.

Exercise 9. Prove that the set of all decimals (1.11.1) is perfect and nowhere dense. We know that this set is not enumerable.

The example of this exercise reveals a general property of perfect sets.

THEOREM 12. *A non-empty perfect set is non-enumerable.*

PROOF. Suppose, on the contrary, that E is non-empty and perfect, and that its points could be arranged as a sequence (P_k). We construct, by induction, a subsequence (Q_p) of (P_k) and closed spheres K_p of centres Q_p and radii ρ_p, as follows.

First $Q_1 = P_1$, and K_1 has radius $\rho_1 = 1$. If K_p has been defined, then Q_{p+1} is the *first* point ($\neq Q_p$) of the P_k that falls into the interior (K_p) of K_p; Q_p is a limiting point of E, since E is dense in itself, and hence Q_{p+1} exists. Next $\rho_{p+1} \leq 2^{-(p+1)}$ and so small that $K_{p+1} \subset (K_p)$ and K_{p+1} excludes Q_p.

By construction, no point P_k preceding Q_p falls into K_p, so that no point of E belongs to ΠK_p. On the other hand, by Theorem 11 and since $\rho_p \to 0$, this product consists of exactly one point P^* which, as limiting point of the Q_n, belongs to E, since E is closed. This contradiction proves the theorem.

1.16. Borel's covering theorem. The following theorem is of fundamental importance in general analysis. It is due to *E. Borel*.

THEOREM 13. *Suppose that* E *is bounded and closed, and that to each point* P *of* E *is attached, by some prescription, an open set* $\mathrm{O} = \mathrm{O}(\mathrm{P})$ *containing* P.

Then there exists a finite number of points $\mathrm{P}_1, \mathrm{P}_2, \ldots \mathrm{P}_m$ *of* E *such that the sum of the sets* $\mathrm{O}(\mathrm{P}_1)$, $\mathrm{O}(\mathrm{P}_2), \ldots \mathrm{O}(\mathrm{P}_m)$ *"covers" (contains)* E.

PROOF. We use Fig. 3 and the notations there, assuming that E is a plane set. Suppose that the theorem were not true. Now $E \subset Q$ where Q is a closed square of side a. At least one of the first four sub-squares (of sides $\frac{1}{2}a$), say Q_1, will have the property that an infinity of the given open sets $O(P)$ is required to cover $E \,.\, Q_1$. Clearly, $E \,.\, Q_1$ must be an infinite set. Again, at least one of the four sub-squares (of sides $\frac{1}{4}a$) of Q_1, say Q_2, must have the corresponding property; and $E \,.\, Q_2$ is infinite. Continuing in this way we obtain a descending sequence of closed squares Q_k (of side $2^{-k}a$), and each $E \,.\, Q_k$ can be covered only by an infinity of the $O(P)$. Now, the Q_k converge to a single point P^*. Since the $E \,.\, Q_k$ are infinite, P^* is a limiting point of E; and since E is closed, P^* belongs to E. But then $O(P^*)$, which as an open set contains with P^* some neighbourhood of P^*, will clearly cover Q_k and hence $E \,.\, Q_k$ if k is large enough. This gives the desired contradiction.

Obviously, the theorem is not true for unbounded sets. For, the $O(P)$ may be bounded sets so that a finite sum of them is also bounded. The condition that E be closed is also essential. Let Q be a limiting point of E which does not belong to E. If, for every P of E, $O(P)$ is the open sphere of centre P and radius $\frac{1}{2}PQ$, then, clearly, no finite sum of the O can cover E.

1.17. Functions of real variables. A real function

$$y = f(x_1, x_2, \ldots x_n) \quad . \quad . \quad (1.17.1)$$

of n real variables x_i is conveniently written in the form

$$y = f(P), \quad . \quad . \quad . \quad . \quad (1.17.2)$$

where P is a point in the space E_n. The function will be defined in a certain set E of this space.

More generally, given a system

$$y_j = f_j(x_1, x_2, \ldots x_n), \quad j = 1, 2, \ldots m, \quad . \quad (1.17.3)$$

of m functions of n variables, defined in a set E of $\mathsf{E} = \mathsf{E}_n$, we see that to each point $P = (x_i)$ of E there corresponds a point $P^* = (y_j)$ in the space $\mathsf{E}^* = \mathsf{E}_m^*$ of m dimensions. We write, therefore, (1.17.3) in the short form

$$P^* = T(P) \quad . \quad . \quad . \quad (1.17.4)$$

The point P^* is called the *image of* P, and the set E^* of all the P^* is called the *image of* E by the transformation $T(P)$. The case $m = 1$ is (1.17.2). It should be noted that the relation between P and its image P^* is not necessarily a one-one relation: the same P^* may be image of several points P.

The transformation $P^* = T(P)$ is said to be *continuous at the point* P_0 of E if

$$T(P_k) \to T(P_0), \quad \text{or} \quad P_k^* \to P_0^*, \quad . \quad (1.17.5)$$

whenever $P_k \to P_0$, the P_k being points of E.

An equivalent definition is: the transformation $P^* = T(P)$ is continuous at P_0 if, given an arbitrary (small) positive ϵ, there exists a $\delta = \delta(\epsilon, P_0)$, that is, depending on ϵ and P_0, such that

$$P^*P_0^* < \epsilon \quad \text{whenever} \quad PP_0 < \delta, \quad . \quad (1.17.6)$$

P and P_0 being points of E, and the distances taken in E^* and E, respectively.

Note that, if P_0 is an isolated point of E, *any* transformation defined in E is continuous at P_0.

The transformation $T(P)$ is said to be *continuous in E* if it is continuous at every point of E.

It should be noted that the points P_k in (1.17.5), or P in (1.17.6), are restricted to E. For instance, continuity of $f(x)$ in $\langle 0, 1\rangle$ means, by definition, " one-sided " continuity at the end-points.

All this is familiar in the case (1.17.2) and extends readily to the general case.

The definition (1.17.4) comprises the usual definition of curves and surfaces in " parameter form ". If $n=1$ and $m=3$, say, then (1.17.3) is

$$y_1=f_1(x), \quad y_2=f_2(x), \quad y_3=f_3(x). \quad . \quad (1.17.7)$$

If the three functions are continuous in a linear closed interval $\langle a, b\rangle$ (not a point), we have the definition of a *curve* in space; x is the parameter. More generally, a curve in E_m^* may be defined as the continuous image of a linear closed interval.

Again, if $n=2$ and $m=3$, the continuous image of a plane bounded region (see below) gives a *surface* in space, and this definition may also be extended to a general space.

Curves and surfaces, thus defined, may exhibit some unexpected aspects. A curve may reduce to a single point P^*, the defining functions $f_i(x)$ being constant. On the other hand, there exist curves which fill the whole of a given cube (*Peano curves*). This example shows that the same set E^* may be interpreted as curve or surface: the parameter representation determines the interpretation.

A curve for which the relation between x and P^* is a one-one relation (except perhaps for the end-points a and b of the parameter interval whose images may coincide: the curve is then closed) is called a *Jordan curve*. Jordan curves are the simplest type of curves: they have no " multiple " points. *Jordan surfaces* may be defined in a similar way.

A set E is said to be *connected* if any two of its points can be joined by a curve in E.

An open connected set D is usually called a *domain*; and its closure $\bar{D}$ is called a *region*. Open spheres and intervals are domains; closed spheres and intervals (other than points) are regions. In linear space these are the only domains and regions.

The continuous image E^{**} of a continuous image E^* of a set E is, clearly, itself a continuous image of E. In particular, the continuous image of a curve is again a curve; and similarly for surfaces.

From this it follows that *the continuous image of a connected set is again a connected set.*

THEOREM 14. *The continuous image* $\mathbf{E}^*$ *of a bounded closed set* $\mathbf{E}$ *is a bounded closed set.*

PROOF. (i) Let Q be a point of E and let Q^* be its image. Let $K^* = K^*(Q^*)$ be a neighbourhood of Q^*, of radius 1, say. By (1.17.6), there exists a neighbourhood $K = K(Q)$ of Q, of radius $\delta = \delta(Q)$, so that $P^* \in K^*$ whenever $P \in E \,.\, K$, the point P^* being the image of P. By Theorem 13, a finite number of such spheres $K(Q)$ covers E. The corresponding spheres $K^*(Q^*)$ cover E^*, so that E^* is bounded.

(ii) Let R^* be a limiting point of E^*. Then there exists a sequence of points $P_k^*(\neq R^*)$ of E^*, such that $P_k^* \to R^*$. Let $P_k^* = T(P_k)$.†

By Theorem 10, the P_k have a limiting point S, since E is bounded. Hence there exists a sub-sequence (P_{k_ν}) of the P_k such that $P_{k_\nu} \to S$. Since E is closed, $S \in E$. Hence, by (1.17.5),

$$P_{k_\nu}^* = T(P_{k_\nu}) \to T(S).$$

On the other hand, $P_{k_\nu}^* \to R^*$, so that R^* is the image of S. Hence $R^* \in E^*$: that is, E^* is closed.

† We choose *one* such P_k; there may be several.

COROLLARY. *Curves and surfaces, as continuous images of bounded regions, are bounded closed connected sets.*

Note that the continuous image of a bounded open set need neither be bounded nor open. Thus the image of $(0, 1)$ by $y = 1/x$ is not bounded, while its image by $y = \sin 1/x$ is the closed interval $\langle -1, 1\rangle$.

THEOREM 15. *Suppose that* E *is a bounded, closed, and connected set ; and that* $y = f(P)$ *is continuous in* E.

Then f *is bounded and attains its minimum* m *and its maximum* M *in* E. *It also takes in* E *every value between* m *and* M.

PROOF. The image E^* of E by f is a bounded, closed, and connected linear set. Hence E^* is the closed interval $\langle m, M\rangle$.

The number δ in (1.17.6) depends, apart from ϵ, on the point P_0 in question. The transformation $P^* = T(P)$ is said to be *uniformly continuous* in E if, given any $\epsilon (>0)$, this $\delta = \delta(\epsilon)$ can be chosen the *same* for all P in E.

We close this chapter with the following important result.

THEOREM 16. *Suppose that* E *is bounded and closed, and that* $P^* = T(P)$ *is continuous in* E. *Then* $T(P)$ *is uniformly continuous in* E.

PROOF. Let $\epsilon (>0)$ be given. To each point Q of E there corresponds, by (1.17.6), a neighbourhood $K_\delta(Q)$, of radius $\delta = \delta(Q)$, such that $P^*Q^* < \frac{1}{2}\epsilon$ whenever $P \varepsilon E \,.\, K_\delta$. The spheres $K_{\frac{1}{2}\delta}(Q)$ have, *a fortiori*, the same property. By Theorem 13, E can be covered by a finite number of these spheres $K_{\frac{1}{2}\delta}$. Let $Q_1, Q_2, \ldots Q_p$ be their centres and let $\frac{1}{2}\delta^*$ be the smallest of their radii. This δ^* will depend on ϵ only.

Now let P_0 be any point of E. It will be covered by $K_{\frac{1}{2}\delta_r}(Q_r)$, say. Consider the sphere $K_{\frac{1}{2}\delta^*}(P_0)$. Since $\delta^* \le \delta_r$, this sphere lies clearly inside $K_{\delta_r}(Q_r)$. Hence, by the definition of the latter sphere, $PP_0 < \frac{1}{2}\delta^*$ implies $P^*Q_r^* < \frac{1}{2}\epsilon$ whenever P belongs to E. In particular, $P_0^*Q_r^* < \frac{1}{2}\epsilon$. Hence, by (1.2.3), $PP_0 < \frac{1}{2}\delta^*$ implies

$$P^*P_0^* \leq P^*Q_r^* + Q_r^*P_0^* < \epsilon,$$

whenever P belongs to E. This proves the theorem, since δ^* depends on ϵ only, and not on P_0.

Solutions to Exercises

Ex. 1. The $\underline{\lim}\ C_k$ is the open circle with centre (0, 0) and radius 1; the $\overline{\lim}\ C_k$ is the same circle closed, with the exception of the two points (0, 1) and (0, −1).

Ex. 2. The $\underline{\lim}\ I_k$ may be either the closed interval $\langle 0, l \rangle$ or the half-closed interval $\langle 0, l)$. The former case will be when "eventually" all $a_k \geq l$. Similarly, $\overline{\lim}\ I_k$ is either $\langle 0, L \rangle$ or $\langle 0, L)$.

Ex. 3. By **(1.7.3)**, ${}_c\Pi_k = \sum_k^\infty {}_cE_j$. Hence

$${}_c\underline{\lim}\ E_k = {}_c\Sigma\Pi_k = \Pi\, {}_c\Pi_k = \Pi \sum_k^\infty {}_cE_j = \overline{\lim}\ {}_cE_k.$$

Ex. 4. By the corollary of Theorem 3, the set of algebraic equations with integral coefficients and of given degree n is enumerable. Since any such equation has at most n roots, the set of corresponding algebraic numbers is enumerable. The proposition now follows from Theorem 2.

Ex. 5. Determine the co-ordinate x_1 of P_k as maximum value in F_k. Then, for this x_1 fixed, determine x_2 as maximum value in F_k; and so on.

Ex. 6. There exists a sequence of points P_k and Q_k of E such that $P_kQ_k \to d$. Choosing, if necessary, a sub-sequence, we may assume that $P_k \to P^*$ and that $Q_k \to Q^*$. Consider P^*. Clearly, it cannot be an exterior point of E. Nor can it be an interior point. For, then a neighbourhood $K_\rho(P^*) \subset E$ would exist, and in it, for each k, a point R_k, such that $R_kQ_k = P^*Q_k + \frac{1}{2}\rho$, say. But $P^*Q_k > d - \frac{1}{2}\rho$ for large k, so that $R_kQ_k > d$. This is against the definition of d. Hence P^* is a frontier point; and so is Q^*.

Ex. 7. The proof is similar to that of Ex. 6. The proposition need not be true when both sets are unbounded: an example are the plane curves $y=0$ and $y=e^x$.

Ex. 8. We use the notations of § 1.15. Any number x of $\langle 0, 1\rangle$ can be represented in the form $x=\sum_1^\infty \epsilon_k 3^{-k}$ where ϵ_k is either 0, 1, or 2. Let $\langle i_k\rangle=\langle a_k, b_k\rangle$ be the closure of i_k. The points of $\langle i_1\rangle$ are characterized by $\epsilon_1=1$, the points of $\langle i_2\rangle$ and $\langle i_3\rangle$ by $\epsilon_1\neq 1$, $\epsilon_2=1$. The next group $\langle i_3\rangle$ to $\langle i_7\rangle$ has $\epsilon_1\neq 1$, $\epsilon_2\neq 1$, $\epsilon_3=1$; and so on. The end-points a_k and b_k have a second representation in which all $\epsilon_j\neq 1$. For the b_k this is obvious: $b_1=\frac{2}{3}$; $b_2=\frac{2}{9}$; $b_3=\frac{2}{3}+\frac{2}{9}$; and so on. Also $a_1=\frac{1}{3}$ $=2\Sigma 3^{-(k+1)}$; $a_2=\frac{1}{9}=2\Sigma 3^{-(k+2)}$; $a_3=\frac{2}{3}+\frac{1}{9}=\frac{2}{3}+2\Sigma 3^{-(k+2)}$; and so on. Hence on omitting the intervals i_k from $\langle 0, 1\rangle$ we retain exactly the points of the form (1.15.1).

The point $\frac{1}{4}=2\Sigma 3^{-2k}$ belongs to this set.

Ex. 9. Any limiting point of points (1.11.1) is, clearly, again such a point: E is closed. Also $\xi=.\ \epsilon_1\ \epsilon_2\ \ldots\ \epsilon_k\ \ldots$ is the limit of the points $\xi_k=.\ \epsilon_1\ \epsilon_2\ \ldots\ \epsilon_k$: E is dense in itself. Hence E is perfect.

Let (a, b) be any open interval containing a certain point $\xi=.\ \epsilon_1\ \epsilon_2\ \ldots\ \epsilon_k\ \ldots$ of our set. Then, if k is large enough, it will also contain the interval (α, β) where $\alpha=.\ \epsilon_1\ \epsilon_2\ \ldots\ \epsilon_k 2$ and $\beta=.\ \epsilon_1\ \epsilon_2\ \ldots\ \epsilon_k 9$. This interval contains no point of the set. E is nowhere dense.

CHAPTER II

CONTENT

2.1. The problem of volume. The length and area of a plane polygon, and the surface area and the volume of a polyhedron in space are familiar geometrical concepts, all easy of definition. We also have more or less " intuition " of the meaning of such expressions as the length of a curve, the area of a plane domain, the area of a surface in space, or the volume of a body. Roughly speaking, we think of some sort of approximation to these more general geometrical configurations by means of the elementary " polyhedral " ones. Thus the length of a curve is thought of as the limit of the lengths of approximating polygons, or the volume of a body as the limit of the volumes of approximating polyhedra. The definition of the perimeter of a circle, and of the area enclosed, by means of approximating polygons, will be well known to the reader.

Naturally, it is a major mathematical problem, of the greatest importance in the application of Geometry to Science, to give a precise meaning to this vague idea of approximation. In ancient Mathematics such approximation, or " exhaustion " as it is sometimes called, was a well-established geometrical principle for the calculation of such lengths, areas, and volumes as occurred in applications. The modern theory of " measure " is but the conscious and systematic application of this principle to general sets of points.

In this book we restrict ourselves to the theory of " volume ". It will be convenient to employ the common name " *volume* " to include also the " length " of a linear set and the "area" of a plane set. This is in accordance with our earlier notation (1.3.2) in the case of intervals. When specialising we shall, of course, retain the usual words length and area.

2.2. Postulates. We consider throughout a given space $\mathsf{E}=\mathsf{E}_n$ of n dimensions, and we wish to establish the notion of volume for a "general" set E in E. That is, we wish to attach to E a certain number $V(E)$ as its volume. If this name is to correspond to our intuitive idea of volume, these numbers $V(E)$ must satisfy the following four postulates:

I. *Volume is non-negative:* $\mathrm{V}(\mathrm{E})\geq 0$.

II. *Volume is additive, i.e. the volume of a finite sum of mutually exclusive sets equals the sum of the volumes of these sets:*

$$V(E_1+E_2+\ldots+E_m) = V(E_1)+V(E_2)+\ldots+V(E_m), \quad (2.2.1)$$

if $E_i \,.\, E_k=0$. This implies

$$V(E_1-E_2)=V(E_1)-V(E_2) \quad . \quad . \quad (2.2.2)$$

and, in conjunction with Postulate I,

$$V(E_2)\leq V(E_1) \quad \text{if} \quad E_2\subset E_1. \quad . \quad . \quad (2.2.3)$$

III. *For an "elementary" set* P *(linear interval, plane polygon, polyhedron in space, general interval)* $\mathrm{V}(\mathsf{P})$ *has its elementary value.*

IV. *Congruent sets have the same volume.*

The last postulate is of main importance for the ordinary spaces ($n\leq 3$), where congruence is an elementary notion. In a general E_n, the set E is said to be *congruent* to E^*, in symbols $E\sim E^*$, if the points $P=(x_i)$ of E correspond to the points $P^*=(x_i{}^*)$ of E^* by a normal orthogonal linear transformation

$$x_i{}^*=a_{i1}x_1+a_{i2}x_2+\ldots+a_{in}x_n+b_i, \quad 1\leq i\leq n, \quad (2.2.4)$$

where

$$a_{i1}a_{j1}+a_{i2}a_{j2}+\ .\ .\ .+a_{in}a_{jn}=\begin{cases}1 \text{ if } i=j\\ 0 \text{ if } i\neq j\end{cases},\ 1\leq i,\ j\leq n\quad . \qquad (2.2.5)$$

Postulate IV also implies that the volume $V(E)$ does not depend on the choice of the Cartesian co-ordinate system employed, once a unit length has been chosen.

We do not know offhand whether a really comprehensive definition of volume is possible. As a matter of fact, none of the subsequent definitions is comprehensive. It is, however, a tacitly assumed " fifth Postulate " that a definition of volume should cover as many sets E as possible.

In order to satisfy Postulate III any general definition of volume must be built upon the volumes of certain elementary sets which we may call *primary sets* (of the definition) : a general set E will have to be approximated by these primary sets. The effectiveness of the definition will depend on the initial choice of these sets.

In this chapter we develop the theory of a definition which is based on the most " natural " choice of primitive sets : approximation by polyhedral sets. This is the ancient principle of exhaustion, though its final form was established only towards the end of the last century by *G. Peano* (1887) and *C. Jordan* (1892). We shall see that this definition has, from the purely mathematical point of view, considerable deficiencies, and it will be replaced, in the next chapter, by a more satisfactory definition. Nevertheless, it has now become " classical ", and it is certainly sufficient for all purposes of application to Science. The volume thus defined is usually known as the Peano-Jordan *content*, and we shall, accordingly, replace $V(E)$ by $c(E)$ throughout this chapter.

2.3. Interval sums. We start with closed intervals $I=\langle a_i,\ b_i\rangle$ and their volumes $|\ I\ |=\Pi(b_i-a_i)$, as defined in (1.3.2). Some or all of the edges b_i-a_i may be zero, in which case $|\ I\ |=0$.

As primary sets we take all finite sums

$$S = I_1 + I_2 + \ldots + I_m \quad . \quad . \quad . \quad (2.3.1)$$

of closed intervals I_k. The I_k may overlap. We call these sets *interval sums*.

Any finite set is an S, the "intervals" I_k being the points of the set. Fig. 5 shows a more general interval

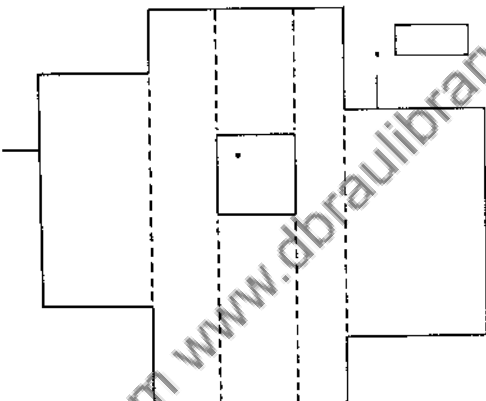

FIG. 5. Plane interval sum.

sum in the plane. For formal reasons, the null set O is also considered as a set S.

In the plane it is clear (see the dotted lines in Fig. 5), and in a general E_n it is an elementary proposition, that any interval sum S can be expressed, in an infinity of ways, as a sum

$$S = J_1 + J_2 + \ldots + J_m, \quad . \quad . \quad . \quad (2.3.2)$$

where any two of the closed intervals J_k are *separate*: that is, they have at most frontier points in common. It is then an elementary proposition of co-ordinate algebra that the number

$$| S | = | J_1 | + | J_2 | + \ldots + | J_m | \quad . \quad . \quad (2.3.3)$$

does not depend on the choice of the J_k. We call $|S|$ the *volume* of the interval sum S.

We shall also take for granted the following elementary property of interval sums:

The sum S_1+S_2 *and the product* $S_1 . S_2$ *of two interval sums* S_1 *and* S_2 *are (possibly empty) interval sums.*

This extends to finite sums and products of interval sums. Further

$$|S_1+S_2|+|S_1 . S_2|=|S_1|+|S_2|. \quad . \quad (2.3.4)$$

This equality expresses the simple fact that each of the interval sums S_1, S_2 and S_1+S_2 covers $S_1 . S_2$ once. The proof is elementary.

The frontier S_f of an interval sum S is itself an interval sum, and $|S_f|=0$. It follows easily, from this and (2.3.4), that

$$|S_1+S_2+\ldots+S_m|=|S_1|+|S_2|+\ldots+|S_m| \quad (2.3.5)$$

whenever the S_k are mutually *separate* (have at most frontier points in common).

The difference S_1-S_2 of two interval sums need not be an interval sum. Thus the difference of two (different) intervals is not closed and hence is not an interval sum. But $\overline{S_1-S_2}$, the closure of S_1-S_2, is an interval sum; and S_2 and $\overline{S_1-S_2}$ are separate, so that, by (2.3.5),

$$|\overline{S_1-S_2}|=|S_1|-|S_2|. \quad . \quad . \quad (2.3.6)$$

Finally, we have $|S_2|\leq|S_1|$ if $S_2\subset S_1$.

2.4. Outer content of bounded sets. Given a set E in E, we wish to approximate to E by interval sums "from outside", that is, by sums $S\supset E$. Whatever the definition of the volume may be, we must have $|S|\geq V(E)$, by (2.2.3) and Postulate III, since the S are elementary sets. It is, therefore, natural to take

$$\bar{c}(E)=\inf_{S\supset E}|S| \quad . \quad . \quad . \quad (2.4.1)$$

as the first step towards the definition of the content. This number $\bar{c}(E)$ is called the *outer content* of E.

It should be noted that this definition presupposes that E is bounded, because any S is a bounded set. We make this assumption throughout this paragraph. Clearly,

$$\bar{c}(S) = |S|, \qquad (2.4.2)$$

so that, in particular, $\bar{c}(E) = 0$ when E is a finite set. Next, $\bar{c}(E) \geq 0$ and

$$\bar{c}(E_2) \leq \bar{c}(E_1) \quad \text{if} \quad E_2 \subset E_1 \qquad (2.4.3)$$

For, any S containing E_1 contains E_2.

Hence $\bar{c}(E)$ satisfies Postulate I and (2.2.3).

If $\bar{E}$ is the closure of E, then

$$\bar{c}(\bar{E}) = \bar{c}(E). \qquad (2.4.4)$$

For, $E \subset \bar{E}$ and thus $\bar{c}(E) \leq \bar{c}(\bar{E})$. Conversely, $S \supset E$ implies $S \supset \bar{E}$ since S is closed. Hence $\bar{c}(\bar{E}) \leq \bar{c}(E)$.

Let $\mathsf{R}_I = \mathsf{R} \,.\, I$ where R is the rational set of § 1.10 and I is a closed interval. The closure of both R_I and $I - \mathsf{R}_I$ is I. Hence

$$\bar{c}(\mathsf{R}_I) = \bar{c}(I - \mathsf{R}_I) = |I|. \qquad (2.4.5)$$

THEOREM 17.

$$\bar{c}(E_1 + E_2) + \bar{c}(E_1 \,.\, E_2) \leq \bar{c}(E_1) + \bar{c}(E_2). \qquad (2.4.6)$$

PROOF. If $S_1 \supset E_1$ and $S_2 \supset E_2$ then $S_1 + S_2 \supset E_1 + E_2$ and $S_1 \,.\, S_2 \supset E_1 \,.\, E_2$. Hence

$$\bar{c}(E_1 + E_2) + \bar{c}(E_1 \,.\, E_2) \leq |S_1 + S_2| + |S_1 \,.\, S_2| = |S_1| + |S_2|,$$

by (2.4.1) and (2.3.4). Since $|S_1|$ and $|S_2|$ can be chosen arbitrarily close to $\bar{c}(E_1)$ and $\bar{c}(E_2)$ respectively, we obtain (2.4.6).

It is an easy corollary that

$$\bar{c}(E_1 + E_2 + \ldots + E_m) \leq \bar{c}(E_1) + \bar{c}(E_2) + \ldots + \bar{c}(E_m). \qquad (2.4.7)$$

Next,

$$\bar{c}(E_1 + E_2) = \bar{c}(E_1) \quad \text{if} \quad \bar{c}(E_2) = 0. \qquad (2.4.8)$$

For, in this case,

$$\bar{c}(E_1) \leq \bar{c}(E_1 + E_2) \leq \bar{c}(E_1) + \bar{c}(E_2) = \bar{c}(E_1).$$

The outer content is not additive. Thus $I = R_I + (I - R_I)$, and R_I and $I - R_I$ are exclusive. Yet $\bar{c}(R_I) + \bar{c}(I - R_I) = 2|I|$, by (2.4.5).

We shall say that two sets E_1 and E_2 are *separated by an interval sum* S_0, if $E_1 \subset S_0$, say, and E_2 is outside S_0, except perhaps for common frontier points with S_0.

THEOREM 18. *If* E_1 *and* E_2 *are separated by some interval sum* S_0, *then*

$$\bar{c}(E_1 + E_2) = \bar{c}(E_1) + \bar{c}(E_2). \qquad (2.4.9)$$

PROOF. Let $S \supset E_1 + E_2$. Then $S_1 = S \cdot S_0 \supset E_1$

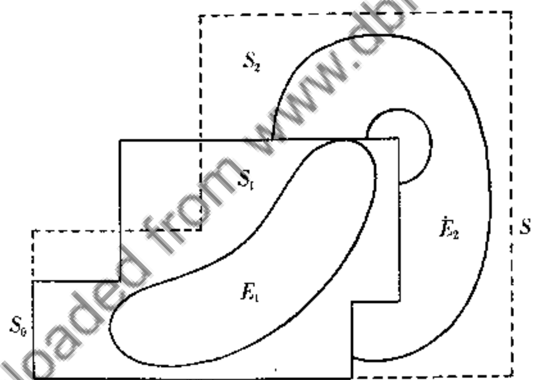

FIG. 6

and $S_2 = \overline{S - S_1} \supset E_2$ (Fig. 6.). Hence, by (2.3.6),

$$|S| = |S_1| + |S_2| \geq \bar{c}(E_1) + \bar{c}(E_2).$$

Since $S \supset E_1 + E_2$, but is otherwise arbitrary, we obtain $\bar{c}(E_1 + E_2) \geq \bar{c}(E_1) + \bar{c}(E_2)$. The opposite inequality holds by (2.4.7).

2.5. Outer content (general). We wish to extend the definition of the outer content to unbounded sets.

Suppose that an ascending sequence of closed intervals I_k is given, such that

$$I_k \uparrow \mathsf{E}\,; \qquad . \quad . \quad . \quad . \quad (2.5.1)$$

that is, the sequence of the I_k fills the whole space. If

$$E^{(k)} = E \,.\, I_k, \quad \bar{c}_k = \bar{c}(E^{(k)}), \qquad . \quad . \quad (2.5.2)$$

then $E^{(k)} \uparrow E$ and $\bar{c}_k$ increases. We define the *outer content* of E by

$$\bar{c}(E) = \lim \bar{c}(E^{(k)}). \qquad . \quad . \quad . \quad (2.5.3)$$

This outer content may be infinite.

For bounded sets E this definition agrees with our former one since then $E^{(k)} = E$ for large k.

To justify our definition we have still to show that it does not depend on the choice of the I_k. Consider a second ascending sequence of intervals $I_k{}^*$ filling the space. Then, with obvious notations,† $E^{(k)*} \subset E^{(K)}$ for fixed k and sufficiently large K. Hence $\bar{c}_k{}^* \le \bar{c}_K \le \bar{c}(E)$, and thus $\bar{c}^*(E) \le \bar{c}(E)$. The opposite inequality is, of course, also true.

We leave it to the reader, as an easy exercise, to verify that all the properties of the outer content, obtained in the preceding paragraph, remain valid for unbounded sets.

2.6. Inner content. We next approximate to E by interval sums "from inside" and define accordingly

$$\underline{c}(E) = \sup_{S \subset E} |\, S \,|. \qquad . \quad . \quad . \quad (2.6.1)$$

as the *inner content* of E. The set E need not be bounded. If E is unbounded, its inner content may be infinite. Clearly,

$$\underline{c}(S) = \underline{c}(S_i) = |\, S \,|. \qquad . \quad . \quad . \quad (2.6.2)$$

Also $\underline{c}(E) \ge 0$ and

$$\underline{c}(E_2) \le \underline{c}(E_1) \quad \text{if} \quad E_2 \subset E_1. \qquad . \quad . \quad (2.6.3)$$

† The * relates to the sequence of the $I_k{}^*$.

For, any S contained in E_2 is also contained in E_1.

If E_i is the interior of E, then

$$\underline{c}(E_i) = \underline{c}(E). \quad . \quad . \quad . \quad . \quad (2.6.4)$$

For, let $S \subset E$. Then $S_i \subset E_i$, so that, by (2.6.2),

$$|S| = \underline{c}(S_i) \leq \underline{c}(E_i).$$

Since $S \subset E$, but is otherwise arbitrary, this implies $\underline{c}(E) \leq \underline{c}(E_i)$. The opposite inequality follows from $E_i \subset E$.

To (2.4.5) corresponds

$$\underline{c}(R_I) = \underline{c}(I - R_I) = 0. \quad . \quad . \quad . \quad (2.6.5)$$

For, the interiors of both sets are empty.

THEOREM 19.

$$\underline{c}(E_1 + E_2) + \underline{c}(E_1 \,.\, E_2) \geq \underline{c}(E_1) + \underline{c}(E_2). \quad . \quad (2.6.6)$$

The proof is much the same as that of Theorem 17 except that interior interval sums are used.

THEOREM 20. *If* E_1 *and* E_2 *are separated by some interval sum* S_0, *then*

$$\underline{c}(E_1 + E_2) = \underline{c}(E_1) + \underline{c}(E_2). \quad . \quad . \quad (2.6.7)$$

PROOF. (Fig. 6). First, $\underline{c}(E_1 + E_2) \geq \underline{c}(E_1) + \underline{c}(E_2)$, by (2.6.6), since $\underline{c}(E_1 \,.\, E_2) = 0$.

Let $S \subset E_1 + E_2$. Then $S_1 = S \,.\, S_0 \subset E_1$. Also $(S_2) \subset E_2$, where (S_2) is the interior of $S_2 = \overline{S - S_1}$. Hence, by (2.3.6) and (2.6.2),

$$|S| = |S_1| + |S_2| = |S_1| + \underline{c}((S_2)) \leq \underline{c}(E_1) + \underline{c}(E_2),$$

which implies $\underline{c}(E_1 + E_2) \leq \underline{c}(E_1) + \underline{c}(E_2)$ and completes the proof.

We observe next that

$$\underline{c}(E) \leq \bar{c}(E). \quad . \quad . \quad . \quad . \quad (2.6.8)$$

For, $|S| = \bar{c}(S) \leq \bar{c}(E)$ whenever $S \subset E$.

THEOREM 21. *If* $E \subset I$ *then*

$$\underline{c}(E) = |I| - \bar{c}(I - E). \qquad (2.6.9)$$

PROOF. Let $S_1 \subset E$ so that $I - S_1 \supset I - E$. Now S_1 and $I - S_1$ are separated by S_1. Hence, by (2.4.2) and Theorem 18,

$$|I| - |S_1| = \bar{c}(I - S_1) \geq \bar{c}(I - E), \quad |S_1| \leq |I| - \bar{c}(I - E),$$

from which follows

$$\underline{c}(E) \leq |I| - \bar{c}(I - E).$$

Next, let $I \supset S_2 \supset I - E$ so that $I - S_2 \subset E$. Now S_2 and $I - S_2$ are separated by S_2. Hence, by (2.6.2) and Theorem 20,

$$|I| - |S_2| = \underline{c}(I - S_2) \leq \underline{c}(E),$$

which implies

$$\underline{c}(E) \geq |I| - \bar{c}(I - E).$$

This completes the proof.

Exercise 10. Prove that $\underline{c}(E_1 + E_2) = \underline{c}(E_1)$ if $\bar{c}(E_2) = 0$. Give an example to show that $\underline{c}(E_2) = 0$ is not sufficient for this.

Exercise 11. If $E^{(k)}$ is defined as in (2.5.2), prove that

$$\underline{c}(E) = \lim \underline{c}(E^{(k)}) \qquad (2.6.10)$$

2.7. Content. We know that $\underline{c}(E) \leq \bar{c}(E)$. From this it is clear how content must be defined: approximation from inside and outside should give the same result. We say, therefore, that the set E has *content* $c(E)$ if

$$\underline{c}(E) = \bar{c}(E) \quad (= c(E)). \qquad (2.7.1.)$$

There is, however, an important restriction. *If* $c(E) = \infty$ *we shall not say that* E *has content except when every product* E . I, *where* I *is any closed interval, has a (finite) content.*

The reason for this restriction is that $c(E) = \infty$ clearly implies $c(E_1) = \infty$ whenever $E_1 \supset E$. Such a set E_1 may be as "irregular" as we please, while the existence of content implies, as we shall see, a certain regularity of the set. Also most of the following theorems cease to hold without this restriction.

Note that, in order to prove (2.7.1), it is sufficient to show that $\underline{c}(E) \geq \bar{c}(E)$.

Any interval sum S has content and

$$c(S) = |S| \quad . \quad . \quad . \quad . \qquad (2.7.2.)$$

Also $c(E) \geq 0$ and

$$c(E_2) \leq c(E_1) \quad \text{if} \quad E_2 \subset E_1, \quad . \quad . \qquad (2.7.3)$$

so that Postulate I and (2.2.3) are satisfied.

If $\bar{c}(\mathrm{E}) = 0$ *then* $c(\mathrm{E}) = 0$. Thus the frontier S_f of an S, or any subset of S_f, has content zero. Any finite set has content zero.

The space E itself has infinite content.

THEOREM 22. *If* E_1 *and* E_2 *have contents, then so have* $\mathrm{E}_1 + \mathrm{E}_2$ *and* $\mathrm{E}_1 \,.\, \mathrm{E}_2$. *Also* †

$$c(E_1 + E_2) + c(E_1 \,.\, E_2) = c(E_1) + c(E_2). \quad . \qquad (2.7.4)$$

PROOF. (i) First,

$$\begin{aligned} \underline{c}(E_1 + E_2) + \underline{c}(E_1 \,.\, E_2) &\geq c(E_1) + c(E_2) \\ &\geq \bar{c}(E_1 + E_2) + \bar{c}(E_1 \,.\, E_2), \quad . \quad . \qquad (a) \end{aligned}$$

by Theorems 17 and 19. Since neither term on the left-hand side is greater than the corresponding term on the right, we conclude that

$$\underline{c}(E_1 + E_2) = \bar{c}(E_1 + E_2), \quad \underline{c}(E_1 \,.\, E_2) = \bar{c}(E_1 \,.\, E_2),$$

provided that all these numbers are finite. In this case $E_1 + E_2$ and $E_1 \,.\, E_2$ have finite contents; and (2.7.4) follows from (a).

(ii) If a set E has content then, with the notation (2.5.2), all sets $E^{(k)}$ have contents. This follows either from (i) if $c(E)$ is finite; or it is part of the definition of infinite

† For infinite contents this has the obvious meaning.

content. The converse is also true: if *all* sets $E^{(k)}$ have contents then E has content, and

$$c(E) = \lim c(E^{(k)}). \quad . \quad . \quad . \quad (2.7.5)$$

This follows from (2.5.3) and (2.6.10).

(iii) Now, consider the general case of the theorem. By (ii) the sets $E_1^{(k)}$ and $E_2^{(k)}$ have contents. By (i), their sum and product, that is, $(E_1+E_2)^{(k)}$ and $(E_1 . E_2)^{(k)}$ have also contents: and these satisfy (2.7.4). Hence E_1+E_2 and $E_1 . E_2$ have contents which satisfy (2.7.4), as is seen on using (2.7.5).

THEOREM 23. *Content is additive (it satisfies Postulate II):*

$$c(E_1+E_2+\ldots+E_m) = c(E_1)+c(E_2)+\ldots+c(E_m), \quad (2.7.6)$$

provided that the E_k *are mutually exclusive and have contents.*

This is a corollary of Theorem 22.

Exercise 12. Suppose that $\underline{c}(E)=\infty$, and that all sets $E^{(k)}$ have contents for some *given* sequence of intervals $I_k (\uparrow E)$. Show that E has infinite content.

THEOREM 24. *If* E_1 *and* E_2 *have contents, and* $E_2 \subset E_1$, *then* $E_1 - E_2$ *has content. Also*

$$c(E_1 - E_2) = c(E_1) - c(E_2), \quad . \quad . \quad (2.7.7)$$

provided that $c(E_2) < \infty$. †

PROOF. (i) Suppose, first, that E_1 is bounded, and let $E_2 \subset E_1 \subset I$. If $D = E_1 - E_2$, then, by Theorems 17 and 21,

$$\begin{aligned} \bar{c}(I-D) &= \bar{c}((I-E_1)+E_2) \leq \bar{c}(I-E_1)+c(E_2) \\ &= |I| - c(E_1) + c(E_2), \\ \underline{c}(D) &= |I| - \bar{c}(I-D) \geq c(E_1) - c(E_2). \end{aligned}$$

On the other hand, E_2 and $I - E_1$ are exclusive. Hence, by Theorems 19 and 21,

† $\infty - \infty$ is meaningless.

$$\underline{c}(I-D)=\underline{c}((I-E_1)+E_2)\geq \underline{c}(I-E_1)+c(E_2)$$
$$=|I|-c(E_1)+c(E_2),$$
$$\bar{c}(D)=|I|-\underline{c}(I-D)\leq c(E_1)-c(E_2).$$

It follows that D has a content which satisfies (2.7.7).

(ii) If E_1 is unbounded, we consider once more the bounded sets $E_1^{(k)}$ and $E_2^{(k)}$. Their difference is $D^{(k)}$ and has, by (i), a content which satisfies (2.7.7). On letting $k\to\infty$ we obtain (2.7.7), provided that $c(E_2)$ is finite.

THEOREM 25. *If* E *has content, then its interior* E_i *and its closure* $\bar{E}$ *have the same content :*

$$c(E_i)=c(E)=c(\bar{E}). \qquad (2.7.8)$$

This is an easy consequence of (2.4.4) and (2.6.4).

THEOREM 26. *A set* E *has content if, and only if, its frontier* E_f *has content zero.*

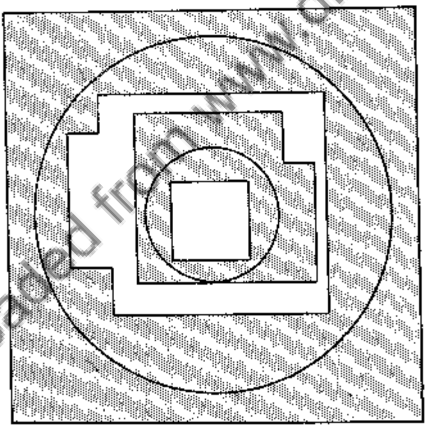

FIG. 7

PROOF. We may assume that E is bounded ; the extension to unbounded sets follows the now usual lines.†

† Note that $(E_f)^{(k)}\subset(E^{(k)})_f$.

(i) If E has content then $E_f = \bar{E} - E_i$ has content zero, by Theorems 24 and 25.

(ii) Suppose that $c(E_f) = 0$. Given any $\epsilon (>0)$, there exists an $S \supset E_f$ such that $|S| < \epsilon$.

The (open) complement of S consists of a finite number of connected parts. Each of these must totally belong either to E_i or to E_e. For, otherwise it would contain frontier points of E not covered by S. (Compare Fig. 7, where the shaded part is S and E is a circular ring.) The set $S^* = S + E_i$ is, therefore, itself an interval sum; its frontier is part of S_f. Also $E \subset S^*$ and $S^* - S \subset E$. Hence

$$\bar{c}(E) \leq |S^*| = c(S^* - S) + |S| < \underline{c}(E) + \epsilon.$$

Since ϵ is arbitrarily small, this implies $\bar{c}(E) \leq \underline{c}(E)$ and so $\underline{c}(E) = \bar{c}(E)$.

Exercise 13. Prove that the formula (2.7.6) holds whenever the E_k are mutually *separate* and have contents.

2.8. Content of elementary sets. Any efficient definition of volume must cover *elementary* sets like polyhedra, spheres, ellipsoids, etc., as they occur in applications. We have, therefore, to investigate whether the content satisfies this test of efficiency.

Consider a function

$$y = f(x_1, x_2, \ldots x_n) = f(P), \qquad (2.8.1)$$

defined in a set E in $\mathsf{E} = \mathsf{E}_n$. The set of points

$$P^* = (x_1, x_2, \ldots x_n, y) = G(P) \qquad (2.8.2)$$

in $\mathsf{E}^* = \mathsf{E}^*_{n+1}$, where $P = (x_i)$ is a point in E and $y = f(P)$, is called the *graph* of the function $y = f(P)$.

THEOREM 27. *If* G *(in* E^**) is the graph of a function* y = f(P), *continuous in a closed set* E *(in* E*), then* c(G) = 0.

PROOF. We may suppose that E is bounded. Let A be the cube $\langle |x_i| \leq a \rangle$ in E, and let $E \subset \langle |x_i| \leq \frac{1}{2}a \rangle$. Given

$\epsilon(>0)$, we can find, for every point P_0 of E, a closed cube Q_{P_0}, of centre P_0 and of edge $\delta=\delta(\epsilon)$ such that

$$|f(P)-f(P_0)|\leq\epsilon \quad . \quad . \quad . \quad . \quad \text{(a)}$$

whenever $P \varepsilon E \,.\, Q_{P_0}$. We may assume that $\delta \leq a$, so that $Q_{P_0} \subset A$.

Now, Theorem 13 is applicable: the open sets there are the interiors of the cubes Q_P. A finite number of such cubes, $Q_1, Q_2, \ldots Q_r$ say, will cover E, and their sum will be contained in A.

By (a), that part of the graph which corresponds to the points of $E \,.\, Q_k$ is, clearly, covered by a closed cube $Q_k{}^*$, in E^*, of volume $|Q_k{}^*|\leq 2\epsilon|Q_k|$. The sum of these $Q_k{}^*$ covers the whole graph G. The "projection" of this sum on the subspace E is the sum of the Q_k, so that, plainly,

$$|Q_1{}^*+Q_2{}^*+\ldots+Q_r{}^*|\leq 2\epsilon|Q_1+Q_2+\ldots+Q_r|\leq 2\epsilon|A|.$$

Hence $\bar{c}(G)\leq 2\epsilon|A|$. This proves the theorem since ϵ is arbitrarily small.

THEOREM 28. *Suppose that the frontier of a set* E* (*in* E^*) *is the sum of a finite number of graphs* (*of functions continuous in closed sets*). *Then* E* *has content.*

This is a corollary of Theorems 26 and 27. It shows that all the elementary sets in ordinary space which are likely to occur in applications have contents. Thus, the circumference of a sphere is the sum of two graphs.

Part of the frontier of an elementary set in space may consist of a surface generated by straight lines perpendicular to the (x, y)-plane (cylindrical surface). Such a part is congruent to a graph and has content zero, according to our next theorem. A body of this type will, therefore, also have content.

A "general" interval $\mathfrak{J}$, that is, a set congruent to an I, is of the type covered by Theorem 28, and hence has content. If $\mathfrak{J}\sim I$ then $|\mathfrak{J}|=|I|$ is its elementary volume. Also

$$c(\mathfrak{J})=|\mathfrak{J}|. \quad . \quad . \quad . \quad . \quad (2.8.3)$$

For, if $S_1 \subset \mathfrak{J} \subset S_2$, then it is an elementary proposition of co-ordinate algebra that $|S_1| \leq |\mathfrak{J}| \leq |S_2|$. This implies (2.8.3).

By a similar argument it could be shown, without having recourse to Theorem 28, that $\mathfrak{J}$ has content.

Let $\mathfrak{S}$ be a "general" interval sum, that is, a set congruent to an S. Then

$$c(\mathfrak{S}) = |\mathfrak{S}|, \quad . \quad . \quad . \quad . \quad (2.8.4)$$

where $|\mathfrak{S}|$ is the elementary volume of $\mathfrak{S}$.

This follows from (2.8.3) and from the representation (2.3.2) of an interval sum. Next,

$$\underline{c}(E) = \sup_{\mathfrak{S} \subset E} |\mathfrak{S}|, \quad \bar{c}(E) = \inf_{\mathfrak{S} \supset E} |\mathfrak{S}|, \quad . \quad . \quad (2.8.5)$$

where the second formula holds for bounded E.

For, let $l = \sup |\mathfrak{S}|$, where $\mathfrak{S} \subset E$. First, $\mathfrak{S} \subset E$ implies $|\mathfrak{S}| = c(\mathfrak{S}) \leq \underline{c}(E)$ so that $l \leq \underline{c}(E)$. On the other hand, $\underline{c}(E) \leq l$ since every S is an $\mathfrak{S}$. The proof for the second formula is similar.

The general interval sums $\mathfrak{S}$, occurring in (2.8.5), need no longer have their edges parallel to the co-ordinate axes. This clearly implies the following result.

Theorem 29. *Congruent sets have the same outer and inner contents. In particular, the content satisfies Postulate IV.*

In the case of the plane it is easy to verify that content also satisfies the remaining Postulate III; that is, that plane polygons have their elementary areas as contents. The existence of these contents follows from Theorem 28.

It may suffice to sketch the proof. Since a polygon is the sum of a finite number of separate triangles, it is enough to show that a closed triangle Δ has the correct content. Now, on adding to Δ an adjacent congruent triangle Δ^*, a "general" interval $\mathfrak{J}$ is obtained. Hence, by Theorems 22 and 29 (see Exercise 13),

$$|\mathfrak{J}| = c(\mathfrak{J}) = c(\Delta) + c(\Delta^*) = 2c(\Delta)$$

as desired.

In space the verification that the content satisfies Postulate III is best obtained by the results of the integral calculus, as based on the notion of content.

Exercise 14. Let P denote a plane polygon, or a polyhedron in space. Then (in plane or space)

$$\underline{c}(E) = \sup_{P \subset E} |P|, \quad \bar{c}(E) = \inf_{P \supset E} |P|, \quad . \quad . \quad (2.8.6)$$

where $|P|$ denotes the elementary area, or volume, of P. The second formula holds for bounded E only.

These formulae show that our original definitions (2.4.1) and (2.6.1) actually amount to an approximation by polygons, or polyhedra.

2.9. Deficiencies. The definition of content is, in spite of its apparent efficiency, not really satisfactory from the purely mathematical point of view. The main application of "volume" to mathematical analysis is the theory of the integral (or vice versa). Here it is desirable that a really satisfactory definition of volume should have the following limiting property:

$$V(E_k) \to V(E) \quad \text{if} \quad E_k \uparrow E. \quad . \quad . \quad (2.9.1)$$

We cannot expect (2.9.1) to hold for a descending sequence of sets, at least not without some restriction. Thus the linear sets $\langle x \geq k \rangle$, of infinite lengths, descend as $k \to \infty$ to the empty set whose length is zero.

It would be a consequence of (2.9.1) that *any enumerable set* E *should have volume zero.* For, if E consists of the sequence (P_k), and if $E_k = \{P_1, P_2, \ldots P_k\}$, then $E_k \uparrow E$. Since $V(E_k) = 0$, by Postulate III, it follows from (2.9.1) that $V(E) = 0$.

The content does not satisfy (2.9.1). Consider the enumerable set R_I of § 2.4. By (2.4.5) and (2.6.5) this set has no content when $|I|>0$. The rational set R itself has also no content: in fact, $\underline{c}(R)=0$ and $\bar{c}(R)=\infty$. The sets $I-R_I$ and ${}_cR$ are examples of non-enumerable sets without content.

Exercise 15. Open and closed sets are, in many respects, the simplest types of sets. One would wish that an efficient definition of volume should cover these sets. The definition of content does not do so.

Let (P_k), $k\geq 1$, denote the sequence of the points of $R_{(I)}$, the set of the rational points contained in the open interval (I). Let Q_k be an open sphere of centre P_k, contained in (I), and of volume $|Q_k|\leq 2^{-(k+1)}|I|$.

Show that the open set $O=\Sigma Q_k$ has no content if $|I|>0$. The closed set $I-O$ has also no content.

Solutions to Exercises

Ex. 10. (i) We may assume that $E_1+E_2\subset I$. By (2.4.8), we have $\bar{c}(I-E_1)=\bar{c}(I-(E_1+E_2))$, since the two sets differ by a set of outer content zero. Hence,

$$\underline{c}(E_1+E_2)=|I|-\bar{c}(I-(E_1+E_2))=|I|-\bar{c}(I-E_1)=\underline{c}(E_1).$$

(ii) The interval I is the sum of R_I and $I-R_I$, both of inner content zero, by (2.6.5). The content of I, however, is positive.

Ex. 11. Let l be the limit (2.6.10). First, $l\leq\underline{c}(E)$ since $E^{(k)}\subset E$. On the other hand, $S\subset E$ implies $S\subset E^{(k)}$, and thus $|S|\leq\underline{c}(E^{(k)})$, for all large k. It follows that $\underline{c}(E)\leq l$.

Ex. 12. First, $\underline{c}(E)=\bar{c}(E)=\infty$. We have to show that $E\,.\,I$ has content for every interval I. Now, $I\subset I_k$ and $E\,.\,I=E^{(k)}\,.\,I$ for large k. Hence $E\,.\,I$ has content, by Theorem 22.

Ex. 13. If E_1 and E_2 are separate, then $E_1\,.\,E_2$ is the common part of their frontiers. Hence $c(E_1\,.\,E_2)=0$, by Theorem 26, and thus $c(E_1+E_2)=c(E_1)+c(E_2)$, by (2.7.4). The general formula follows by induction.

Ex. 14. Let l be the sup in (2.8.6). Then $\underline{c}(E)\leq l$ since every S is also a P. On the other hand, $P \subset E$ implies $|P| \leq \underline{c}(E)$; and, therefore, $l \leq \underline{c}(E)$. The proof for $\bar{c}(E)$ is similar.

Ex. 15. First, $O \supset R_{(I)}$ so that $\bar{c}(O) \geq \bar{c}(R_{(I)}) = |I| > 0$. On the other hand, if $S \subset O$, then, by Theorem 13, S can be covered by a finite number of the spheres Q_k. Hence

$$|S| \leq \Sigma |Q_k| \leq |I| \,.\, \Sigma 2^{-(k+1)} = \tfrac{1}{2}|I|,$$

so that $\underline{c}(O) \leq \frac{1}{2}|I|$.

CHAPTER III

MEASURE

3.1. A new postulate. The main reason why the Peano-Jordan content is, mathematically, not a satisfactory definition of volume is that it does not satisfy the limit relation (2.9.1). If we apply the latter to the ascending sequence of sums $E_1 + E_2 + \ldots + E_k$ of mutually exclusive sets E_k, we are led, in conjunction with Postulate II, to formulate the following new postulate for a satisfactory definition of volume.

Postulate II*. *Volume is fully additive; i.e., the volume of a finite or infinite sum* † *of mutually exclusive sets equals the sum of the volumes of these sets:*

$$V(E_1 + E_2 + \ldots + E_k + \ldots) \\ = V(E_1) + V(E_2) + \ldots + V(E_k) + \ldots \quad . \quad (3.1.1)$$

We know that content does not satisfy this postulate: if the E_k have contents, their infinite sum need not have a content. Thus the rational set R, considered as the sum of its (enumerable) points, which have contents zero, has itself no content (§ 2.9).

3.2. Interval sets. It is clear that, in order to satisfy Postulate II*, the volumes of the primary sets employed in the definition of the general volume should themselves satisfy this postulate. In particular, an infinite sum of primary sets should be a primary set itself. Now, an

† Infinite sums comprise finite sums: the E_k are eventually empty.

infinite sum of interval sums S is not necessarily an interval sum. It is, therefore, natural to extend the class of primary sets so as to include infinite sums of interval sums. The new primary sets are, accordingly, of the form

$$\Sigma = I_1 + I_2 + \ldots + I_k + \ldots = \Sigma I_k, \quad . \quad (3.2.1)$$

where the I_k, as usual, denote closed intervals.† We call these sets *interval sets*. They were first considered by *E. Borel* (1898). Interval sums S are special interval sets, the I_k eventually being " empty " intervals.

Any enumerable set is an interval set: the I_k are points or empty.

A finite or infinite sum of interval sets is an interval set.

This is obvious. Next, we prove :

A finite product of interval sets is a (possibly empty) interval set.

For, if $\Sigma = \Sigma I_k$ and $\Sigma^* = \Sigma I_l^*$, then $\Sigma \,.\, \Sigma^* = \sum_{k,l} I_k \,.\, I_l^*$; and $I_k \,.\, I_l^*$ is a (possibly empty) interval.

Theorem 30. *Any open set is an interval set.*

Proof. We may assume that the open set O is neither empty nor the whole space. The rational points in O, *i.e.* the points in $O.R$, can be arranged as a sequence (P_k). Each P_k is centre of a closed sphere $K_k \subset O$ of radius half the distance of P_k from the frontier O_f. Let Q_k be the closed cube of centre P_k inscribed to K_k.

If $P \in O$ then, choosing P_k near enough to P, Q_k will cover P. Hence $O \subset \Sigma Q_k$. On the other hand, $\Sigma Q_k \subset O$. Hence $O = \Sigma Q_k$, and O is an interval set.

A closed set need not be an interval set.

Take as example the plane segment

$$L = \{0 \leq x \leq 1, \quad y = x\}.$$

If L were a plane interval set, its constituent intervals would have to be points. Since L is not enumerable, this is impossible.

† We admit now the empty set as an " interval ".

By the same argument no (non-empty) perfect and nowhere dense set can be an interval set. For it is not enumerable (Theorem 12).

The difference of two interval sets need not be an interval set.

For, let O be a plane open set containing the above segment L. Then $O_1 = O - L$ is open, and $L = O - O_1$ appears as the difference of two interval sets.

An infinite product of interval sets need not be an interval set.

Divide L into 2^k parts of equal length, and let S_k be the interval sum consisting of the 2^k squares with these parts as diagonals. Clearly, $S_k \supset S_{k+1}$ and $L = \Pi S_k$.

3.3. Volume of interval sets. If we wish to employ interval sets as primary sets we must first assign a volume to them. We need the following generalisation of (2.3.2).

THEOREM 31. *Any interval set can be represented in the form*

$$\Sigma = \sum_1^\infty J_k, \qquad (3.3.1)$$

where the J_k *are mutually separate closed intervals.*

PROOF. This is true for an interval sum S. We also note that $I - (I \,.\, S)$, the complement with respect to I of the interior $(I \,.\, S)$ of the interval sum $I \,.\, S$, is an interval sum.

Let $\Sigma = \sum_1^\infty I_k$ and $S_m = \sum_1^m I_k$. We construct (3.3.1) by induction.

Put $J_1 = I_1$, and if a representation

$$S_m = \sum_1^{K_m} J_k$$

has been found, take any representation

$$D_m = I_{m+1} - (I_{m+1} \,.\, S_m) = \sum_{K_m+1}^{K_{m+1}} J_k$$

of this interval sum D_m. Since D_m and S_m are separate, the same will hold for any two of the J_k, $1 \leq k \leq K_{m+1}$.

Their sum is S_{m+1}, and the inductive construction is complete.

Of course, there are an infinity of representations (3.3.1).

We define now the *volume* of the interval set Σ as the (possibly infinite) number

$$|\Sigma| = \sum_1^\infty |J_k|. \quad . \quad . \quad . \quad (3.3.2)$$

To justify this definition, we have to show that it does not depend on the choice of the representation (3.3.1). This is a consequence of the following theorem.

THEOREM 32. *Suppose that*

$$\sum_1^\infty J_k \subset \sum_1^\infty I_k, \quad . \quad . \quad . \quad (3.3.3)$$

where the J_k *are mutually separate. Then*

$$\sum_1^\infty |J_k| \leq \sum_1^\infty |I_k|. \quad . \quad . \quad . \quad (3.3.4)$$

In particular,

$$\sum_1^\infty |J_k| = \sum_1^\infty |J_k^*|, \quad . \quad . \quad . \quad (3.3.5)$$

if $\Sigma J_k = \Sigma J_k^*$.

PROOF. The theorem is elementary for interval sums S.

Now, given an $\epsilon (>0)$, let (I_k) be an open interval, concentric with, and similar to, I_k; and let it be of volume $|(I_k)| = |I_k| + 2^{-k}\epsilon$. The interval sum $S_m = \sum_1^m J_k$ is covered by the sum of all these (I_k). By Theorem 13 it is even covered by a finite number of them, say by the first k_m of the (I_k). Hence

$$|S_m| = \sum_1^m |J_k| \leq \sum_1^{k_m} |(I_k)| \leq \sum_1^\infty |(I_k)| = \sum_1^\infty |I_k| + \epsilon,$$

for every m. It follows that

$$\sum_1^\infty |J_k| \le \sum_1^\infty |I_k| + \epsilon$$

for every $\epsilon(>0)$. This proves (3.3.4).

Immediate consequences of Theorem 32 are

$$|\Sigma_2| \le |\Sigma_1| \quad \text{if} \quad \Sigma_2 \subset \Sigma_1 \qquad . \qquad . \quad (3.3.6)$$

and

$$|\Sigma_1 + \Sigma_2 + \ldots + \Sigma_k + \ldots| \\ \le |\Sigma_1| + |\Sigma_2| + \ldots + |\Sigma_k| + \ldots . \quad . \quad (3.3.7)$$

Also $|\mathbf{E}| = 0$ *if* $\mathbf{E}$ *is enumerable.* For, an enumerable set is a Σ consisting of points as intervals.

We prove next that

$$|\Sigma| = \underline{c}(\Sigma). \quad . \quad . \quad . \quad . \quad (3.3.8)$$

For, if $S \subset \Sigma$, then $|S| \le |\Sigma|$, by (3.3.6). Hence $\underline{c}(\Sigma) \le |\Sigma|$. On the other hand, $S_m = \sum_1^m J_k \subset \Sigma$ and $|S_m| = \sum_1^m |J_k| \to |\Sigma|$ as $m \to \infty$, so that $\underline{c}(\Sigma) \ge |\Sigma|$.

THEOREM 33.

$$|\Sigma_1 + \Sigma_2| + |\Sigma_1 . \Sigma_2| = |\Sigma_1| + |\Sigma_2|. \quad . \quad (3.3.9)$$

In particular,

$$|\Sigma_1 + \Sigma_2| = |\Sigma_1| + |\Sigma_2| \quad \text{if} \quad |\Sigma_1 . \Sigma_2| = 0. \quad (3.3.10)$$

PROOF. First, $|\Sigma_1 + \Sigma_2| + |\Sigma_1 . \Sigma_2| \ge |\Sigma_1| + |\Sigma_2|$, by (3.3.8) and Theorem 19. The opposite inequality is certainly true when $|\Sigma_1|$ or $|\Sigma_2|$ is infinite. If both are finite, we write, for $p = 1$ or 2,

$$\Sigma_p = \sum_1^\infty J_k^{(p)}, \quad S_m^{(p)} = \sum_1^m J_k^{(p)}, \quad R_m^{(p)} = \sum_{m+1}^\infty J_k^{(p)},$$

so that $|S_m^{(p)}| \to |\Sigma_p|$ and $|R_m^{(p)}| \to 0$ as $m \to \infty$.
Now,

$$|\Sigma_1 + \Sigma_2| \le |S_m^{(1)} + S_m^{(2)}| + |R_m^{(1)}| + |R_m^{(2)}|$$

by (3.3.7) ; and

$$|\Sigma_1 . \Sigma_2| \leq |S_m^{(1)} . S_m^{(2)}| + |R_m^{(1)}| + |R_m^{(2)}|,$$

since $\Sigma_1 . \Sigma_2 \subset S_m^{(1)} . S_m^{(2)} + R_m^{(1)} + R_m^{(2)}$. On adding, we obtain

$$|\Sigma_1 + \Sigma_2| + |\Sigma_1 . \Sigma_2| \leq |S_m^{(1)}| + |S_m^{(2)}| + 2(|R_m^{(1)}| + |R_m^{(2)}|),$$

by (2.3.4). Finally, as $m \to \infty$, we find the desired opposite inequality.

3.4. Outer measure. The definition of volume based on interval sets as primary sets is due to *H. Lebesgue* (1902), a pupil of *Borel*. This volume is called *measure*.

First we define, in close analogy to (2.4.1), the *outer measure* of a set E, in a given space E, as the number

$$\bar{m}(E) = \inf_{\Sigma \supset E} |\Sigma|. \quad . \quad . \quad . \quad (3.4.1)$$

The set E need not be bounded, and $\bar{m}(E)$ may be infinite. Clearly,

$$\bar{m}(\Sigma) = |\Sigma|. \quad . \quad . \quad . \quad . \quad (3.4.2)$$

In particular, *any enumerable set has outer measure zero.* Also $\bar{m}(E) \geq 0$ and

$$\bar{m}(E_2) \leq \bar{m}(E_1) \quad \text{if} \quad E_2 \subset E_1. \quad . \quad . \quad (3.4.3)$$

Next,

$$\bar{m}(E_1 + E_2 + \ldots + E_k + \ldots) \leq \bar{m}(E_1) + \bar{m}(E_2) + \ldots + \bar{m}(E_k) + \ldots . \quad . \quad (3.4.4)$$

For, given $\epsilon(>0)$, take sets $\Sigma_k \supset E_k$ such that $|\Sigma_k| \leq \bar{m}(E_k) + 2^{-k}\epsilon$. Then $\Sigma = \sum_{k=1}^{\infty} \Sigma_k \supset \sum_{k=1}^{\infty} E_k$, so that, by (3.4.3) and (3.3.7),

$$\bar{m}\left(\sum_{k=1}^{\infty} E_k\right) \leq \bar{m}(\Sigma) = \left|\sum_{k=1}^{\infty}\right| \leq \sum |\Sigma_k| \leq \sum_{k=1}^{\infty} \bar{m}(E_k) + \epsilon,$$

for every $\epsilon(>0)$.

Theorem 34.

$$\bar{m}(E_1 + E_2) + \bar{m}(E_1 . E_2) \leq \bar{m}(E_1) + \bar{m}(E_2). \quad . \quad . \quad (3.4.5)$$

The proof is the same as that of the corresponding Theorem 17 except that we use interval sets Σ instead of interval sums S.

Also

$$\bar{m}(E_1+E_2)=\bar{m}(E_1) \quad \text{if} \quad \bar{m}(E_2)=0, \qquad (3.4.6)$$

with a proof similar to that of (2.4.8).

The following definition of outer measure is equivalent to (3.4.1); in it the general sets Σ are restricted to be open sets O:

$$\bar{m}(E)=\inf_{O\supset E}|O|=\inf_{O\supset E}\underline{c}(O). \qquad (3.4.7)$$

First, we note that $|O|=\underline{c}(O)$, by (3.3.8). Next, if L denotes the above inf, then $L\geq\bar{m}(E)$, since every O is a Σ. On the other hand, let $\Sigma\supset E$, and let $\Sigma=\sum_{k=1}^{\infty}J_k$ where the J_k are separate. If $\epsilon(>0)$ is given, let (I_k) be an open interval, concentric with, and similar to, J_k, and of volume $|(I_k)|=|J_k|+2^{-k}\epsilon$. Then $O=\sum_{k=1}^{\infty}(I_k)$ is an open set containing E. Hence

$$L\leq|O|\leq\sum_{k=1}^{\infty}|J_k|+\epsilon=|\Sigma|+\epsilon.$$

Here ϵ is arbitrarily small, and we conclude that $L\leq|\Sigma|$. Since $\Sigma\supset E$ but is otherwise arbitrary, it follows that $L\leq\bar{m}(E)$.

THEOREM 35. *Congruent sets have the same outer measure.*

PROOF. Both the definitions of an open set O and of its volume $|O|=\underline{c}(O)$ are independent of the choice of the co-ordinates in E, by Theorem 29. The theorem, therefore, follows from (3.4.7).

The following theorem corresponds to Theorem 18.

THEOREM 36. *If* E_1 *and* E_2 *are separated by some interval sum* S_0, *then*

$$\bar{m}(E_1+E_2)=\bar{m}(E_1)+\bar{m}(E_2). \qquad (3.4.8)$$

PROOF. (Compare Fig. 6.) First, we observe that the frontier of S_0, itself an interval sum, has by (3.4.2), outer measure zero. Suppose that $E_1 \subset S_0$. We may also assume that E_1 and E_2 have no points on the frontier of S_0. For, the sets of such points have outer measure zero, as subsets of the frontier of S_0; and their omission does not alter either side of (3.4.8), according to (3.4.6).

Now, let O be an open set containing E_1+E_2. If (S_0) denotes the interior of S_0, then $O_1=O\,.\,(S_0) \supset E_1$; and $O_2=O-\bar{O}_1 \supset E_2$, where $\bar{O}_1$ is the closure of O_1. Both O_1 and O_2 are open and they are exclusive. Also $O_1+O_2 \subset O$, so that, by (3.3.10),

$$\bar{m}(E_1)+\bar{m}(E_2) \leq |\,O_1\,|+|\,O_2\,|=|\,O_1+O_2\,| \leq |\,O\,|.$$

Since $O \supset E_1+E_2$ but is otherwise arbitrary, this implies $\bar{m}(E_1)+\bar{m}(E_2) \leq \bar{m}(E_1+E_2)$, by (3.4.7). The opposite inequality follows from (3.4.4).

It is an easy corollary that

$$\bar{m}(E_1+E_2+\ldots+E_k+\ldots) \\ =\bar{m}(E_1)+\bar{m}(E_2)+\ldots\,\bar{m}(E_k)+\ldots, \qquad (3.4.9)$$

whenever any two of the E_k are separated by some interval sum. For then, for every p,

$$\bar{m}(E_1+E_2+\ldots+E_k+\ldots) \geq \bar{m}(E_1+E_2+\ldots+E_p) \\ =\bar{m}(E_1)+\bar{m}(E_2)+\ldots+\bar{m}(E_p),$$

by an obvious extension of (3.4.8); and this gives, as $p \to \infty$, the opposite of (3.4.4).

Another obvious consequence of Theorem 36 is that

$$\bar{m}(E_1-E_2)=\bar{m}(E_1)-\bar{m}(E_2) \quad . \quad . \quad (3.4.10)$$

whenever $E_2 \subset E_1$, and E_2 and E_1-E_2 are separated by some interval sum; provided, of course, that $\bar{m}(E_2)<\infty$.

Next, let $E^{(k)}$ be defined as in (2.5.2). Then

$$\bar{m}(E)=\lim \bar{m}(E^{(k)}). \quad . \quad . \quad . \quad (3.4.11)$$

For,

$$E=E^{(1)}+(E^{(2)}-E^{(1)})+\ldots+(E^{(k)}-E^{(k-1)})+\ldots.$$

Here any two terms on the right are separated by one of the intervals I_k, and (3.4.10) can be applied to each difference. Hence, by (3.4.9),

$$\bar{m}(E) = \bar{m}(E^{(1)}) + (\bar{m}(E^{(2)}) - \bar{m}(E^{(1)})) + \ldots + (\bar{m}(E^{(k)}) - \bar{m}(E^{(k-1)})) + \ldots$$

which is equivalent to (3.4.11).

Formula (3.4.11) often permits one, in the now customary way, to restrict proofs to bounded sets. Thus the inequality

$$\bar{m}(E) \leq \bar{c}(E) \quad . \quad . \quad . \quad (3.4.12)$$

is plain for bounded sets because every S is a Σ. For unbounded sets it follows from (3.4.11).

We know that $\bar{m}(O) = | O | = \underline{c}(O)$. If F is closed then

$$\bar{m}(F) = \bar{c}(F). \quad . \quad . \quad . \quad (3.4.13)$$

For, we may assume that F is bounded. Let $\Sigma = \sum_{k=1}^{\infty} J_k \supset F$, and let (I_k) be defined as in the proof of Theorem 34. A finite number of these (I_k) will cover F, by Theorem 13. Let S be the sum of their closures I_k. Then $S \supset F$ and $| S | \leq | \Sigma | + \epsilon$. Hence $\bar{c}(F) \leq | \Sigma | + \epsilon$ and so, by the now familiar argument, $\bar{c}(F) \leq \bar{m}(F)$. The opposite inequality also holds, by (3.4.12).

3.5. Inner measure of bounded sets. The reader will expect the definition of inner measure to be analogous to the definition of inner content. This is, however, not the case, for reasons which we shall discuss later. It is rather in analogy to (2.6.9) that we proceed.

Throughout this paragraph we restrict ourselves to bounded sets. The theorems, however, will be stated for general sets; the extension of the proofs to unbounded sets will be obvious from the definition of inner measure for such a set in the next paragraph.

Suppose that $E \subset I$. We define the number

$$\underline{m}(E) = | I | - \bar{m}(I - E) \quad . \quad . \quad . \quad (3.5.1)$$

as the *inner measure* of the *bounded* set E.

To justify this definition we have to show that it is independent of the choice of the interval I. First, let $I_0 \supset I$. The two terms in the sums

$$I_0 - E = (I_0 - I) + (I - E), \quad I_0 = (I_0 - I) + I$$

are separated by I. Hence, by Theorem 36 and (3.4.2),

$$\bar{m}(I_0 - E) = \bar{m}(I_0 - I) + \bar{m}(I - E), \quad |I_0| = \bar{m}(I_0 - I) + |I|.$$

From this follows

$$|I_0| - \bar{m}(I_0 - E) = |I| - \bar{m}(I - E).$$

Now, let I^* be any interval containing E. We can then find an I_0 such that $I_0 \supset I$ and $I_0 \supset I^*$; and the desired result follows from the above.

We observe, first, that $\underline{m}(E) \geq 0$. For, $I - E \subset I$ implies $\bar{m}(I - E) \leq |I|$.

Next,

$$\underline{m}(E_2) \leq \underline{m}(E_1) \quad \text{if} \quad E_2 \subset E_1. \quad . \quad . \quad (3.5.2)$$

For, if $I \supset E_1$, then $I - E_2 \supset I - E_1$ and so $\bar{m}(I - E_2) \geq \bar{m}(I - E_1)$.

THEOREM 37.

$$\underline{m}(E_1 + E_2) + \underline{m}(E_1 \,.\, E_2) \geq \underline{m}(E_1) + \underline{m}(E_2). \quad . \quad (3.5.3)$$

PROOF. Suppose that $E_1 + E_2 \subset I$. Then, by (1.7.2),

$$I - (E_1 + E_2) = (I - E_1) \,.\, (I - E_2),$$
$$I - E_1 \,.\, E_2 = (I - E_1) + (I - E_2).$$

The left-hand side of (3.5.3) equals, therefore,

$$2|I| - \bar{m}[(I - E_1) \,.\, (I - E_2)] - \bar{m}[(I - E_1) + (I - E_2)];$$

and this, by Theorem 34, is not smaller than

$$2|I| - \bar{m}(I - E_1) - \bar{m}(I - E_2)$$

which is the right-hand side of (3.5.3).

Exercise 16. Prove that

$$\underline{m}(E_1 + E_2) = \underline{m}(E_1) \quad \text{if} \quad \bar{m}(E_2) = 0 \quad . \quad . \quad (3.5.4)$$

THEOREM 38. *If* E_1 *and* E_2 *are separated by some interval sum* S_0, *then*

$$\underline{m}(E_1+E_2)=\underline{m}(E_1)+\underline{m}(E_2). \quad . \quad . \quad (3.5.5)$$

PROOF. (Compare Fig. 6.) Suppose that $E_1 \subset S_0$ and that $I \supset S_0+E_2$. The set $I-(S_0+E_2)$ is separated from both S_0 and S_0-E_1 by S_0. Hence, by Theorem 36,

$$\bar{m}(I-(S_0+E_2))+\bar{m}(S_0)=\bar{m}(I-E_2),$$
$$\bar{m}(I-(S_0+E_2))+\bar{m}(S_0-E_1)=\bar{m}(I-(E_1+E_2)),$$

so that

$$\begin{aligned}\bar{m}(S_0)-\bar{m}(S_0-E_1)&=\bar{m}(I-E_2)-\bar{m}(I-(E_1+E_2))\\&=\underline{m}(E_1+E_2)-\underline{m}(E_2). \quad . \quad . \quad (a)\end{aligned}$$

For similar reasons

$$\bar{m}(I-S_0)+\bar{m}(S_0)=|I|, \quad \bar{m}(I-S_0)+\bar{m}(S_0-E_1)=\bar{m}(I-E_1)$$

so that

$$\bar{m}(S_0)-\bar{m}(S_0-E_1)=|I|-\bar{m}(I-E_1)=\underline{m}(E_1). \quad . \quad (b)$$

The two results prove the theorem.

A simple corollary is that

$$\underline{m}(E_1-E_2)=\underline{m}(E_1)-\underline{m}(E_2) \quad . \quad . \quad (3.5.6)$$

whenever $E_2 \subset E_1$ and E_2 and E_1-E_2 are separated by some interval sum S_0, provided that $\underline{m}(E_2)<\infty$.

To (3.4.9) corresponds

$$\begin{aligned}&\underline{m}(E_1+E_2+\ldots+E_k+\ldots)\\&=\underline{m}(E_1)+\underline{m}(E_2)+\ldots+\underline{m}(E_k)+\ldots \quad . \quad (3.5.7)\end{aligned}$$

whenever any two of the E_k are separated by some interval sum.

This is clear for a finite sum. In the infinite case, however, we shall prove it, for the sake of simplicity, only under a somewhat stricter assumption, i.e., that in

$$E=\sum_1^\infty E_k=\sum_1^l E_k+\sum_{l+1}^\infty E_k=A_l+B_l$$

the sets A_l and B_l are separated, for every l, by some interval sum: this implies the former condition. We also assume that E is bounded. By (3.5.5),

$$\underline{m}(E) = \underline{m}(A_l) + \underline{m}(B_l) = \sum_1^l \underline{m}(E_k) + \underline{m}(B_l) \quad . \quad . \quad (a)$$

Also, by (3.4.9), $\bar{m}(E) = \Sigma \bar{m}(E_k) < \infty$ so that $\bar{m}(B_l) = \sum_{l+1}^{\infty} \bar{m}(E_k)$ $\to 0$ as $l \to \infty$. Hence $\underline{m}(B_l) \to 0$, and (a) yields (3.5.7), on letting $l \to \infty$.

Next,

$$\underline{c}(E) \leq \underline{m}(E) \leq \bar{m}(E) \leq \bar{c}(E). \quad . \quad . \quad (3.5.8)$$

For, let $E \subset I$. First, $\bar{m}(I - E) \leq \bar{c}(I - E)$, by (3.4.12), and hence $\underline{m}(E) \geq | I | - \bar{c}(I - E) = \underline{c}(E)$, by Theorem 21. Next, $| I | \leq \bar{m}(E) + \bar{m}(I - E)$, by (3.4.4). In other words, $\underline{m}(E) \leq \bar{m}(E)$. Finally, $\bar{m}(E) \leq \bar{c}(E)$ by (3.4.12).

If Σ is an interval set, then

$$\underline{m}(\Sigma) = \underline{c}(\Sigma) = | \Sigma |. \quad . \quad . \quad . \quad (3.5.9)$$

For, by (3.3.8), (3.4.2), and (3.5.8),

$$| \Sigma | = \underline{c}(\Sigma) \leq \underline{m}(\Sigma) \leq \bar{m}(\Sigma) = | \Sigma |.$$

To the formula (3.4.7) corresponds

$$\underline{m}(E) = \sup_{F \subset E} \bar{c}(F) \quad . \quad . \quad . \quad (3.5.10)$$

where the F are closed sets; in particular,

$$\underline{m}(F) = \bar{c}(F). \quad . \quad . \quad . \quad (3.5.11)$$

For, let $E \subset (I)$ where (I) is the interior of I. If $F \subset E$ then $(I) - F$ is an open set containing $(I) - E$. Hence †

$$\underline{c}(I - F) \geq \underline{c}((I) - F) = \bar{m}((I) - F) \geq \bar{m}((I) - E) = \bar{m}(I - E)$$

by (3.4.6) and (3.4.7), since the frontier of I has outer measure zero. Substituting into $\bar{c}(F) = | I | - \underline{c}(I - F)$ we obtain

$$\bar{c}(F) \leq | I | - \bar{m}(I - E) = \underline{m}(E).$$

This implies $\underline{m}(E) \geq \sup \bar{c}(F)$.

† Actually, $\underline{c}(I - F) = \underline{c}((I) - F)$, by Exercise 10.

On the other hand, given $\epsilon(>0)$, we can choose an open set $O \supset I - E$ such that $|O| \leq \bar{m}(I-E) + \epsilon$. The set $F = I - O \,.\, I = I \,.\, {}_cO$ is closed and contained in E. Hence

$$\bar{c}(F) = |I| - \underline{c}(O \,.\, I) \geq |I| - \underline{c}(O)$$
$$= |I| - |O| \geq |I| - \bar{m}(I-E) - \epsilon = \underline{m}(E) - \epsilon.$$

Since ϵ is arbitrary this implies $\sup \bar{c}(F) \geq \underline{m}(E)$, and (3.5.10) follows.

The definitions of a closed set and of outer content are independent of the choice of co-ordinates in E. Hence (3.5.10) implies the following result.

Theorem 39. *Congruent sets have the same inner measure.*

It is of interest to discuss why the definition

$$\underline{m}(E) = \sup_{\Sigma \subset E} |\Sigma|, \quad . \quad . \quad . \quad (3.5.12)$$

which would correspond to (2.6.1), is not satisfactory.

It is clear that $\underline{m}(E) \leq \underline{m}(E)$ since $|\Sigma| \leq \underline{m}(E)$ if $\Sigma \subset E$. In fact, we can have $\underline{m}(E) < \underline{m}(E)$. As an example, consider the set $I - R_I$ of § 2.4. Since R_I is enumerable, $\bar{m}(R_I) = 0$. Hence

$$\underline{m}(I - R_I) = |I| - \bar{m}(R_I) = |I|.$$

On the other hand, if $\Sigma \subset I - R_I$, then, by (2.6.5),

$$|\Sigma| = \underline{c}(\Sigma) \leq \underline{c}(I - R_I) = 0,$$

so that $\underline{m}(I - R_I) = 0$.

Approximation by interval sets from inside provides, therefore, a less close approach to volume than our definition of inner measure.

By (3.5.10), closed sets may be used as primary sets from inside if the outer content is taken as their volume.

Exercise 17. Prove that, if $\bar{m}(E_1 \,.\, E_2) = 0$, then

$$\underline{m}(E_1 + E_2) \leq \underline{m}(E_1) + \bar{m}(E_2) \leq \bar{m}(E_1 + E_2) \quad . \quad (3.5.13)$$

3.6. Inner measure (general). The inner measure of a general set E is defined as

$$\underline{m}(E) = \lim \underline{m}(E^{(k)}), \quad . \quad . \quad . \quad (3.6.1)$$

where the $E^{(k)}$ are the sets (2.5.2).

We leave it to the reader to verify, in the now familiar way, that this definition is consistent with our previous one in the case of a bounded set; that it is independent of the choice of the I_k; and that the results of the previous paragraph hold for unbounded sets (compare § 2.5).

3.7. Measure. We know that $\underline{m}(E) \leq \bar{m}(E)$. The set E is said to be *measurable* and to have *measure* $m(E)$ if

$$\underline{m}(E) = \bar{m}(E) \quad (= m(E)). \qquad (3.7.1)$$

However, as in the case of infinite content, we add the following restriction: *If* $m(E) = \infty$ *we shall not say that* E *is measurable except when every product* E . I, *where* I *is any closed interval, has (finite) measure.* The reader should keep this restriction well in mind.

First, $m(E) \geq 0$ and

$$m(E_2) \leq m(E_1) \quad \text{if} \quad E_2 \subset E_1. \qquad (3.7.2)$$

Next, $\bar{m}(E) = 0$ implies $m(E) = 0$. *In particular, every enumerable set has measure zero.*

Every interval set Σ, and thus every open set O, is measurable; and

$$m(\Sigma) = |\Sigma|, \qquad (3.7.3)$$

by (3.4.2) and (3.5.9).

Every closed set F is measurable and

$$m(F) = \bar{c}(F), \qquad (3.7.4)$$

by (3.4.13) and (3.5.11).

That measure is more powerful than content, but consistent with it, follows clearly from (3.5.8). We state this as a theorem.

THEOREM 40. *If* E *has content, then it is also measurable and* $m(E) = c(E)$. *The reverse is not true.*

Thus open and closed sets are measurable but need not have contents (Exercise 15); the rational set R has no content but has, as an enumerable set, measure zero.

The following two theorems are proved in the same way as the corresponding theorems on content.

THEOREM 41. *If* E_1 *and* E_2 *are measurable, then* $E_1 + E_2$ *and* $E_1 . E_2$ *are measurable. Also*

$$m(E_1 + E_2) + m(E_1 . E_2) = m(E_1) + m(E_2). \quad (3.7.5)$$

THEOREM 42. *If* E_1 *and* E_2 *are measurable, and* $E_2 \subset E_1$, *then* $E_1 - E_2$ *is measurable. Also*

$$m(E_1 - E_2) = m(E_1) - m(E_2) \quad (3.7.6)$$

provided that $m(E_2) < \infty$.

We observe, finally, that the definition of measure does not depend on the choice of the co-ordinate system in E. For, this is true for both inner and outer measure. These results show that measure satisfies, as content does, all the postulates of § 2.2.

3.8. Limiting properties of measure. We shall now show that measure satisfies the new Postulate II* of § 3.1.

THEOREM 43. *Suppose that the sets* E_k *of a given sequence are measurable. Then their sum*

$$E = E_1 + E_2 + \ldots + E_k + \ldots$$

is also measurable; and

$$m(E) \leq m(E_1) + m(E_2) + \ldots + m(E_k) + \ldots. \quad (3.8.1)$$

If, in addition, $m(E_k . E_l) = 0$ *for* $k \neq l$ *then*

$$m(E) = m(E_1) + m(E_2) + \ldots + m(E_k) + \ldots; \quad (3.8.2)$$

that is, measure is fully additive.

PROOF. We may, as usual, assume that $\bar{m}(E) < \infty$.

(i) We prove the last clause first. By (3.4.4),

$$\bar{m}(E) \leq \Sigma m(E_k).$$

On the other hand,

$$m(E_1 + E_2 + \ldots + E_p) = m(E_1) + m(E_2) + \ldots + m(E_p)$$

for every p, if $m(E_k \, . \, E_l) = 0$ for $k \neq l$. This is an obvious corollary of Theorem 41. Also $m(E_1 + E_2 + \ldots + E_p) \leq \underline{m}(E)$, by (3.5.2). Hence we obtain, on letting $p \to \infty$,

$$\underline{m}(E) \geq \Sigma m(E_k).$$

This proves (3.8.2).

(ii) To prove (3.8.1), we put

$$E_k{}^* = (E_1 + E_2 + \ldots + E_{k-1}) \, . \, E_k$$

for $k \geq 2$. The sets $E_k{}^*$ are measurable, by Theorem 41. Also

$$E = E_1 + (E_2 - E_2{}^*) + \ldots + (E_k - E_k{}^*) + \ldots,$$

the terms of this sum being measurable and mutually exclusive. Hence E is measurable, by (3.8.2); and

$$m(E) = m(E_1) + m(E_2 - E_2{}^*) + \ldots \\ + m(E_k - E_k{}^*) + \ldots \leq \Sigma m(E_k).$$

THEOREM 44. *If the sets* $\mathrm{E_k}$ *are measurable, then their product* $\Pi = \prod_1^\infty \mathrm{E}_k$ *is also measurable.*

PROOF. The complement of Π is ${}_c\Pi = \Sigma_c E_k$, by (1.7.3). Each set ${}_cE_k$ is measurable as complement of E_k with respect to the measurable space E, by Theorem 42. Hence ${}_c\Pi$ is measurable, by Theorem 43; and Π itself is measurable, again by Theorem 42.

The following theorem is of particular importance. It shows that measure has all the desired efficiency as regards limiting operations. It is because of this efficiency that it has widely superseded the older notion of content in pure mathematics.

THEOREM 45. *Suppose that the sets* $\mathrm{E_k}$ *are measurable. Then their limiting sets* $\underline{\lim}\, \mathrm{E_k}$ *and* $\overline{\lim}\, \mathrm{E_k}$ *are also measurable. Also* †

$$m\,(\underline{\lim}\, E_k) \leq \underline{\lim}\, m(E_k). \quad . \quad . \quad (3.8.3)$$

† The inequality (3.8.3) is known as *Fatou's Lemma* (for sets).

Again

$$m(\overline{\lim}\, E_k) \geq \overline{\lim}\, m(E_k), \qquad . \quad . \quad (3.8.4)$$

provided, however, that

$$m(E_p + E_{p+1} + \ldots) < \infty \qquad . \quad . \quad (3.8.5)$$

for some p.

In particular, if $E_k \to E$ *then* E *is measurable ; and*

$$m(E_k) \to m(E), \qquad . \quad . \quad . \quad (3.8.6)$$

provided that (3.8.5) *is satisfied.*

If $E_k \uparrow E$ *then no such restriction is required. If* $E_k \downarrow E$, *then* (3.8.5) *is equivalent to* $m(E_p) < \infty$, *for some* p.

PROOF. (i) First, suppose that $E_k \uparrow E$ so that $m(E_k)$ increases. Then

$$E = E_1 + (E_2 - E_1) + \ldots + (E_k - E_{k-1}) + \ldots,$$

where the terms of the series are mutually exclusive measurable sets, by Theorem 42. Hence E is measurable, by Theorem 43.

Now, if $m(E_p) = \infty$ for some p, then $m(E) = \infty$ and (3.8.6) is obvious. Otherwise

$$m(E) = m(E_1) + (m(E_2) - m(E_1)) + \ldots + (m(E_k) - m(E_{k-1})) + \ldots = \lim m(E_k),$$

by (3.8.2) and (3.7.6).

(ii) Next, let $E_k \downarrow E$ so that $m(E_k)$ decreases. Then $E = \Pi E_k$ is measurable, by Theorem 44. Clearly,

$$E_p = E + (E_p - E_{p+1}) + (E_{p+1} - E_{p+2}) + \ldots,$$

for every p. Hence, if $m(E_p) < \infty$, we find, as above, that

$$m(E_p) = m(E) + (m(E_p) - m(E_{p+1})) + (m(E_{p+1}) - m(E_{p+2})) + \ldots = m(E) + m(E_p) - \lim m(E_k),$$

which is (3.8.6).

(iii) In the general case

$$\Pi_k = \prod_k^{\infty} E_l \uparrow \underline{\lim}\, E_k, \quad \Sigma_k = \sum_k^{\infty} E_l \downarrow \overline{\lim}\, E_k,$$

by (1.8.8).† The sets Π_k and Σ_k are measurable, by Theorems 44 and 43, respectively. Hence $\underline{\lim}\, E_k$ and $\overline{\lim}\, E_k$ are measurable, by what we just have proved. Also

$$m(\Pi_k) \uparrow m(\underline{\lim}\, E_k), \quad m(\Sigma_k) \downarrow m(\overline{\lim} E_k). \quad . \quad (a)$$

The second inequality requires $m(\Sigma_p) < \infty$, for some p, which is (3.8.5).

Now $\Pi_k \subset E_l$ and $\Sigma_k \supset E_l$ for all $l \geq k$. Hence

$$m(\Pi_k) \leq \underline{\lim}\, m(E_k), \quad m(\Sigma_k) \geq \overline{\lim}\, m(E_k) ;$$

and (3.8.3) and (3.8.4) follow from (a).

If $E_k \to E$ then (3.8.6) is an obvious consequence of (3.8.3) and (3.8.4).

We observe that the restriction (3.8.5) cannot be omitted as the example, given after (2.9.1), shows.

We shall later require the following limiting property of the outer measure:

$$\bar{m}(E_k) \uparrow \bar{m}(E) \quad \text{if} \quad E_k \uparrow E. \quad . \quad . \quad (3.8.7)$$

To prove this we choose interval sets $\Sigma_k \supset E_k$ such that $|\Sigma_k| \leq \bar{m}(E_k) + \epsilon$, for given $\epsilon (> 0)$. Now, $\Sigma_l \supset E_k$ for all $l \geq k$, so that $E_k \subset \underline{\lim}\, \Sigma_k$ and thus $E \subset \underline{\lim}\, \Sigma_k$. The latter set is measurable, by Theorem 45. Also, by (3.8.3),

$$\bar{m}(E) \leq m(\underline{\lim}\, \Sigma_k) \leq \underline{\lim}\, |\Sigma_k| \leq \lim \bar{m}(E_k) + \epsilon,$$

for every $\epsilon (> 0)$. Hence $\bar{m}(E) \leq \lim \bar{m}(E_k)$. The opposite inequality is obvious.

Exercise 18. Show that for every set E there exist two measurable sets $E_1 \subset E$ and $E_2 \supset E$ such that $m(E_1) = \underline{m}(E)$ and $m(E_2) = \bar{m}(E)$. The set E_1 is a sum of closed sets and the set E_2 is a product of open sets.

† Here Σ_k does not denote an interval set.

3.9. Non-measurable sets. The results so far may leave the impression that perhaps every set is measurable. This is not the case. In fact, no definition of volume can be totally comprehensive if it is to be fully additive (as measure is).† This is shown by the following example of a non-measurable set, due to *Vitali*, which is based on the Postulates I, II*, III, and IV only. On the other hand, this and any similar known example‡ is rather sophisticated; and we may say that any "reasonable" set is measurable.

We denote by ξ any irrational number and also the number 0. The set $V(\xi)$ is the set of all numbers $\xi + r$, where ξ is fixed and r runs through all rational numbers. Each $V(\xi)$ is enumerable and everywhere dense. Also $V(\xi_1)$ and $V(\xi_2)$ are identical if, and only if, $\xi_1 - \xi_2$ is rational; otherwise $V(\xi_1)$ and $V(\xi_2)$ are exclusive sets. Finally, every number x belongs to some $V(\xi)$.

Now *select* the ξ so as to obtain all *different* sets $V(\xi)$ and *construct* in each a number $\eta = \eta(\xi)$ such that $0 \leq \eta \leq \frac{1}{2}$; for instance, the *first* such η, for a given enumeration of the rational numbers. All these η are different and form a certain set V. We shall prove that V is not measurable.

Suppose, on the contrary, that V were measurable. Let V_k, for fixed $k \geq 2$, be the set of all numbers $\eta + k^{-1}$. Then V_k is congruent to V, and hence $m(V_k) = m(V)$, by Postulate IV. Any two sets V_k and V_l, $k \neq l$, are exclusive. For, a common point would be of the form

$$\eta_1 + \frac{1}{k} = \eta_2 + \frac{1}{l} \quad \text{or} \quad \xi_1 + r_1 + \frac{1}{k} = \xi_2 + r_2 + \frac{1}{l},$$

and $\xi_1 - \xi_2$ would be rational. Similarly, V and V_k are exclusive. Hence, by Postulate II,

$$m(V + V_2 + \ldots + V_k) = km(V) \leq 1,$$

† It has also been proved that no totally comprehensive definition of volume in ordinary space is possible that satisfies the Postulates I-IV (as content does).

‡ They are all based on the axiom of choice (see footnote on p. 20). No example of different construction is known.

since V and every V_k is contained in $\langle 0, 1\rangle$. This holds for every $k \geq 2$. Hence $m(V) = 0$.

On the other hand, let $V_{(r)}$ be the set of all numbers $\eta + r$, where the rational number r is fixed and η runs through V. Then $m(V_{(r)}) = 0$, since $V_{(r)}$ is congruent to V. Also $V_{(r)}$ and $V_{(s)}$ are exclusive when $r \neq s$, by the above argument. Now every real number x belongs to some $V(\xi)$ and hence to some $V_{(r)}$. It follows that $E_1 = \Sigma V_{(r)}$, where the sum is extended over all rational numbers r in some order of enumeration. The whole linear space E_1 would, therefore, have measure zero; and we have arrived at a contradiction (to Postulate III).

Exercise 19. Show that $\underline{m}(V) = 0$.

Exercise 20. Show that the proposition

$$\underline{m}(E_1 + E_2) = \underline{m}(E_1) \quad \text{if} \quad \underline{m}(E_2) = 0$$

is false.

3.10. Vitali's covering theorem. The following theorem, which is due to *G. Vitali* (1904), is of great importance in the finer theory of measure. †

Theorem 46. *Suppose that* $\bar{m}(E) < \infty$, *and that to each point* P *of* E *corresponds a sequence of closed cubes* $Q_k(P)$ *of positive volume, containing* P *and such that* $|Q_k(P)| \to 0$ *as* $k \to \infty$. *Then there exists a finite or infinite sum* Σ *(an interval set) of such cubes, all mutually exclusive, which covers* E *"almost everywhere"; i.e.*

$$\bar{m}(E \,.\, \Sigma) = \bar{m}(E), \quad m(E \,.\, {}_c\Sigma) = 0 \quad . \quad (3.10.1)$$

Proof. It suffices to prove the second equation since by (3.4.6) it implies the first one. Also it is convenient to assume that E is a plane set, the extension to the general case being obvious. Since $\bar{m}(E) < \infty$ we can find an open

† We shall require this theorem for some of the deeper results in the theory of the indefinite integral (Chapter VI). Till then the reader may well omit reading this paragraph.

set $O > E$ such that $m(O) < \infty$. In constructing Σ we shall consider only the aggregate α of those of our squares Q for which $Q \subset O$. Note that, since $|Q_k(P)| \to 0$, *every* open set that contains a point P of E contains eventually all the $Q_k(P)$. We construct the cubes $Q^{(k)}$ of Σ by induction. We choose $Q^{(1)}$ of α arbitrarily. Suppose now that, for some $p \geq 1$,

$$S_p = \sum_1^p Q^{(k)}$$

has been defined. Let α_p be the sub-aggregate of all squares Q of α, contained in the (open) complement ${}_cS_p$. Either α_p is empty when S_p covers E and the theorem is proved; or else the areas $|Q|$ of α_p have a positive finite least upper bound A_p. We then choose a $Q^{(p+1)}$ of α_p (and hence exclusive to all preceding $Q^{(k)}$) such that

$$|Q^{(p+1)}| \geq \tfrac{1}{4} A_p. \quad . \quad . \quad . \quad . \quad \text{(a)}$$

This implies: (i) that the sides of any Q of α_p are at most double those of $Q^{(p+1)}$; and therefore: (ii) if such a Q has points in common with $Q^{(p+1)}$ it is covered by $Q_*^{(p+1)}$, a square concentric with $Q^{(p+1)}$ and having its sides five times as long as $Q^{(p+1)}$.

We may assume that the interval set Σ thus defined has an infinity of terms $Q^{(k)}$. Its measure is finite:

$$m(\Sigma) = \sum_1^\infty |Q^{(k)}| \leq m(O) < \infty. \quad . \quad . \quad . \quad \text{(b)}$$

Now, given $\epsilon > 0$, there exists an $N = N(\epsilon)$ such that

$$\sum_{N+1}^\infty |Q^{(k)}| < \epsilon. \quad . \quad . \quad . \quad . \quad . \quad \text{(c)}$$

Let $E^* = E \,.\, {}_c\Sigma$. If E^* is empty then the theorem is proved. Otherwise, let P be *any* point of E^*. *A fortiori*, $P \,\varepsilon\, {}_cS_N$ and hence there exists a $Q' = Q_k(P)$ belonging to α_N.

By (a) and (b), $A_p \to 0$ as $p \to \infty$. Hence Q' cannot belong to all α_p, and there is a *first* $p_0 > N$ such that Q' does not belong to α_{p_0}. It follows that Q', belonging to α_{p_0-1}, must have points in common with $Q^{(p_0)}$, and hence is covered by $Q_*^{(p_0)}$. Therefore

$$E^* \subset \sum_{N+1}^{\infty} Q_*^{(k)}; \quad \bar{m}(E^*) \leq 25 \sum_{N+1}^{\infty} |Q^{(k)}| < 25\epsilon,$$

by (c). Since $\epsilon > 0$ is arbitrary, the theorem is proved.

Corollary. *Let $0 < \bar{m}(E) < \infty$. Then, given ϵ, $0 < \epsilon < 1$, there exists a finite sum S (an interval sum) of our cubes, all mutually exclusive, so that*

$$|S| < (1+\epsilon)\bar{m}(E); \quad \bar{m}(E.S) > (1-\epsilon)\bar{m}(E); \quad \bar{m}(E - E.S) < \epsilon\bar{m}(E). \qquad (3.10.2)$$

Proof. We choose $O \supset E$ so that $m(O) < (1+\epsilon)\bar{m}(E)$, and construct Σ as above. By (b), every partial sum S_m of Σ satisfies the first inequality (3.10.2).

Next choose $m = m(\epsilon)$ so that

$$\sum_{m+1}^{\infty} |Q^{(k)}| < \epsilon\bar{m}(E), \qquad (d)$$

and consider $S = S_m$. Now $E.\Sigma = E.S + E.\sum_{m+1}^{\infty} Q^{(k)}$ and $E = E.S + (E - E.S)$. Since E and $E.\Sigma$ differ at most by a set of measure zero, $E - E.S$ and $E.\sum_{m+1}^{\infty} Q^{(k)}$ differ in the same way.

Hence, by (d),

$$\bar{m}(E - E.S) = \bar{m}(E.\sum_{m+1}^{\infty} Q^{(k)}) < \epsilon\bar{m}(E).$$

Finally,

$$\bar{m}(E) \leq \bar{m}(E.S) + \bar{m}(E - E.S) < \bar{m}(E.S) + \epsilon\bar{m}(E),$$

$$\bar{m}(E.S) > (1-\epsilon)\bar{m}(E).$$

Solutions to Exercises

Ex. 16. See solution to Ex. 10, (i).

Ex. 17. Let $E = E_1 + E_2$ and $I \supset E$. Then

$$(I - E_1) + (E_1 \, . \, E_2) = (I - E) + E_2.$$

Hence, by (3.4.5) and (3.4.6),

$$\bar{m}(I - E_1) = \bar{m}[(I - E_1) + E_1 \, . \, E_2] \leq \bar{m}(I - E) + \bar{m}(E_2). \quad \text{(a)}$$

Also, by (3.5.1),

$$\underline{m}(E) - \underline{m}(E_1) = \bar{m}(I - E_1) - \bar{m}(I - E).$$

This and (a) prove the left-hand part of (3.5.13). The proof of the other part is similar.

Ex. 18. By (3.5.10), we can find a sequence of closed sets $F_k \subset E$ such that $m(F_k) \to \underline{m}(E)$. Then $E_1 = \Sigma F_k \subset E$ and $m(E_1) = \underline{m}(E)$. The construction of the set E_2 is similarly based on (3.4.7).

Ex. 19. Using the notations of § 3.9, we have $\underline{m}(V_k) = \underline{m}(V)$ since V_k and V are congruent. Hence, by (3.5.3),

$$k\underline{m}(V) \leq \underline{m}(V + V_2 + \ldots + V_k) \leq 1$$

for every $k \geq 2$, since the V_k and V are mutually exclusive and their sum is contained in $\langle 0, 1 \rangle$. This gives $\underline{m}(V) = 0$ on letting $k \to \infty$.

Ex. 20. If V is Vitali's set (§ 3.9), then V is contained in $I = \langle 0, \frac{1}{2} \rangle$. Also $\underline{m}(V) = 0$ and $\bar{m}(V) > 0$. It follows that

$$\underline{m}(I - V) < \tfrac{1}{2},$$

so that

$$\underline{m}((I - V) + V) = |\, I \,| = \tfrac{1}{2} > \underline{m}(I - V) + \underline{m}(V).$$

PART II

THE INTEGRAL

CHAPTER IV

RIEMANN'S INTEGRAL

4.1. Volume and integral. In ancient Greek mathematics the method which we now call integration appears as a geometrical method for the exhaustion of elementary volumes (or areas) (*Archimedes*, third century B.C.) With the development of co-ordinate geometry in the seventeenth century integration tends to take the form of a " summing up " process for functions defining elementary geometrical configurations. As the power of the new analytical methods increases, the original geometrical idea becomes absorbed, and is almost hidden, in the purely analytical definition of the *definite integral*. The reader will be familiar, for instance, with the now classical definition of the integral as given by *B. Riemann* (1854). The determination of volume thus becomes merely one of the principal applications of the integral calculus. Also, at this stage, volume is accepted as a primitive geometrical concept.

Towards the end of the last century the problem of volume was thoroughly investigated by *Peano*, *Jordan*, *Borel*, *Lebesgue*, and others. We have given an account of their results in the first part of this book. It seems to-day more natural, and also more consistent with the historical background, to approach the theory of integration once more from the original geometrical point of view. Volume then no longer appears as a mere application of the integral calculus ; the integral itself is now the volume of a certain set derived from the function to be integrated. The classical analytical formulae which served for the definition of the integral take again their natural places in this geometrical theory. Their general importance for the calculus remains, of course, quite unimpaired.

Let E be a set in a given space $\mathsf{E} = \mathsf{E}_n$ of n dimensions; and let, first,

$$y = f(P) = f(x_1, x_2, \ldots x_n) \quad . \quad . \quad (4.1.1)$$

be a non-negative function (of n real variables x_i) defined in E. The value $y = \infty$ for $f(P)$ is admitted.†

We consider, in the $(n+1)$-dimensional space $\mathsf{E}^* = \mathsf{E}^*_{n+1}$, the set Y that consists of all points

$$P^* = (x_1, x_2, \ldots x_n, x_{n+1}) = (P\,;\; x_{n+1}), \quad . \quad (4.1.2)$$

where $P \,\varepsilon\, E$ and $0 \leq x_{n+1} \leq y (= f(P))$. This set $Y = Y(f)$ is called the *ordinate set* of f over E. It consists of all the linear "ordinate intervals" $\langle P, 0 \leq x_{n+1} \leq y \rangle$ ‡ in E^*, erected "over" each point P of E (in the subspace E of E^*). Y is bounded "below" by the given set E, and "above" by the graph of $f(P)$ (§ 2.8).

Now suppose that Y has a volume $V(Y)$, in E^*, defined in some way. We shall then say that the function $f(P)$ has a (*definite*) integral over E, this integral being defined (and written) as

$$\int_E f(P)dP = \int_E f dP = V(Y) \quad . \quad . \quad (4.1.3)$$

The integral may be infinite. Also its existence depends, of course, on the underlying notion of volume.

In the case $n = 1$ it is usual to write the integral in the form

$$\int_E f(x)dx \quad \text{or, more specially,} \quad \int_a^b f(x)dx, \quad . \quad (4.1.4)$$

if $E = \langle a, b \rangle$. It is the area, supposed to exist, of the plane ordinate set of the non-negative function $y = f(x)$, defined in the linear set E. Similarly, in the cases $n = 2$ and $n = 3$, the usual notations are

† The "numbers" ∞ and $-\infty$ are used in the way, "obvious" in real arithmetic.

‡ If $y = \infty$, this is a half-line.

$$\iint\limits_E f(x, y)dxdy, \quad \iiint\limits_E f(x, y, z)dxdydz. \quad . \quad (4.1.5)$$

We shall, of course, keep these notations when dealing with these cases. For a general n, the short notation (4.1.3) is preferable to that of an n-times repeated symbol of integration.

The case of an infinite value of the integral is mainly of "formal" interest. We reserve the word *integrable* (over E) for a *function whose integral is finite.* The reader should keep this well in mind. So a non-negative function may have an (infinite) integral and yet may not be integrable.

So far we have restricted ourselves to non-negative functions. Now consider a general function $y=f(P)$ defined in E. We write

$$f_+(P) = \max\,(0, f(P)), \quad f_-(P) = \min\,(0, f(P))\,; \quad . \quad (4.1.6)$$

that is, $f_+ = f$ when $f \geq 0$, and $f_+ = 0$ otherwise; and $f_- = f$ when $f \leq 0$, and $f_- = 0$ otherwise. Both f_+ and $-f_-$ are non-negative in E, and

$$f(P) = f_+(P) + f_-(P), \quad |f(P)| = f_+(P) - f_-(P). \quad (4.1.7)$$

We shall say that a general f is *integrable* over E, if both the ordinate sets of f_+ and $-f_-$ have *finite* volumes; that is, if both f_+ and $-f_-$ are integrable. The *integral* of f over E is then defined as

$$\int\limits_E f dP = \int\limits_E f_+ dP - \int\limits_E (-f_-) dP. \quad . \quad . \quad (4.1.8)$$

The integral thus measures the volume of the ordinate set of a general function $f(P)$ algebraically: the part "above" E is counted positive, the part "below" negative.

It is important to note that our definition of the integral implies that $|$ f(P) $|$ *is integrable whenever* f(P) *is integrable*. It is a so-called *absolute* integral. In fact, if f is integrable, then the ordinate set of $|f|$ has a finite volume as the sum of the volumes of the ordinate sets of f_+ and $-f_-$†; that is,

$$\int_E |f| dP = \int_E f_+ dP + \int_E (-f_-) dP. \quad . \quad . \quad (4.1.9)$$

It follows from (4.1.8) and (4.1.9) that

$$\left| \int_E f dP \right| \leq \int_E |f| \, dP, \quad . \quad . \quad . \quad (4.1.10)$$

whenever f is integrable.

Consider the four integrals

$$\int_0^1 \frac{dx}{\sqrt{x}}, \quad \int_0^1 \frac{dx}{x}, \quad \int_0^T \frac{\sin x}{x} dx, \quad \int_0^\infty \frac{\sin x}{x} dx. \quad . \quad (4.1.11)$$

With any of our later definitions the first three integrals will exist, though the second is infinite: the function $1/x$ is not integrable over $\langle 0, 1 \rangle$. The last integral does not exist since the integral of $|\sin x/x|$ over $\langle 0, \infty \rangle$ is infinite. This integral is a so-called *improper* integral, or *Cauchy integral.* It exists in the sense that

$$\int_0^\infty \frac{\sin x}{x} dx = \lim_{T \to \infty} \int_0^T \frac{\sin x}{x} dx. \quad . \quad . \quad (4.1.12)$$

Such integrals are not covered by the theory of the absolute integral as discussed in this book.

The notion of the integral of a non-negative function is equivalent to the notion of volume of the ordinate set of this function. The more effective the definition of volume is, the more effective will be the definition of the integral based on it. In this chapter we shall discuss the theory of the integral based on the Peano-Jordan content (Chapter II). This definition will prove to be equivalent to the classical definition of the integral as given by *B. Riemann* (1854).‡

† By Postulate II, the part where the ordinate sets of f_+ and $-f_-$ overlap, that is, where $f_+ = f_- = 0$, is a subset of **E**; and hence has volume zero (in **E***), by Postulate III.

‡ Riemann considers bounded functions in bounded sets only. This restriction, however, is not essential for our "geometrical" approach.

It is true that the Riemann integral has been widely superseded, in modern analysis, by the more powerful integral of *Lebesgue* (which is based on measure and which we shall discuss in the next chapter). Nevertheless, its importance, both for less advanced analysis and for application to Science, remains undiminished.

4.2. Lower and upper integrals. Let $y=f(P)$ be a non-negative function defined in E, and let Y be its ordinate set. We write

$$\underline{\int_E} f dP = \underline{c}(Y), \quad \overline{\int_E} f dP = \bar{c}(Y) \qquad . \quad . \quad (4.2.1)$$

for the inner and outer contents of Y (in E^*), and call these contents the *lower and upper Riemann integrals* (R-integrals), respectively. These integrals were first considered by *G. Darboux* (1875). They are non-negative and always exist, though they may be infinite. Clearly,

$$\underline{\int_E} f dP \leq \overline{\int_E} f dP. \qquad . \quad . \quad . \quad (4.2.2)$$

Also, if $0 \leq f \leq g$ in E, then

$$\underline{\int_E} f dP \leq \underline{\int_E} g dP, \quad \overline{\int_E} f dP \leq \overline{\int_E} g dP. \quad . \quad . \quad (4.2.3)$$

For, the ordinate set of f is then part of that of g.

Similarly, the integrals of f over a subset E_1 of E are at most equal to the corresponding integrals over E.

If $a \geq 0$, let

$$f_{[a]} = \min(f, a), \qquad . \quad . \quad . \quad (4.2.4)$$

so that $f_{[a]} = f$ when $f \leq a$, and $f_{[a]} = a$ otherwise. Then, as $a \uparrow \infty$,

$$\underline{\int_{E_{[a]}}} f_{[a]} dP \uparrow \underline{\int_E} f dP, \quad \overline{\int_{E_{[a]}}} f_{[a]} dP \uparrow \overline{\int_E} f dP \qquad . \quad (4.2.5)$$

where $E_{[a]}$ is the product of E by the $(n+1)$-dimensional cube $Q_a = \langle\, |x_i| \leq a \rangle$. This follows from (2.5.3) and (2.6.10) since the ordinate set of $f_{[a]}$ over $E_{[a]}$ is the product of Y by Q_a.

The formulae (4.2.5) are important for extending results, first obtained for bounded functions and bounded sets, to the general case.

Next, if $\alpha \geq 0$, then

$$\underline{\int_E} \alpha f dP = \alpha \underline{\int_E} f dP, \quad \overline{\int_E} \alpha f dP = \alpha \overline{\int_E} f dP. \qquad (4.2.6)$$

For, let first E and f be bounded. Then any inner, or outer, interval sum S of Y becomes a corresponding interval sum S^* for the ordinate set of αf, on multiplying, in an obvious way, the x_{n+1}-side of each constituent interval by α; and any S^* can thus be obtained. Clearly, $|S^*| = \alpha|S|$, which proves the formulae. In the general case, the results follow easily on using (4.2.5).

The set E lies in the n-dimensional subspace E of E^*. We denote the inner and outer contents of E, in E, by $\underline{\gamma}(E)$ and $\bar{\gamma}(E)$, respectively.

THEOREM 47.

$$\underline{\int_E} dP = \underline{\gamma}(E), \quad \overline{\int_E} dP = \bar{\gamma}(E). \qquad (4.2.7)$$

PROOF. Here $f \equiv 1$ in E. To prove the first formula, take any interval sum $S = \sum_1^m I_k$ contained in E. Let I_k^* be the interval $\langle I_k;\ 0 \leq x_{n+1} \leq 1 \rangle$, of "height 1" over I_k, in E^*. Clearly, $S^* = \sum_1^m I_k^* \subset Y$; and any $S_1^* \subset Y$ is contained in some such S^*. Also $|S^*|$, in E^*, equals $|S|$, in E.

The proof for the second formula is similar; it can be assumed that E is bounded.

We prove as an application:

If the upper R-integral of f *is finite then* $\gamma(E_\infty) = 0$, *where* E_∞ *is the subset of* E *in which* $\mathrm{f} = \infty$.

For, $f \geq M$ in E_∞, for every $M(>0)$. Hence

$$M\bar{\gamma}(E_\infty) = \bar{\int\limits_{E_\infty}} M dP \leq \bar{\int\limits_{E_\infty}} f dP \leq \bar{\int\limits_{E}} f dP < \infty,$$

by (4.2.3), (4.2.6), and (4.2.7). This gives $\bar{\gamma}(E_\infty) = 0$ on letting $M \to \infty$.

By (2.4.8) and Exercise 10, we may omit from Y any subset of content zero, in E^*, without altering $\underline{c}(Y)$ and $\bar{c}(Y)$. In particular, the subspace E, and any set in it, has content zero in E^*. We may, therefore, change the definition of the ordinate set Y, by taking only the half-open intervals $(P\,;\ 0 < x_{n+1} \leq y)$ over E. Or, we may add to E, or omit from E, any set in E where $f = 0$, without changing the lower or upper R-integrals. Let

$$f^*(P) = \begin{cases} f(P) & \text{when } P \,\varepsilon\, E \\ 0 & \text{otherwise} \end{cases} \qquad . \quad . \quad (4.2.8)$$

This function is defined in the whole space E. We call it the *extension* of $f(P)$, or the *extended function* ; and we have

$$\underline{\int\limits_{E_1}} f^* dP = \underline{\int\limits_{E}} f dP, \quad \bar{\int\limits_{E_1}} f^* dP = \bar{\int\limits_{E}} f dP, \quad . \quad . \quad (4.2.9)$$

whenever $E_1 \supset E$. The set E_1 may be the whole space E.

4.3. Riemann's integral. From now on we use c and γ to denote contents in E^* and E, respectively.

Given, in the first place, is a non-negative function $f(P)$ in E. If its ordinate set Y has a content, then we write

$$\int\limits_{E} f dP = c(Y), \quad . \quad . \quad . \quad . \quad (4.3.1)$$

and call this content the *Riemann integral* (R-integral) of f over E. This integral is non-negative and may be infinite.

The existence of the R-integral implies that the lower and upper integrals (4.2.1) should be equal. If the integral is finite, this is also sufficient. If, however, $c(Y)=\infty$ then all sets $Y\,.\,I$ must have finite contents if the R-integral is to exist (§ 2.7); in particular, $f_{[a]}$ must have an integral over $E_{[a]}$ (see (4.2.5)). In any case, we shall have

$$\int_{E[a]} f_{[a]}dP \uparrow \int_{E} fdP \quad . \quad . \quad . \quad (4.3.2)$$

as $a \uparrow \infty$.

Whatever the set E be,

$$\int_{E} 0dP = 0. \quad . \quad . \quad . \quad . \quad (4.3.3)$$

For, $Y=E$ in this case, so that $c(Y)=0$.

A non-negative function $f(P)$ is said to be R-*integrable* over E, if its integral is *finite*.

The definition extends to a general function:

A general function $f(P)$ is R-*integrable* over E if the two non-negative components f_+ and $-f_-$, of (4.1.6), are R-integrable, and then *the* R-*integral of* f *is defined by* (4.1.8).

The function f may have values ∞ and $-\infty$. However, if f is integrable, then the set where this happens has content zero, in E. This follows from what we proved after Theorem 47.

We define the extension $f^*(P)$ of a general function $f(P)$ as in (4.2.8). Then

$$\int_{E_1} f^*dP = \int_{E} fdP, \quad . \quad . \quad . \quad (4.3.4)$$

whenever $E_1 \supset E$. For, this is true for both f_+ and $-f_-$, by (4.2.9). The formula (4.3.4) is important in that it often permits replacement of E by a bigger set of simpler type (for instance, by the whole space E).

Next, if $f \leq g$ and both functions have R-integrals over E, then

$$\int_{E} fdP \leq \int_{E} gdP. \, . \quad . \quad . \quad . \quad (4.3.5)$$

For, $f_+ \leq g_+$ and $(-f_-) \geq (-g_-)$; and we may apply (4.2.3) to the integrals in (4.1.8).

THEOREM 48. *If* f *is* R-*integrable over* E, *then so is* αf ; *and* †

$$\int_E \alpha f dP = \alpha \int_E f dP. \quad . \quad . \quad . \quad (4.3.6)$$

PROOF. If $\alpha \geq 0$, then $(\alpha f)_+ = \alpha f_+$ and $(\alpha f)_- = \alpha f_-$. If $\alpha \leq 0$, then $(\alpha f)_+ = -|\alpha| f_-$ and $(\alpha f)_- = -|\alpha| f_+$. In both cases the theorem clearly follows from (4.2.6) and (4.1.8).

THEOREM 49. *If* $\gamma(\mathrm{E}) = 0$, *then any function* f(P) *is* R-*integrable over* E ; *and its integral is zero.*‡

PROOF. We may assume that $f \geq 0$; and, because of (4.3.2), we may assume that $f \leq a$, say. Then

$$0 \leq \bar{\int_E} f dP \leq \bar{\int_E} a dP = a\gamma(E) = 0,$$

by (4.2.3) and (4.2.7). Hence the upper integral and, therefore, the integral of f is zero.

The *characteristic function* $\chi(P)$ of a set E is defined as $\chi(P) \equiv 1$ in E, and $\chi(P) = 0$ otherwise. Thus $\chi(P)$ is the extension of $f = 1$ from E into E.

THEOREM 50. *The characteristic function of a set* E *has an* R-*integral over* E (*or* E) *if, and only if,* E *has content ; and then*

$$\int_{\mathrm{E}} \chi dP = \int_E dP = \gamma(E). \quad . \quad . \quad . \quad (4.3.7)$$

The proof follows from (4.2.7) and (4.2.9).

We know that the rational set R has no content. Hence its characteristic function gives an example of a function which has no R-integral.

More generally, $\int_E \alpha dP = \alpha\gamma(E)$ when E has content (finite content, if $\alpha < 0$).

We also note that the function $f \equiv \infty$ has an R-integral

† If $f \geq 0$ and $\alpha > 0$, then it suffices that f has an integral (which may be infinite). If $\alpha = 0$, then αf is not defined where $|f| = \infty$. This, however, is irrelevant (see Theorem 54).

‡ Note that $|f|$ may be infinite throughout E.

over E if E has content. This follows from Theorem 50 and (4.3.2). However, $f \equiv \infty$ will not be integrable unless $\gamma(E)=0$.

THEOREM 51. *Suppose that* f *has an* R-*integral over* E. *Then it has also an* R-*integral over any subset* E_1, *provided that* E_1 *has content.*

PROOF. We may assume that $f \geq 0$. If Y_1 is the ordinate set of f over E_1, then $Y_1 = Y.Y_1^*$ where Y_1^* is the ordinate set of $g \equiv \infty$ over E_1. Both $c(Y)$ and $c(Y_1^*)$ exist. Hence Y_1 has content, by Theorem 22.

THEOREM 52. *Suppose that* $E = E_1 + E_2 + \ldots + E_m$ *where* $\gamma(E_k \,.\, E_l) = 0$ *for any* $k \neq l$. *If* f *is* R-*integrable over each* E_k, *then it is also* R-*integrable over* E ; *and*†

$$\int_E f dP = \int_{E_1} f dP + \int_{E_2} f dP + \ldots + \int_{E_m} f dP. \qquad (4.3.8)$$

PROOF. We may assume that $f \geq 0$. Then

$$Y = Y_1 + Y_2 + \ldots + Y_m$$

where Y_k is the ordinate set of f over E_k. Now, $Y_k \,.\, Y_l$ is the ordinate set of f over $E_k \,.\, E_l$; if $k \neq l$, it has content zero, by Theorem 49, since $\gamma(E_k \,.\, E_l) = 0$. Also all the Y_k have contents. Hence Y has content, and (4.3.8) holds, by Theorem 22.

The proof is similar for the following theorem.

THEOREM 53. *Suppose that* $E_1 \subset E$, *and that* f *is* R-*integrable over* E_1, *and has an* R-*integral over* E. *Then* f *has also an* R-*integral over* $E - E_1$; *and*

$$\int_{E-E_1} f dP = \int_E f dP - \int_{E_1} f dP. \qquad (4.3.9)$$

Note that the integral over E_1 is finite, by assumption.

The following theorem asserts that two functions which differ only in a set of content zero must be considered as equivalent in the theory of the R-integral.

† If $f \geq 0$, then it suffices for (4.3.8) that all the integrals over the E_k exist.

THEOREM 54. *Suppose that* f *has an* R*-integral over* E, *and that* f = g *except perhaps in a subset* E_1 *of content zero (in* E*). Then* g *has an* R*-integral over* E, *equal to that of* f.

PROOF. The integrals of f, or g, over E_1 are zero, by Theorem 49. Hence

$$\int_E f dP = \int_{E-E_1} f dP = \int_{E-E_1} g dP = \int_E g dP,$$

by (4.3.9) (for f) and (4.3.8) (for g).

Our next theorem is important for the applications.

THEOREM 55. *Suppose that* E *is bounded, closed, and has content. If* f *is continuous in* E, *then* f *is* R*-integrable over* E.

PROOF. First, f is bounded, by Theorem 14. We may also assume that $f \geq 0$. For, clearly, if f is continuous in E, then so are f_+ and $-f_-$.

The ordinate set Y of f is bounded. Since E is closed, the frontier of Y consists of (i) the set E itself, at the bottom; (ii) the graph G of f, on top; and (iii) the ordinate set Y_1 of f over the frontier E_f of E. Now, $c(E) = 0$; also $c(G) = 0$, by Theorem 27; and $c(Y_1) = 0$, by Theorem 49, since $\gamma(E_f) = 0$ by Theorem 26. It follows that the frontier of Y has content zero, so that $c(Y)$ exists, again by Theorem 26.

Exercise 21. For a general f the *lower and upper integrals* are defined as

$$\underline{\int}_E f dP = \underline{\int}_E f_+ dP - \overline{\int}_E (-f_-) dP\;;\quad \overline{\int}_E f dP = \overline{\int}_E f_+ dP - \underline{\int}_E (-f_-) dP, \quad . \qquad (4.3.10)$$

assuming that all the integrals are finite.

Prove that the lower integral is not greater than the upper integral; and that f is integrable if, and only if, the lower integral equals the upper one.

Exercise 22. Let R_I be the set of all rational numbers r_k in $I = \langle 0, 1 \rangle$ in some order of enumeration. If $f(r_k) = 1/k$, show that the integral of f over R_I is zero.

This example shows that a positive function may be integrable over a set without content.

4.4. Riemann's sums. The reader will be familiar with the main properties of Riemann's integral, as they are proved in elementary accounts of the integral calculus.† Of these we have not yet proved, for instance, that $f+g$ is integrable when f and g are; and that the integral of

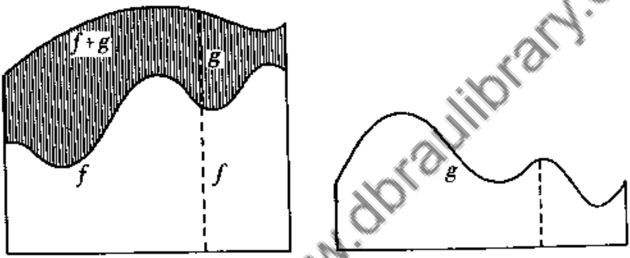

FIG. 8

$f+g$ equals the sum of the integrals of f and g. This is, from our point of approach, not really an elementary proposition.

Take e.g. two non-negative functions $f(x)$ and $g(x)$, defined in the interval $\langle a, b\rangle$; and draw the ordinate sets of f and $f+g$ (Fig. 8). The latter consists of the ordinate set of f and the shaded part on top of it. This part is not an ordinate set, though it appears "plausible" that it should have the same content as the ordinate set of g. However, the proof of this is not obvious.‡

To overcome this difficulty *Riemann* approached the concept of the integral in a purely analytical way. We shall now attach his method to our geometrical theory.

It is essential for Riemann's definition that *both the set* E *and the function* f *in it should be bounded*. We shall,

† Cf. Gillespie, *Integration*, chap. v.
‡ In § 5.7 we shall give a direct "geometrical" proof for the more general L-integral.

therefore, assume the same in this paragraph. However, the main result, that is Theorem 59, remains valid without this restriction. We shall state it, therefore, in the general form. It would be possible, though somewhat troublesome, to prove it thus, making use of the approximations (4.2.5). We prefer, however, to leave the general case unproved for the present; it will follow easily from the theory of the L-integral, in the next chapter.

We suppose then that $f(P)$ is a bounded function defined in a bounded set E. Without loss of generality we may assume that E is a closed interval I. For, we may extend f from E into such an interval, according to (4.3.4).

Now consider a "division"

$$(D) \qquad I = J_1 + J_2 + \ldots + J_m$$

of I into mutually separate closed intervals J_k; and let

$$m_k = \inf_{P \varepsilon J_k} f(P), \quad M_k = \sup_{P \varepsilon J_k} f(P), \qquad (4.4.1)$$

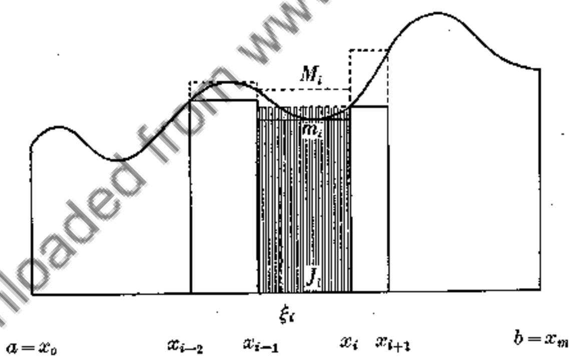

FIG. 9. Riemann division

where the bounds are taken with respect to all points P of a fixed J_k. Let also (Fig. 9)

$$s(D) = \sum_1^m m_k |J_k|, \quad S(D) = \sum_1^m M_k |J_k|. \qquad (4.4.2)$$

G

Clearly, $s(D) \leq S(D)$. On splitting some, or all, of the J_k into mutually separate sub-intervals,

$$J_k = J_k^{(1)} + J_k^{(2)} + \ldots + J_k^{(s_k)},$$

we obtain a "subdivision" D' of D. Now, with obvious notations, $m_k \leq m_k^{(i)} \leq M_k^{(i)} \leq M_k$ for all $1 \leq i \leq s_k$, so that

$$m_k |J_k| \leq \sum_1^{s_k} m_k^{(i)} |J_k^{(i)}| \leq \sum_1^{s_k} M_k^{(i)} |J_k^{(i)}| \leq M_k |J_k|.$$

It follows that, for every subdivision D' of D,

$$s(D) \leq s(D') \leq S(D') \leq S(D). \quad . \quad . \quad (4.4.3)$$

Next, if $f(P) \geq 0$, then

$$0 \leq s(D) \leq \underline{\int_I} f dP \leq \overline{\int_I} f dP \leq S(D). \quad . \quad . \quad (4.4.4)$$

For, it is plain that $s(D)$ is the volume of an interval sum contained in Y, the intervals of which have "height" m_k over J_k; and, similarly, $S(D)$ is the volume of an interval sum containing Y.

Also, again if $f \geq 0$,

$$\underline{\int_I} f dP = \sup s(D), \quad \overline{\int_I} f dP = \inf S(D), \quad . \quad (4.4.5)$$

the bounds being taken with respect to all divisions D of I.

For, let S be an interval sum (in E^*) contained in Y. By increasing S, if necessary, we may assume that

$$S = J_1^* + J_2^* + \ldots + J_m^*$$

where the J_k^* are mutually separate intervals, reaching down to I. Then the $J_k = J_k^* . I$ are mutually separate sub-intervals of I. If D is a division of I comprising these J_k, then, clearly, $|S| \leq s(D)$. It follows that $\underline{v}(Y) \leq \sup s(D)$; the opposite is true by (4.4.4). This proves the first formula (4.4.5). The proof for the second formula is similar.

If f is a general function, then $f = f_+ - (-f_-)$; and

$$m_k = m_k^+ - M_k^-, \quad M_k = M_k^+ - m_k^-,$$

where the upper signs + − refer to f_+ and $(-f_-)$, respectively.† Hence

$$s(D)=s^+(D)-S^-(D), \quad S(D)=S^+(D)-s^-(D), \quad . \quad (4.4.6)$$

where the sums on the right are taken for f_+ and $-f_-$, respectively.

Now let

$$\Delta(D)=S(D)-s(D)=\sum_1^m \sigma_k |J_k|, \quad . \quad . \quad (4.4.7)$$

where $\sigma_k = M_k - m_k$ is the *oscillation* of f in I. The following theorem is known as *Riemann's Criterion.*

THEOREM 56. *Suppose that* f(P) *is bounded in* I. *In order that* f(P) *be* R-*integrable over* I, *it is necessary and sufficient that there exists, for every positive* ϵ, *a division* D *of* I *such that* $\Delta(\mathrm{D}) \le \epsilon$.

PROOF. Suppose, first, that $f \ge 0$. If $\Delta(D) \le \epsilon$, for every $\epsilon(>0)$ and suitable D, then f is integrable, by (4.4.4): the lower and upper integrals are equal.

Conversely, if f is integrable, then, by (4.4.5), we can find divisions D_1 and D_2 such that

$$s(D_1) \ge \int_I f dP - \frac{\epsilon}{2}, \quad S(D_2) \le \int_I f dP + \frac{\epsilon}{2}.$$

If D is a common subdivision of D_1 and D_2, then these inequalities will also hold for $s(D)$ and $S(D)$, respectively, by (4.4.3). Hence $\Delta(D) \le \epsilon$, and the theorem is proved.

For a general f, with obvious notations,

$$\Delta(D)=\Delta^+(D)+\Delta^-(D), \quad . \quad . \quad (4.4.8)$$

by (4.4.6). Hence $\Delta(D) \le \epsilon$ implies both $\Delta^+(D) \le \epsilon$ and $\Delta^-(D) \le \epsilon$: both f_+ and $-f_-$ are integrable; and so is f. Conversely, if f is integrable, then both f_+ and $-f_-$ are integrable. There exists a D ‡ such that both $\Delta^+(D) \le \frac{1}{2}\epsilon$ and $\Delta^-(D) \le \frac{1}{2}\epsilon$; and then $\Delta(D) \le \epsilon$.

† One of the terms on the right always vanishes.

‡ A common subdivision of such divisions for f_+ and $-f_-$.

We add two applications:

(i) *If* f(P) *is continuous in* I, *then it is* R-*integrable over* I *(a special case of Theorem* 55).

For, f is bounded, by Theorem 15; and it is uniformly continuous in I, by Theorem 16. This clearly implies $\sigma_k \leq \epsilon$ for all k, whenever the division D is " fine " enough; that is, whenever the sides of all the constituent intervals J_k are small enough. We then have $\Delta(D) \leq \epsilon |I|$.

(ii) *If* f(x) *is monotone and bounded in* ⟨a, b⟩, *then it is* R-*integrable over* ⟨a, b⟩.

We may assume that $f(x)$ increases in $I = \langle a, b \rangle$. Now divide I into m equal parts by the points $\xi_k = a + k(b-a)/m$, $0 \leq k \leq m$. Each part $J_k = \langle \xi_{k-1}, \xi_k \rangle$, $1 \leq k \leq m$, has then the length $(b-a)/m$, and $\sigma_k = f(\xi_k) - f(\xi_{k-1})$ is the oscillation of f in it. Hence

$$\Delta(D) = \frac{b-a}{m} \sum_1^m (f(\xi_k) - f(\xi_{k-1})) = \frac{b-a}{m}(f(b) - f(a)) \leq \epsilon,$$

if m is large enough.

The m intervals J_k of a division D determine a finite number of edges. We denote the maximal length of these edges by $L(D)$. The following fundamental theorem is due to *Darboux*.

THEOREM 57. *Suppose that* f(P) *is bounded in* I. *Then* f(P) *is* R-*integrable over* I *if, and only if,* $\Delta(D) \leq \epsilon$ *for every given* $\epsilon (>0)$, *whenever* $L(D) \leq \Lambda = \Lambda(\epsilon)$; *that is to say, if, and only if,*

$$\Delta(D) \to 0, \qquad (4.4.9)$$

for every sequence of divisions D *such that* $L(D) \to 0$.

PROOF. It is convenient to assume that $n = 2$, so that I is a plane interval. The extension to a general n will be obvious: if $n = 1$, the proof would be slightly simpler.

Because of Theorem 56 we need prove the necessity of the conditions only. Let $z = f(x, y)$ be defined in the interval $I = \langle a \leq x \leq b, \alpha \leq y \leq \beta \rangle$, and let $|f| \leq M$ there.

Suppose, then, that f is integrable. Given $\epsilon(>0)$, there exists, by Theorem 56, a division D_0 of I into m_0 intervals $J_r^{(0)}$, such that

$$\Delta(D_0)=\sum_1^{m_0}\sigma_r^{(0)}|\,J_r^{(0)}\,|\le\frac{\epsilon}{2},$$

where $\sigma_r^{(0)}$ is the oscillation of f in $J_r^{(0)}$. Let

$$J_r^{(0)}=\langle\xi_r\le x\le\xi_r',\quad \eta_r\le y\le\eta_r'\rangle,$$

where $a\le\xi_r\le\xi_r'\le b,\quad \alpha\le\eta_r\le\eta_r'\le\beta$, and $1\le r\le m_0$.

Now consider also an *arbitrary* division D of I into m intervals J_k, and the corresponding sum

$$\Delta(D)=\sum_1^{m}\sigma_k|\,J_k\,|.$$

If a J_k is contained in some $J_r^{(0)}$, then, clearly, $\sigma_k\le\sigma_r^{(0)}$; and the contribution of all such J_k to $\Delta(D)$ is not greater than $\Delta(D_0)$.

Any of the remaining intervals J_k must have points in common with at least one of the lines $x=\xi_r$, $x=\xi_r'$, $y=\eta_r$, or $y=\eta_r'$. Consider, for instance, one such line $x=\xi_r$, and suppose that $\xi_r>a$. All intervals J_k which have points in common with this line (which may either cross the interior of J_k or border it from one side) contribute to $\Delta(D)$ not more than $4M(\beta-\alpha)L(D)$. For, clearly, $\sigma_k\le 2M$, and the height of I is $\beta-\alpha$ (the total length of the sides parallel to the y-axis is at most $2(\beta-\alpha)$; there may be two such intervals separated by our line $x=\xi_r$). If $\xi_r=a$, the contribution is at most $2M(\beta-\alpha)L(D)$. Hence the contribution of all J_k which have points in common with some of the lines $x=\xi_r$ is at most $4m_0M(\beta-\alpha)L(D)$.

A similar argument applies to the other lines mentioned, and we find that

$$\Delta(D)\le\Delta(D_0)+8m_0M[(\beta-\alpha)+(b-a)]L(D).$$

Hence, if

$$L(D)\le\frac{\epsilon}{16m_0M[(\beta-\alpha)+(b-a)]}=\Lambda,$$

a number which depends on ϵ only, we shall have

$$\Delta(D) \leq \Delta(D_0) + \frac{\epsilon}{2} \leq \epsilon,$$

as desired.

Our next theorem contains Riemann's definition of the integral by means of the sums (4.4.10). These sums are known as *Riemann sums* (Fig. 9).

THEOREM 58. *Suppose that* f(P) *is bounded in* I. *Then* f(P) *is* R-*integrable over* I *if, and only if, the sums*

$$\mathsf{S}(D) = \sum_1^m f(P_k)|J_k| \quad . \quad . \quad . \quad (4.4.10)$$

have a limit whenever $L(D) \to 0$, *the* P_k *being arbitrary points in* J_k *and the limit being independent of their choice. In fact,*

$$\mathsf{S}(D) \to \int_I f(P)dP. \quad . \quad . \quad . \quad (4.4.11)$$

PROOF. Suppose, first, that the sums $\mathsf{S}(D)$ have a limit, independent of the P_k, whenever $L(D) \to 0$. Then the sums $s(D)$ and $S(D)$ must have the same limit, since the P_k can be chosen so that $f(P_k)$ is arbitrarily close to m_k, or M_k. Hence $\Delta(D) \to 0$, and f is integrable, by Theorem 57.

Conversely, if f is integrable, then $\Delta(D) \to 0$ when $L(D) \to 0$, again by Theorem 57. Suppose that f is non-negative. Then $s(D)$ and $S(D)$ tend to the integral of f, by (4.4.4). This is also true for $\mathsf{S}(D)$, since $s(D) \leq \mathsf{S}(D) \leq S(D)$.

For a general f, we apply the theorem to f_+ and $-f_-$.

Exercise 23. Prove the formulae (4.4.5) for a general bounded f, using the definition (4.3.10). We state, in this connection, that

$$s(D) \to \underline{\int_I} f(P)dP, \quad S(D) \to \overline{\int_I} f(P)dP, \quad . \quad (4.4.12)$$

as $L(D) \to 0$.

The Riemann sums find their main application in the proof of the following result.

THEOREM 59. *Suppose that* f(P) *and* g(P) *are* R-*integrable over* E. *Then* (i) f + g *and* f − g *are* R-*integrable over* E, *and*

$$\int_E (f \pm g)dP = \int_E f dP \pm \int_E g dP. \quad . \quad . \quad (4.4.13)$$

(ii) fg *is* R-*integrable over* E, *provided that* g *is bounded.*

(iii) f/g *is* R-*integrable over* E, *provided that* $|g| \geq$ d *for some positive number* d.

PROOF. We prove the theorem here only in the case when E is a bounded set, and f and g are bounded. For the general case see § 5.8. The set E can be taken as an interval I.

(i) We have

$$\sum_1^m (f(P_k) \pm g(P_k))|J_k| = \sum_1^m f(P_k)|J_k| \pm \sum_1^m g(P_k)|J_k|;$$

and the desired result follows from Theorem 58 when $L(D) \to 0$.

(ii) Let $|f| \leq M$, $|g| \leq M$. Then

$$|f(P)g(P) - f(Q)g(Q)| \leq |f(P)|\,|g(P) - g(Q)| + |g(Q)|\,|f(P) - f(Q)|.$$

Hence, with obvious notations,

$$\sigma_k(fg) \leq M(\sigma_k(f) + \sigma_k(g)),$$

so that

$$\Delta(D, fg) \leq M[\Delta(D, f) + \Delta(D, g)];$$

and fg is integrable, by Theorem 57, since the right-hand side tends to zero as $L(D) \to 0$.

(iii) If $|g| \geq d > 0$, then

$$\left|\frac{1}{g(P)} - \frac{1}{g(Q)}\right| = \left|\frac{g(Q) - g(P)}{g(P)g(Q)}\right| \leq \frac{|g(Q) - g(P)|}{d^2},$$

so that $\sigma_k(1/g) \leq d^{-2}\sigma_k(g)$; and $1/g$ is integrable, again by Theorem 57.

Finally, $f/g = f(1/g)$ is integrable, by (ii).

Some more of the familiar properties of the R-integral (for instance, the mean-value theorems) will be contained in the corresponding theorems for the Lebesgue integral. Direct proofs are very similar to those given later for the L-integral.

4.5. Deficiencies of the *R*-integral. It is of great importance, in the applications of the integral calculus, that integration should be efficient as regards limiting processes.

Consider a sequence of functions $f_k(P)$, integrable over a set E, and converging there to a function $f(P)$. It is desirable that we should have

$$\lim \int_E f_k(P)dP = \int_E \lim f_k(P)dP = \int_E f(P)dP \; ; \quad (4.5.1)$$

that is, it should be permissible to interchange the limiting process and the process of integration.

The formula (4.5.1) is not generally true for the R-integral. First, the limit function f need not be integrable. Thus we know that the characteristic function $\chi(x)$ of the (linear) rational set R is not R-integrable (over the whole line). Yet it is the limit of the integrable functions $\chi_k(x)$, where χ_k is the characteristic function of the finite set $\{r_1, r_2 \dots r_k\}$, the r_k being the rational numbers in some order of enumeration.

Secondly, the formula (4.5.1) need not hold even when the limit f of the functions f_k is itself integrable over E. Take as $f_k(x)$ a continuous function which is equal to zero when $1/k \leq x \leq 1$, and whose graph forms, over $\langle 0, 1/k \rangle$, an isosceles triangle Δ of height $2k$. Clearly,

$$\int_0^1 f_k(x)dx = c(\Delta) = 1,$$

so that the limit to the left in (4.5.1) is 1. On the other hand, $f_k(x) \to 0$ throughout $\langle 0, 1 \rangle$, so that the integral of the limiting function $f(\equiv 0)$ is nought.

The following theorem is the main practicable test for the validity of (4.5.1) in the case of the R-integral. However, for many applications its conditions are too restrictive.

THEOREM 60. *Suppose that* E *is bounded; that the functions* $f_k(P)$ *are* R*-integrable over* E; *and that* $f_k(P) \to f(P)$ *uniformly in* E.

Then f(P) *is* R*-integrable over* E, *and* (4.5.1) *holds.*

PROOF. We may assume that E is an interval I. Also, clearly,

$$(f_k)_+ \to f_+, \quad (f_k)_- \to f_-,$$

uniformly in I. We may, therefore, assume that $f_k \geq 0$ (and $f \geq 0$). Now, $f_k - \epsilon_k \leq f \leq f_k + \epsilon_k$, where $\epsilon_k \to 0$, uniformly in I. Hence,

$$\int_I f_k dP - \epsilon_k |I| \leq \underline{\int_I} f dP \leq \overline{\int_I} f dP \leq \int_I f_k dP + \epsilon_k |I|,$$

by (4.2.3). Taking $k=1$, we see that the upper integral of f is finite. Then, letting $k \to \infty$, we conclude, first, that f is integrable; and also, that the limit of the integrals of the f_k exists and that it equals the integral of f.

The theorem is not true for an unbounded set. Thus, if $f_k(x) = 1/x$ in $\langle 1, k \rangle$, $f_k(x) = 0$ for $x \geq k$, then $f_k(x)$ is integrable over $\langle 1, \infty \rangle$; and $f_k(x) \to 1/x$ uniformly in $\langle 1, \infty \rangle$. Yet $f(x) = 1/x$ is not integrable over $\langle 1, \infty \rangle$, though it has an (infinite) R-integral.

A limiting function without R-integral is obtained, in a similar way, on replacing $1/x$ by $\sin x/x$.

In § 5.8 we shall obtain a further test concerning (4.5.1) (*Arzelà's* test).

Further deficiencies of the R-integral, in connection with the theory of the indefinite integral, will be discussed in Chapter VI.

Solutions to Exercises

Ex. 21. If the lower integral of f equals the upper one, then

$$\underline{\int_E} f_+ dP - \overline{\int_E} f_+ dP = \overline{\int_E} (-f_-) dP - \underline{\int_E} (-f_-) dP.$$

Both sides are of opposite signs and hence are nought: f_+ and $-f_-$ are integrable, and so is f. The rest of the exercise is obvious.

Ex. 22. Let I_k be the segment $\langle x = r_k, 0 \leq y \leq 1/k \rangle$ in the (x, y)-plane, that is, a plane interval of area zero; and let J_k be a closed plane interval of height $1/k$ over $\langle 0, 1 \rangle$. The interval sum

$$S_k = I_1 + I_2 + \ldots + I_k + J_k$$

plainly contains the ordinate set Y of f. Also $|S_k| = |J_k| = 1/k$. Hence $\bar{c}(Y) \leq 1/k$, for every k. This implies $c(Y) = 0$.

Ex. 23. By (4.3.10) and (4.4.5), applied to f_+ and $-f_-$,

$$\int_I f dP = \underline{\int_I} f_+ dP - \overline{\int_I} (-f_-) dP = \sup s^+(D) - \inf S^-(D) = A - B,$$

say. By (4.4.6), we have $s(D) = s^+(D) - S^-(D)$. Plainly, $\sup s(D) \leq A - B$. On the other hand, given $\epsilon (>0)$, we can find divisions D_1 and D_2 such that $s^+(D_1) \geq A - \epsilon/2$ and $S^-(D_2) \leq B + \epsilon/2$. By (4.4.3), a common subdivision D will satisfy the same inequalities, so that $s(D) \geq A - B - \epsilon$, for every ϵ. This implies $\sup s(D) \geq A - B$, and completes the proof of the first formula (4.4.5). The proof for the second formula is similar.

CHAPTER V

LEBESGUE'S INTEGRAL

5.1. Lower and upper L-integrals. The classical Riemann integral is based on content as the underlying notion of volume. Now we replace content by measure and so obtain the more powerful modern definition of the integral as given by *H. Lebesgue* (1902).

We start as in § 4.2. Let Y be the ordinate set of a non-negative function $y = f(P)$, defined in a set E of the space $\mathsf{E} = \mathsf{E}_n$. We write †

$$\underline{\int_E} f dP = \underline{m}(Y), \quad \overline{\int_E} f dP = \overline{m}(Y) \quad . \quad . \quad (5.1.1)$$

for the inner and outer measures of Y (in $\mathsf{E}^* = \mathsf{E}^*_{n+1}$), and call these measures *the lower and upper Lebesgue integrals* (L-integrals) of $f(P)$ over E, respectively. They are non-negative and may be infinite.

As in § 4.2 we have

$$\underline{\int_E} f dP \leq \overline{\int_E} f dP \; ; \quad . \quad . \quad . \quad (5.1.2)$$

and, if $0 \leq f \leq g$ in E, then

$$\underline{\int_E} f dP \leq \underline{\int_E} g dP, \quad \overline{\int_E} f dP \leq \overline{\int_E} g dP \; ; \quad . \quad . \quad (5.1.3)$$

and

$$\underline{\int_{E_{[a]}}} f_{[a]} dP \uparrow \underline{\int_E} f dP, \quad \overline{\int_{E_{[a]}}} f_{[a]} dP \uparrow \overline{\int_E} f dP \quad . \quad (5.1.4)$$

as $a \uparrow \infty$. [Compare (4.2.2)-(4.2.5)].

† We use the same symbol of integration for both the R- and the L-integrals, unless a distinction becomes unavoidable.

Next, if $\alpha \geq 0$, then

$$\underline{\int_E} \alpha f dP = \alpha \underline{\int_E} f dP, \quad \overline{\int_E} \alpha f dP = \alpha \overline{\int_E} f dP. \quad . \quad (5.1.5)$$

The proof is similar to that of (4.2.6). E and f may be assumed as bounded. For the proof of the second formula the outer interval sums are to be replaced by outer interval sets.

For the proof of the first formula we use closed sets F contained in Y, in view of (3.5.10). Now, let αF denote the set of all points $(P\,;\ \alpha x_{n+1})$ where P belongs to E and $(P\,;\ x_{n+1})$ belongs to F. Then αF is a closed subset of αY, the ordinate set of αf; plainly, any closed subset of αY can thus be obtained. Also, by the argument for (4.2.6), $\bar{c}(\alpha F) = \alpha \bar{c}(F)$. Hence, by (3.5.10), $\underline{m}(\alpha Y) = \alpha \underline{m}(Y)$ which is the desired result.

The set E lies in the n-dimensional subspace E of E^*. We denote the inner and outer measures of E, in E, by $\underline{\mu}(E)$ and $\bar{\mu}(E)$, respectively.

THEOREM 61.

$$\underline{\int_E} dP = \underline{\mu}(E), \quad \overline{\int_E} dP = \bar{\mu}(E). \quad . \quad . \quad (5.1.6)$$

This is proved as Theorem 47; outer interval sums are replaced by outer interval sets, and inner interval sums by closed subsets of E. As a corollary we have again:

If the upper L*-integral of* f *is finite, then* $\mu(\mathrm{E}_\infty) = 0$, *where* E_∞ *is the subset of* E *in which* $\mathrm{f} = \infty$.

By (3.4.6) and (3.5.4) we may omit from Y any subset of measure zero in E^* (for instance, the "bottom" E) without altering $\underline{m}(Y)$ and $\bar{m}(Y)$; and we may omit from E, or add to E, any set in E where $f \equiv 0$. In particular, whenever $E_1 \supset E$,

$$\underline{\int_{E_1}} f^* dP = \underline{\int_E} f dP, \quad \overline{\int_{E_1}} f^* dP = \overline{\int_E} f dP, \quad . \quad (5.1.7)$$

where f^* is the extension of f.

5.2. Lebesgue's integral. From now on we use m and μ to denote measures in E* and E, respectively.

Let $f(P)$ be a non-negative function in E. If its ordinate set is measurable, then we write

$$\int_E f(P)dP = m(Y) \quad . \quad . \quad . \quad (5.2.1)$$

and call this measure the *Lebesgue integral* (L-integral) of f over E. The integral is non-negative and may be infinite.

The existence of the L-integral implies that the lower and upper integrals (5.1.1) should be equal. If the integral is finite this is also sufficient (compare remark after (4.3.1)).

As in § **4.3** we have

$$\int_{E_{[a]}} f_{[a]}dP \uparrow \int_E fdP \quad . \quad . \quad . \quad (5.2.2)$$

as $a \uparrow \infty$; and

$$\int_E 0 \; dP = 0 \quad . \quad . \quad . \quad . \quad (5.2.3)$$

over any set E. We shall also require the reverse of (5.2.3).

THEOREM 62. *Suppose that* $f \geq 0$, *and that its* L-*integral over* E *vanishes. Then* $f \equiv 0$ *in* E *except perhaps for a subset of measure zero.*

PROOF. Let $a \geq 0$, and denote by E_a the subset of E where $f > a$. Then

$$a\bar{\mu}(E_a) = a \bar{\int_{E_a}} dP \leq \bar{\int_{E_a}} fdP \leq \int_E fdP = 0$$

by (5.1.3), (5.1.5), and (5.1.6). Hence $\mu(E_a) = 0$ for every $a > 0$. Now $E_a \uparrow E_0$ when $a \downarrow 0$. By Theorem 45,

$$\mu(E_a) \uparrow \mu(E_0).$$

Hence E_0, the set where $f > 0$, has measure zero.

A non-negative function $f(P)$ is said to be L-*integrable* over E, if its integral is finite.

The definition extends to a general function :

A general function $f(P)$ is *L-integrable* over E if the two components f_+ and $-f_-$, of (4.1.6), are L-integrable; and then *the* L-*integral of* f *is defined by* (4.1.8).

The function f may have values ∞ and $-\infty$. However, if f is integrable, then the set where this happens has measure zero. This follows from the corollary of Theorem 61.

The next formulae and theorems (up to (5.2.9) inclusive) are proved in exactly the same way as the corresponding results for the R-integral. First,

$$\int_{E_1} f^*(P)dP = \int_E f(P)dP, \quad . \quad . \quad . \quad (5.2.4)$$

whenever $E_1 \supset E$; and

$$\int_E f dP \leq \int_E g dP, \quad . \quad . \quad . \quad (5.2.5)$$

if f and g have L-integrals over E, and $f \leq g$. Next,

THEOREM 63. *If* f *is* L-*integrable over* E, *then* αf *is* L-*integrable, and* †

$$\int_E \alpha f dP = \alpha \int_E f dP. \quad . \quad . \quad . \quad (5.2.6)$$

THEOREM 64. *If* $\mu(\mathrm{E}) = 0$, *then any function* f(P) *is* L-*integrable over* E; *and its integral is zero.*

THEOREM 65. *The characteristic function* $\chi(\mathrm{P})$ *of a set* E *has an* L-*integral over* E (*or* E) *if, and only if,* E *is measurable; and then*

$$\int_{\mathsf{E}} \chi dP = \int_E dP = \mu(E). \quad . \quad . \quad . \quad (5.2.7)$$

Thus, the characteristic function of Vitali's set V [§ 3.9] is an example of a non-negative function without L-integral. On the other hand, the characteristic function of the rational set R is L-integrable; its integral vanishes since $\mu(R) = 0$. We recall that this function has no R-integral.

† If $f \geq 0$ and $\alpha > 0$, then it suffices that f has an integral (compare footnote to Theorem 48).

THEOREM 66. *Suppose that* f *has an* L-*integral over* E. *Then it has also an* L-*integral over any measurable subset* E_1.

We omit, for the present, the theorem that corresponds to Theorem 52; we shall soon prove it in a more general form.†

THEOREM 67. *Suppose that* $E_1 \subset E$, *and that* f *is* L-*integrable over* E_1, *and has an* L-*integral over* E. *Then* f *has also an* L-*integral over* $E - E_1$, *and*

$$\int_{E-E_1} f dP = \int_E f dP - \int_{E_1} f dP. \quad . \quad . \quad (5.2.8)$$

THEOREM 68. *Suppose that* f *has an* L-*integral over* E, *and that* f = g *except perhaps in a subset* E_1 *of measure zero. Then* g *has an* L-*integral over* E, *equal to that of* f.

Two such functions f and g must be considered as equivalent in the theory of the L-integral.

We shall say that a certain property holds *almost everywhere* (or *p.p.*) ‡ in a set E, if it holds throughout E except perhaps for a subset of measure zero. Thus a function f, which is L-integrable over E, is finite p.p. in E. If $f = g$ p.p. in E, then f and g have the same L-integrals: they are equivalent.

The values of a function f in a set of measure zero are of no relevance as regards integration: f can be changed arbitrarily in such a set, or it may not be defined there at all. In fact, it will be convenient to speak of the L-integral of a function f over E even if f is only defined p.p. in E.

Next, we emphasize once more that our integrals are absolute integrals: $|f|$ has also an integral, and

$$\left| \int_E f dP \right| \leq \int_E |f| dP. \quad . \quad . \quad . \quad (5.2.9)$$

We close this paragraph with the following generalised analogue of Theorem 52, the importance of which lies in the fact that we can split E into an infinite sequence of subsets.

† The proof for the exact analogue is as before.

‡ *p.p.* indicates the French *presque partout.*

THEOREM 69. *Suppose that* $E = E_1 + E_2 + \ldots + E_k + \ldots$ *is a finite or infinite sum of sets, and that* $\mu(E_k \,.\, E_l) = 0$ *for any two* $k \neq l$. *If* f *is* L-*integrable over each* E_k, *then it is* L-*integrable over* E, *and* †

$$\int_E f dP = \int_{E_1} f dP + \int_{E_2} f dP + \ldots + \int_{E_k} f dP + \ldots; \qquad (5.2.10)$$

provided, however (in the case of an infinite sum), that

$$\int_{E_1} |f| dP + \int_{E_2} |f| dP + \ldots + \int_{E_k} |f| dP + \ldots < \infty. \quad (5.2.11)$$

PROOF. We may assume that $f \geq 0$. Then $Y = \Sigma Y_k$ is the ordinate set of f over E, where Y_k is the ordinate set over E_k. Now $Y_k \,.\, Y_l$ is the ordinate set over $E_k \,.\, E_l$. It has measure zero (in E*) when $k \neq l$, by Theorem 64, since $\mu(E_k \,.\, E_l) = 0$. Also all the Y_k are measurable. Hence Y is measurable, and $m(Y) = \Sigma m(Y_k)$, by (3.8.2). This is (5.2.10). The (non-negative) f has an integral over E; and if (5.2.11) is satisfied, this integral is finite and f is integrable.

Exercise 24. For a general f the lower and upper integrals are defined by (4.3.10). Prove Exercise 21 for the L-integral.

5.3. Limiting theorems. In § 4.5 we have discussed the deficiencies of the R-integral as regards limiting processes. One of the great advantages of the L-integral is its smooth applicability in this respect. Our first theorem concerns non-negative functions and is a straightforward application of Theorem 45 to ordinate sets.

THEOREM 70. *Suppose that the non-negative functions* f_k *have* L-*integrals over* E. *Then the functions* $\underline{\lim}\, f_k$ *and* $\overline{\lim}\, f_k$ *have also* L-*integrals and* ‡

$$\int_E \underline{\lim}\, f_k dP \leq \underline{\lim} \int_E f_k dP. \quad . \quad . \quad . \quad (5.3.1)$$

† If $f \geq 0$ and the integrals over the E_k exist, then the integral of f exists, and (5.2.10) holds without the condition (5.2.11).

‡ (5.3.1) is known as *Fatou's Lemma* (for integrals).

Also

$$\int_E \overline{\lim}\, f_k dP \geq \overline{\lim} \int_E f_k dP, \quad . \quad . \quad (5.3.2)$$

provided, however, that

$$f_k(P) \leq \Phi(P), \quad . \quad . \quad . \quad (5.3.3)$$

where Φ *has a finite upper* L-*integral over* E.

In particular, if $f_k \to f$, *then* f *has an* L-*integral over* E. *Also*

$$\int_E f_k dP \to \int_E f dP, \quad . \quad . \quad . \quad (5.3.4)$$

provided that (5.3.3) *is satisfied.*

If $f_k \uparrow f$ *then* (5.3.3) *is not required. If* $f_k \downarrow f$ *this restriction is equivalent to* f_p *being* L-*integrable, for some* p.

Proof. The Y_k are measurable. By Theorem 45, $\underline{\lim}\ Y_k$ and $\overline{\lim}\ Y_k$ are also measurable, and

$$m(\underline{\lim}\ Y_k) \leq \underline{\lim}\ m(Y_k), \quad m(\overline{\lim}\ Y_k) \geq \overline{\lim}\ m(Y_k). \qquad \text{(a)}$$

The second inequality requires the condition

$$m(Y_p + Y_{p+1} + \ldots) < \infty \quad . \quad . \quad . \quad \text{(b)}$$

for some p.

Next, let L and Λ be the ordinate sets of $\underline{\lim}\ f_k$ and $\overline{\lim}\ f_k$, respectively †; and, for $0 < \theta < 1$, let θL and $\theta\Lambda$ be the ordinate sets of $\theta\ \underline{\lim}\ f_k$ and $\theta\ \overline{\lim}\ f_k$.

Now $\underline{\lim}\ Y_k$ is the set of all points which belong "eventually" to all Y_k. Clearly, $\theta L \subset \underline{\lim}\ Y_k \subset L$. Hence, by (5.1.5),

$$\theta \bar{m}(L) \leq m(\underline{\lim}\ Y_k) \leq \underline{m}(L).$$

On letting $\theta \uparrow 1$ we find that L is measurable and $m(L) = m(\underline{\lim}\ Y_k)$. The first inequality (a) now gives (5.3.1).

Similarly, $\overline{\lim}\ Y_k$ is the set of all points which belong to an infinity of the Y_k. Since $\theta\Lambda \subset \overline{\lim}\ Y_k \subset \Lambda$, we conclude as before that Λ is measurable and that $m(\Lambda) = m(\overline{\lim}\ Y_k)$.

† It is not necessarily true that $L = \underline{\lim}\ Y_k$, or $\Lambda = \overline{\lim}\ Y_k$.

The second inequality (a) now gives (5.3.2.). If (5.3.3) is satisfied, then $Y_k \subset Y^*$, the ordinate set of Φ, so that

$$m(Y_1 + Y_2 + \ldots + Y_k + \ldots) \leq \bar{m}(Y^*) < \infty,$$

and the condition (b) is satisfied.

The rest of the theorem follows at once from the corresponding parts of Theorem 45.

It is easy to extend the formula (5.3.4) to general functions. The following theorem is one of the most important results of the integral calculus. It is known as *Lebesgue's Test*.

THEOREM 71. *Suppose that the functions* $\mathrm{f_k}$ *are* L-*integrable over* E, *and that*

$$|f_k(P)| \leq \Phi(P) \quad . \quad . \quad . \quad (5.3.5)$$

is satisfied, where Φ *has a finite upper* L-*integral over* E.

If $\mathrm{f_k} \to \mathrm{f}$ *p.p. in* E, *then* f *is* L-*integrable over* E, *and*

$$\int_E f_k dP \to \int_E f dP. \quad . \quad . \quad . \quad (5.3.6)$$

PROOF. We may assume that $f_k \to f$ throughout E. Now, $f_k \to f$ implies $(f_k)_+ \to f_+$ and $(f_k)_- \to f_-$. We may, therefore, assume that $f_k \geq 0$. By Theorem 70, f has an L-integral over E. Also (5.3.6) holds, because of (5.3.5); and it follows that the integral of f is not greater than the upper integral of Φ. Hence f is integrable.

That some restriction of the type (5.3.5) is necessary for the validity of (5.3.6) is shown by the second example after (4.5.1). (Here (5.3.5) is, clearly, not satisfied.)

We shall say that a sequence of functions f_k converges *dominatedly* to f in E, if $f_k \to f$ and if (5.3.5) is satisfied in E for some such Φ. Lebesgue's Test asserts that dominated convergence p.p. justifies the limit relation (5.3.6). In particular, a restriction

$$|f_k(P)| \leq M, \quad . \quad . \quad . \quad (5.3.7)$$

where M does not depend on k, is sufficient, provided that E has finite outer measure.

For then, by (5.1.5) and (5.1.6), M has a finite upper integral over E. In the case (5.3.7) we speak of *bounded convergence* in E.

We remark that the formula (5.2.2) is a special case of (5.3.4) : the extension of $f_{[a]}$, from $E_{[a]}$ into E, tends increasingly to $f(\geq 0)$.

Similarly, when the non-negative function f has an L-integral over E, then

$$\int_E f_{[a]}dP \uparrow \int_E f dP \quad . \quad . \quad . \quad (5.3.8)$$

as $a \uparrow \infty$ (the ordinate set of $f_{[a]}$, as product of the measurable sets Y and $\langle \mathsf{E},\ 0 \leq x_{n+1} \leq a \rangle$, is measurable : $f_{[a]}$ has an integral over E).

THEOREM 72. *Suppose that* f *is* L-*integrable over* E. *Then, given an arbitrary positive* ϵ, *there exists a* $\delta = \delta(\epsilon)$ *such that*

$$\int_{E_1} |f| dP < \epsilon \quad . \quad . \quad . \quad . \quad (5.3.9)$$

for every measurable subset E_1 *of* E *for which* $\mu(\mathrm{E}_1) < \delta$.

PROOF. First, $|f|$ is L-integrable over E, and, by Theorems 66 and 67, also over E_1 and $E - E_1$. The same holds for $|f|_{[a]}$. Hence, by (5.2.8),

$$\begin{aligned} 0 \leq \int_{E_1} |f| dP - \int_{E_1} |f|_{[a]} dP &= \int_E |f| dP - \int_E |f|_{[a]} dP \\ &\quad - \Big[\int_{E-E_1} |f| dP - \int_{E-E_1} |f|_{[a]} dP \Big] \\ &\leq \int_E |f| dP - \int_E |f|_{[a]} dP. \end{aligned}$$

The right-hand side, by (5.3.8) applied to $|f|$, is at most $\frac{1}{2}\epsilon$ if $a = A(\epsilon)$ is large enough. Here A does not depend on E_1 (although it does depend on E). Hence

$$\int_{E_1} |f| dP \leq \int_{E_1} |f|_{[A]} dP + \tfrac{1}{2}\epsilon \leq A \int_{E_1} dP + \tfrac{1}{2}\epsilon = A\mu(E_1) + \tfrac{1}{2}\epsilon < \epsilon,$$

when $\mu(E_1) < \frac{1}{2}\epsilon A^{-1} = \delta(\epsilon)$. This proves the Theorem.

Exercise 25. If $f_k(x) = 1/2k$ for $|x| \leq k$, and $f_k(x) = 0$ otherwise, then $f_k(x) \to 0$ boundedly for all x. But

$$\int_{-\infty}^{\infty} f_k(x)dx = 1.$$

Why does (5.3.6) not hold?

5.4. Measurable functions. We shall see, in the next paragraph, that the following class of functions is of fundamental importance in the theory of the L-integral.

A function $f(P)$, defined in the whole space E, is said to be *measurable* if the sets $[f > a]$, that is the sets where $f > a$, are measurable for every real number a.

A function $f(P)$, defined in E, is said to be *measurable in* E if its extension $f^*(P)$ is measurable.

Exercise 26. If E is measurable, then f is measurable in E if, and only if, all the sets $[f > a]$ are measurable.

If f is measurable then all the sets

$$[f \geq a], \quad [f = a], \quad [f = \infty] \qquad . \quad . \quad (5.4.1)$$

are measurable. For, $[f > b] \downarrow [f \geq a]$ as $b \uparrow a$, so that the latter set is measurable, by Theorem 45; $[f = a] = [f \geq a] - [f > a]$ is the difference of two measurable sets; and $[f > a] \downarrow [f = \infty]$ as $a \uparrow \infty$.

Also all the sets

$$[f < a], \quad [f \leq a], \quad [f = -\infty] \qquad . \quad . \quad (5.4.2)$$

are measurable. For, $[f < a] = {}_c[f \geq a]$; $[f \leq a] = {}_c[f > a]$; and $[f < a] \downarrow [f = -\infty]$ as $a \downarrow -\infty$.

Exercise 27. If f is measurable, prove that all the sets

$$[a < f < b], \quad [a \leq f \leq b], \quad [a \leq f < b], \quad [a < f \leq b] \quad . \quad (5.4.3)$$

are measurable.

Exercise 28. Prove that f is measurable when all the sets of one of the following types are measurable: (i) $[f \geq a]$; (ii) $[f < a]$; (iii) $[f \leq a]$; (iv) $[f > r]$ where r is rational.

If f *is measurable, then* f + c *is measurable.* For, $[f+c>a] = [f > a - c]$.

If f *is measurable in* E, *then* | f | *is measurable in* E.

For, $[|f|^* > a] = [f^* > a] + [f^* < -a]$ when $a \geq 0$, and $[|f|^* > a] = E$ when $a < 0$.

If f *is measurable in* E, *then* cf, *and* $|f|^\alpha$ *for* $\alpha > 0$ *(in particular,* f^2*), are measurable in* E. We leave the simple proof to the reader.

Next, *if* f *is measurable in* E, *and* f ≠ 0 *p.p.*,† *then* 1/f *is measurable in* E.

For, we may assume that $f \neq 0$ throughout E.† Then

$$[(1/f)^* > a] = [0 < f^* < 1/a]$$

when $a > 0$;

$$[(1/f)^* > 0] = [0 < f^* < \infty];$$

and $[(1/f)^* > a] = [f^* \geq 0] + [f^* < 1/a]$ when $a < 0$. All these sets are measurable.

If f *and* g *are measurable, then the sets*

$$[f > g], \quad [f \geq g], \quad [f = g] \quad . \quad . \quad . \quad (5.4.4)$$

are measurable.

For, at a point where $f > g$, there exists a rational number r such that $f > r > g$, and conversely. Hence

$$[f > g] = \Sigma[f > r_k] \cdot [g < r_k],$$

where the sum is extended over all rational numbers in some order of enumeration. Each term of the sum is measurable, by Theorem 41; and so is the sum, by Theorem 43. Also $[f \geq g] = {}_c[f < g]$ and $[f = g] = [f \geq g] - [f > g]$.

THEOREM 73. *If* f *and* g *are measurable and finite p.p. in* E,‡ *then* f + g, f − g, *and* fg *are measurable in* E.

† Sets of measure zero are irrelevant; $1/f$ is defined p.p.

‡ In order to define $f \pm g$, fg, and f/g p.p.

If f *and* g *are measurable in* E, *and if* f *is finite and* $g \neq 0$ *p.p. in* E,‡ *then* f/g *is measurable in* E.

PROOF. If g^* is measurable, then $-g^*$ and $a+(-g^*) = a-g^*$ are measurable. If f^* is also measurable, then $[(f+g)^*>a]=[f^*+g^*>a]=[f^*>a-g^*]$ is measurable, by (5.4.4). Hence $f+g$ is measurable in E; and so is $f-g = f+(-g)$.

Also, as stated before, $f^2=|f|^2$, g^2, and $(f+g)^2$ are measurable in E. Hence $fg=\frac{1}{2}[(f+g)^2-f^2-g^2]$ is measurable in E.

Finally, under the assumptions of the last clause, $1/g$ is measurable in E. Hence $f/g=f(1/g)$ is measurable in E.

Perhaps the most important property of measurable functions is contained in the following theorem.

THEOREM 74. *Suppose that the functions* $f_k(P)$, *of a given sequence, are measurable in* E. *Then the functions*

$$\underline{\phi}(P)=\inf f_k(P), \quad \bar{\phi}(P)=\sup f_k(P) \qquad . \quad (5.4.5)$$

and

$$\underline{\Phi}(P)=\underline{\lim}\, f_k(P), \quad \bar{\Phi}(P)=\overline{\lim}\, f_k(P) \qquad . \quad (5.4.6)$$

are also measurable in E.

In particular, if $f_k \to f$ *p.p. in* E, *then* f *is measurable in* E.

PROOF. On considering the extensions we may assume that the f_k are measurable (in E)

(i) If $\bar{\phi}>a$ at a given point P, then at least one $f_k>a$; the converse is also true. Hence $[\bar{\phi}>a]=\Sigma[f_k>a]$. It follows that the sets $[\bar{\phi}>a]$ are measurable: $\bar{\phi}$ is measurable; and $\underline{\phi}=-\sup(-f_k)$ is also measurable.

(ii) By (i) the function $\bar{\phi}_k=\sup(f_k, f_{k+1}\ldots)$ is measurable; and so is $\bar{\Phi}$, since $\bar{\Phi}=\inf \bar{\phi}_k$, by (1.8.3). Finally, $\underline{\Phi}=-\overline{\lim}\,(-f_k)$ is measurable.

We note that, when f is measurable in E, then $f_+ = \text{Max}\,(f,0)$ and $f_- = \text{Min}\,(f,0)$ are measurable in E, by (5.4.5.)

The converse is also true, by Theorem 73, since $f = f_+ + f_-$, this sum being defined in E.

The simplest class of measurable functions is that of the continuous functions. Let O be an open set, say, and let f be continuous in O. Then the sets $[f > a]$ are, clearly, open and thus measurable. Hence f is measurable in O, by Exercise 26.

By Theorem 74 we can ascend from continuous functions through limiting processes to more involved classes of measurable functions. Thus the function

$$\chi(x) = \lim_{k \to \infty} [\lim_{m \to \infty} \cos (k!\,\pi x)^{2m}]. \quad . \quad (5.4.7)$$

is measurable as a double limit of continuous functions. The inner limit has the value 1 when x is a rational number such that, in reduced form, its denominator is at most $k!$; and the inner limit is nought otherwise. Hence $\chi(x)$ is the characteristic function of the linear rational set. The example also shows that such apparently "artificial" functions can be obtained in a "natural" way from "simple" functions: the distinction between simple and artificial functions can hardly be maintained in analysis.

Exercise 29. Prove that, if f is continuous in a closed set F, then it is measurable in F.

Exercise 30. Prove that, if f is continuous p.p., then it is measurable.

Exercise 31. Let C be the set where the measurable functions f_k have a finite limit f. Show that C is measurable.

The importance of uniform convergence in analysis is well known (see Theorem 60). We conclude this paragraph by showing that *any* convergent sequence of measurable functions is, in a certain sense, "nearly" uniformly convergent. This is the gist of the following theorem of *Egoroff* (1911).

THEOREM 75. *Suppose that the functions* f_k *are measurable in a set* E *of finite measure; and that* $f_k \to f$ *p.p. in* E, *where* f *is finite. Then, given an arbitrary positive* ϵ, *there exists a subset* $E_1 = E_1(\epsilon)$ *of* E, *of measure* $\mu(E_1) > \mu(E) - \epsilon$, *such that* $f_k \to f$ *uniformly in* E_1.

PROOF. We may assume that $f_k \to f$ throughout E, where f is finite. Then f and $|f - f_k|$ will be measurable in E. Also all the sets which occur in this proof will be measurable since E is measurable (Exercise 26).

Consider a sequence of numbers $\eta_p \to 0$. For fixed p the sets

$$E_{K,p} = \prod_{k=K}^{\infty} [|f_k - f| < \eta_p]$$

increase to E when $K \uparrow \infty$. For, $f_k - f \to 0$ in E and, therefore, every point of E will belong to all sets $[|f_k - f| < \eta_p]$ for which k is sufficiently large. It follows that $\mu(E_{K,p}) \uparrow \mu(E)$ when $K \uparrow \infty$, and that $\mu(E) - \mu(E_{K,p}) < 2^{-p}\epsilon$, when $K \geq K_p(\epsilon) = K_p$, the ϵ being fixed.

Next, let

$$E_1 = \prod_{p=1}^{\infty} E_{K_p,p}.$$

Whenever P belongs to E_1, it belongs to all $E_{K_p,p}$. Hence, and since $\eta_p \to 0$, we can find, for given η, an $\eta_p < \eta$ (and the corresponding K_p) such that, at P, $|f_k - f| < \eta_p < \eta$ for all $k \geq K_p$. This K_p will depend on η only, not on P: that is, $f_k \to f$ uniformly in E_1. Also

$$E - E_1 = \sum_{p=1}^{\infty} (E - E_{K_p,p}),$$

by (1.7.2). Hence

$$\mu(E - E_1) = \mu(E) - \mu(E_1) \leq \sum_{p=1}^{\infty} \mu(E - E_{K_p,p})$$
$$< \epsilon \,.\, \Sigma 2^{-p} = \epsilon,$$

by (3.8.1.) This completes the proof.

5.5. Integrable functions. Suppose, first, that $f(P)$ is a non-negative function, defined in E; and that Y is its ordinate set. Let $a \geq 0$. By E_a, or $[f > a]$, we denote the subset of E where $f > a$; and by Y_a the set $[E_a\,;\ a < x_{n+1} \leq f]$ in E^*.

Clearly, $E_a \uparrow E_0$ and $Y_a \uparrow Y_0$ as $a \downarrow 0$; here E_0 is the set where $f > 0$, and Y_0 is the set $[E_0\,;\ 0 < x_{n+1} \leq f]$. If $a \uparrow \infty$, then $E_a \downarrow E_\infty$, the set where $f = \infty$; and Y_a decreases to the empty set.

Now let $0 \leq a < b$. The set

$$Y_{a,b} = Y_a - Y_b \quad . \quad . \quad . \ (5.5.1)$$

is the product of Y by the "strip" $[E\,;\ a < x_{n+1} \leq b]$. Clearly,

$$[E_b\,;\ a < x_{n+1} < b] \subset Y_{a,b} \subset [E_a\,;\ a \leq x_{n+1} \leq b]. \quad (5.5.2)$$

Hence, by an obvious extension of (5.1.6),

$$(b-a)\bar{\mu}(E_b) \leq \bar{m}(Y_{a,b}) \leq (b-a)\bar{\mu}(E_a)\,; \quad . \ (5.5.3)$$

there are corresponding inequalities for the inner measures. Also

$$\bar{m}(Y_{a,c}) = \bar{m}(Y_{a,b}) + \bar{m}(Y_{b,c}) \quad . \quad . \ (5.5.4)$$

when $0 \leq a < b < c$; and there is a similar equality for the inner measures.†

For, $Y_{a,c} = Y_{a,b} + Y_{b,c}$. Here the two terms are, plainly, separated by an interval, if E is bounded; in which case we can apply Theorems 36 and 38. For unbounded E we approximate, as usual, through the sets $E^{(k)}$ of (2.5.2).

Next, we divide the interval $a \leq y \leq b$ into k equal parts by the points $y_i = a + i\dfrac{b-a}{k}$, $0 \leq i \leq k$, and apply (5.5.3) to each part. By (5.5.4), we obtain on adding

$$\frac{b-a}{k}\sum_{1}^{k} \bar{\mu}(E_{y_i}) \leq \bar{m}(Y_{a,b}) \leq \frac{b-a}{k}\sum_{0}^{k-1} \bar{\mu}(E_{y_i})\,; \quad (5.5.5)$$

there are similar inequalities for the inner measures.

† If $E_a^* = [f \geq a]$, with corresponding definitions of Y_a^* and $Y_{a,b}^*$, then (5.5.3) and (5.5.4) remain valid for these sets. However, it is not true that $E_a^* \uparrow E\ (= E_0^*)$ as $a \downarrow 0$: in fact, $E_a^* \uparrow E_0$.

We can now prove the following theorem which is fundamental for the theory of the L-integral.

THEOREM 76. *A non-negative function* f(P) *has an* L-*integral over* E *if, and only if,* f(P) *is measurable in* E.

PROOF. (i) Suppose that f has an L-integral over E. We have to show that all sets $[f^* > a]$ are measurable. Now $[f^* > a] = \mathsf{E}$ when $a < 0$; and $[f^* > a] = [f > a] = E_a$ when $a \geq 0$. It remains to show that these sets E_a are measurable.

The set $Y_{a,b}$, as product of the measurable set Y by the measurable strip $[\mathsf{E};\ a < x_{n+1} \leq b]$, is measurable in E^*. Hence, by (5.5.3) for inner and outer measures,

$$(b-a)\bar{\mu}(E_b) \leq m(Y_{a,b}) \leq (b-a)\underline{\mu}(E_a),$$

so that $\bar{\mu}(E_b) \leq \underline{\mu}(E_a)$ for all $b > a (\geq 0)$. Now $E_b \uparrow E_a$ as $b \downarrow a$, so that $\bar{\mu}(E_b) \uparrow \bar{\mu}(E_a)$, by (3.8.7). It follows that $\bar{\mu}(E_a) \leq \underline{\mu}(E_a)$. Hence E_a is measurable if E, and thus E_a, is bounded. The case of an unbounded set E is dealt with in the now familiar way.

(ii) If f is measurable in E, then all sets E_a, where $a \geq 0$, are measurable. We may also suppose that E is bounded. Then (5.5.5), for inner and outer measures, becomes

$$\frac{b-a}{k}\sum_{1}^{k}\mu(E_{y_i}) \leq \underline{m}(Y_{a,b}) \leq \bar{m}(Y_{a,b}) \leq \frac{b-a}{k}\sum_{0}^{k-1}\mu(E_{y_i}).$$

The two sides differ by $\dfrac{b-a}{k}(\mu(E_a) - \mu(E_b))$ which tends to zero as $k \to \infty$. It follows that $Y_{a,b}$ is measurable. Also $Y_{0,b} \uparrow Y_0$ as $b \uparrow \infty$, so that Y_0 is measurable, by Theorem 45. Finally, $Y = Y_0 + E$ is measurable, because E, considered as a set in E^*, has measure zero there. Hence f has an integral over E.

It should be noted that *the set* $\mathrm{E}_0 = [\mathrm{f} > 0]$ *is necessarily measurable if the non-negative function* f *has an integral* *over* E. The set E itself need not be measurable (compare (5.2.3)); on the other hand, it is E_0 that matters.

We recall that, if a general function f is measurable in E, then $|f|$ is also measurable in E. Hence, by Theorem 76, $|f|$ has then an L-integral over E.

We can now extend the above theorem to general functions.

THEOREM 77. *A function* f(P) *is* L-*integrable over* E *if, and only if,* f *is measurable in* E, *provided that the integral of* | f(P) | *(which then exists) is finite.*

PROOF. The function f is integrable over E if, and only if, f_+ and $-f_-$ have finite integrals. By Theorem 76, these integrals exist if, and only if, f_+ and $-f_-$ are measurable in E. By the remark after Theorem 74, this, in turn, is equivalent to f being measurable in E. Also the integrals of f_+ and $(-f_-)$ will be finite if, and only if, that of $|f|$ is finite.

Suppose that the non-negative functions f and g have L-integrals over E. It is an immediate consequence of Theorems 73 (footnote) and 76 that then $f+g$ and fg, and also f/g if $g \neq 0$ p.p., have L-integrals as well. For general functions we have the following result.

THEOREM 78. *Suppose that* f(P) *and* g(P) *are* L-*integrable over* E. *Then* f(P) + g(P) *and* f(P) − g(P) *are also* L-*integrable over* E.

If, in addition, g(P) *is bounded p.p. in* E, *then* f(P)g(P) *is* L-*integrable over* E; *and if* | g(P) | $\geq d > 0$ *p.p., then* f(P)/g(P) *is* L-*integrable over* E.

PROOF. First, by Theorem 77, f and g are measurable in E, and they are finite p.p. Also, by Theorem 73, the functions $f+g$, fg, and f/g are measurable; so are their absolute values. Hence, by Theorem 76, the L-integrals of $|f \pm g|$, $|fg|$, and $|f/g|$ exist. It remains to show that they are finite.

In fact, the integral of $|f| + |g|$ exists, and hence

$$\int_E |f \pm g| dP \leq \int_E (|f| + |g|) dP \leq \int_{[|f| \geq |g|]} 2|f| dP + \int_{[|f| \leq |g|]} 2|g| dP$$
$$\leq 2[\int_E |f| dP + \int_E |g| dP] < \infty;$$

and

$$\int_E |fg| dP \leq M \int_E |f| dP < \infty, \quad \int_E |f/g| dP \leq \frac{1}{d} \int_E |f| dP < \infty,$$

if $|g| \leq M$, or $[g] \geq d$ p.p., respectively.

Theorem 79. *If* f(P) *and* g(P) *are* L-*integrable over* E, *then* †

$$\int_E (f(P) \pm g(P)) dP = \int_E f(P) dP \pm \int_E g(P) dP. \quad . \quad (5.5.6)$$

Proof. We consider $f+g$ only. Also we may assume that f and g are finite, and that E is measurable (by extension of the functions, if necessary).

(i) Let $f \geq 0$ and $g = c \geq 0$. The ordinate set of $f+c$ is the sum of the ordinate set C of $g = c$ and of a set Y^* (on top of C) which is congruent to the ordinate set Y of f. Also $C . Y^*$ is congruent to E and hence has measure zero, in E*. It follows that $m(C + Y^*) = m(Y) + m(C)$; that is,

$$\int_E (f + c) dP = \int_E f dP + \int_E c dP. \quad . \quad . \quad . \quad (a)$$

(ii) Let $f \geq 0$ and $g \geq 0$. For $k \geq 1$, $i \geq 0$, let $E_i^{(k)} = [i . 2^{-k} \leq g(P) < (i+1) . 2^{-k}]$, and consider the function $g_k(P)$ in E which has constant value $i . 2^{-k}$ in $E_i^{(k)}$. All the sets $E_i^{(k)}$ are measurable, and $E = \Sigma E_i^{(k)}$. Hence, by (a) and Theorems 66 and 69,

† If $f \geq 0$, $g \geq 0$, then the formula for $f+g$ holds whenever f and g have L-integrals.

$$\int_E (f+g_k)dP = \sum_i \int_{E_i^{(k)}} (f+g_k)dP = \sum_i \{ \int_{E_i^{(k)}} fdP + \int_{E_i^{(k)}} g_k dP \}$$

$$= \int_E fdP + \int_E g_k dP.$$

Also $g_k(P) \uparrow g(P)$ as $k \uparrow \infty$. Hence, by Theorem 70, we obtain (5.5.6) on letting $k \uparrow \infty$.

(iii) In the general case,

$$E = [f \geq 0 \ ; \ g \geq 0] + [f \geq 0 \ ; \ g < 0] + [f < 0 \ ; \ g \geq 0] + [f < 0 \ ; \ g < 0]$$

$$= E_1 + E_2 + E_3 + E_4,$$

say. All these sets are measurable, and, in view of (5.2.10), it suffices to prove (5.5.6) for them separately. The case E_1 has been proved in (ii), and the case E_4 reduced to this on considering $-f$ and $-g$ and using (5.2.6). Similarly, case E_3 reduces to case E_2, which remains to be discussed. Now

$$E_2 = [f \geq -g \ ; \ g < 0] + [0 \leq f < -g \ ; \ g < 0] = E_2{}^* + E_2{}^{**},$$

say. In $E_2{}^*$ we have $f = (f+g) + (-g)$, both terms being non-negative. Hence, by what we have proved in (ii),

$$\int_{E_2{}^*} fdP = \int_{E_2{}^*} (f+g)dP + \int_{E_2{}^*} (-g)dP = \int_{E_2{}^*} (f+g)dP - \int_{E_2{}^*} gdP,$$

which is the desired result. The same argument proves it for $E_2{}^{**}$, where $-g = f + \{-(f+g)\}$. This completes the proof of the theorem.

The addition formula (5.5.6) can be extended to infinite sums. The series

$$f_1(P) + f_2(P) + \ldots + f_k(P) + \ldots = f(P) \quad . \quad (5.5.7)$$

is said to be *dominatedly* (*boundedly*) *convergent in* E, if its partial sums

$$s_k(P) = f_1(P) + f_2(P) + \ldots + f_k(P)$$

converge to $f(P)$ dominatedly (boundedly) in E (§ 5.3). Applying Theorems 71 and 79 (repeatedly) we obtain the following result.

THEOREM 80. *Suppose that the functions* $f_k(P)$ *are* L-*integrable over* E, *and that the series* (5.5.7) *converges dominatedly p.p. in* E.

Then its sum f(P) *is* L-*integrable over* E, *and we may integrate the series term by term: i.e.* †

$$\int_E f_1(P)dP + \int_E f_2(P)dP + \ldots + \int_E f_k(P)dP + \ldots = \int_E f(P)dP. \quad (5.5.8)$$

If E *has finite outer measure, then bounded convergence of* (5.5.7) *is sufficient.*

We conclude this paragraph by the so-called *first mean-value theorem* of the integral calculus.

THEOREM 81. *Suppose that* f(P) *and* g(P) *are* L-*integrable over* E, *and that* f(P)≥0 *and* g(P) *is bounded in* E. *Then*

$$\int_E fg dP = \lambda \int_E f dP, \quad \ldots \quad (5.5.9)$$

where λ *is a certain mean-value between the lower and upper bounds of* g(P) *in* E.

PROOF. Let $m = \inf g$ and $M = \sup g$. Then $mf \leq fg \leq Mf$, since $f \geq 0$. Hence, by (5.2.5), the integral of fg is between m times, and M times, the integral of f.

Exercise 32. THEOREM 82. *Suppose that* E *is bounded and measurable. Suppose also that*

(i) f(P,t) *is* L-*integrable over* E, *for every fixed value of the parameter* t, *where* a<t<b.

(ii) *the partial derivative* $\frac{\partial}{\partial t}$ f(P,t) *exists and is uniformly bounded for all* P *in* E *and all* t *in* (a,b).

Then

$$\frac{d}{dt}\int_E f(P,t)dP = \int_E \frac{\partial}{\partial t} f(P,t)dP \quad \ldots \quad (5.5.10)$$

In particular, the condition (ii) is satisfied when E *is bounded and closed, and when* $\frac{\partial}{\partial t}$ f(P,t), *as function of* P *and* t, *is continuous for* P *in* E *and* t *in* (a,b).

† If the f_k are non-negative, then (5.5.8) holds whenever the f_k have L-integrals.

5.6. Lebesgue's sums. As in the case of the R-integral, the theory of the Lebesgue integral can be based, purely analytically, on certain "division sums" which correspond to Riemann's sums (§ 4.4).† Like these, they give an easy access to the fundamental addition formula (Theorem 79). The main difference is that we now have "divisions" of the linear set of the values y of the integrated function, while Riemann uses divisions of the set E of integration.

First, we consider a non-negative function $y=f(P)$, and a division

$$(D) \qquad 0=y_0<y_1<\ldots<y_i<\ldots, \quad y_i\uparrow\infty$$

of the values y of f. The number

$$\Lambda(D)=\sup\,(y_{i+1}-y_i) \qquad (5.6.1)$$

is called the *maximal length* of the division.

THEOREM 83. *Suppose that the non-negative function* f(P) *has an* L-*integral over* E; *that* f *is finite p.p.; and that the set* [f > 0] *has finite measure.*‡ *Then*

$$\sum_0^\infty (y_{i+1}-y_i)\mu[f>\eta_i]\to\int_E fdP\,; \qquad (5.6.2)$$

and, if f *is integrable, then*

$$\sum_0^\infty \eta_i\mu[y_i<f\le y_{i+1}]\to\int_E fdP, \qquad (5.6.3)$$

where $y_i\le\eta_i\le y_{i+1}$. *The sums are extended over divisions* (D) *such that* $\Lambda(D)\to 0$.

If E *has finite measure, then we may replace the sets in the above sums by* $[f\ge\eta_i]$ *and* $[y_i\le f<y_{i+1}]$, *respectively.*

† *Lebesgue* himself gives both the geometrical and the analytical definition of his integral. He uses, however, mainly the analytical method in the detailed development of the theory. The geometrical method was first used systematically by *C. Carathéodory*.

‡ If f has an integral, then $[f>0]$ is measurable; if f is integrable, it is finite p.p.

PROOF. (i) We may assume that f is finite throughout E. With the notations of § 5.5, we have

$$Y_0 = \sum_0^\infty Y_{y_i, y_{i+1}}, \quad m(Y) = m(Y_0) = \sum_0^\infty m(Y_{y_i, y_{i+1}}), \quad \text{(a)}$$

the latter by Theorem 43. Now, all the sets $[f > \eta_i]$ are measurable, by Theorem 76. Hence, applying (5.5.3) to each $Y_{y_i, y_{i+1}}$ and summing, we obtain

$$\sum_0^\infty (y_{i+1} - y_i)\mu[f > y_{i+1}] \le m(Y) \le \sum_0^\infty (y_{i+1} - y_i)\mu[f > y_i]. \quad \text{(b)}$$

The set $[f > y_i]$ decreases to the empty set as $y_i \uparrow \infty$, since f is finite, so that $\mu[f > y_i] \downarrow 0$, by Theorem 45, since $\mu[f > y_i] < \infty$. Hence the difference of the two sides of (b) is

$$\sum_0^\infty (y_{i+1} - y_i)(\mu[f > y_i] - \mu[f > y_{i+1}])$$
$$\le \Lambda(D) \sum_0^\infty (\mu[f > y_i] - \mu[f > y_{i+1}]) = \Lambda(D)\mu[f > 0], \quad . \quad \text{(c)}$$

and thus tends to zero as $\Lambda(D) \to 0$. This proves (5.6.2) for $\eta_i = y_i$ and $\eta_i = y_{i+1}$, and so, clearly, also for a general η_i.

On using the footnote on p. 111, we can replace the sets $[f > \eta_i]$ by $[f \ge \eta_i]$, provided that $E = [f \ge 0]$ has finite measure.

(ii) By (5.5.3),

$$\tfrac{1}{2}a\mu(E_a) \le m(Y_{\frac{1}{2}a, a}) = m(Y_{\frac{1}{2}a}) - m(Y_a),$$

since f is integrable and so $m(Y_a) < \infty$.

Now Y_a decreases to the empty set as $a \uparrow \infty$. Hence $m(Y_a) \downarrow 0$, by Theorem 45. It follows that

$$a\mu(E_a) \to 0 \quad . \quad . \quad . \quad . \quad (5.6.4)$$

as $a \uparrow \infty$.

Since $y_0 = 0$, we have

$$\sum_0^K (y_{i+1} - y_i)\mu[f > y_{i+1}] = \sum_0^K y_i(\mu[f > y_i] - \mu[f > y_{i+1}])$$
$$+ y_{K+1}\mu[f > y_{K+1}]. \quad . \quad . \quad . \quad \text{(d)}$$

By (5.6.4), the last term tends to zero as $K \to \infty$. Hence

$$\sum_0^\infty (y_{i+1} - y_i)\mu[f > y_{i+1}] = \sum_0^\infty y_i \mu[y_i < f \le y_{i+1}]. \qquad \text{(e)}$$

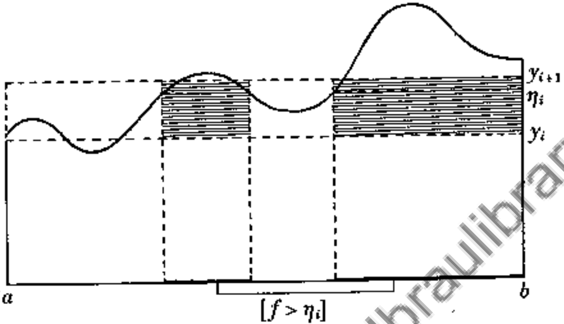

FIG. 10. Lebesgue division (first kind)

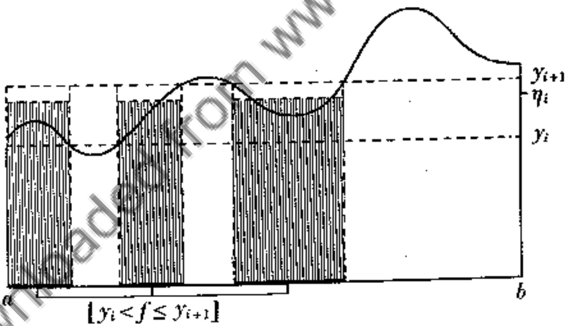

FIG. 11. Lebesgue division (second kind)

This proves (5.6.3) with $\eta_i = y_i$. The proof for $\eta_i = y_{i+1}$, and so for a general η_i, is similar.

Using (5.6.2) with the sets $[f \ge \eta_i]$, we obtain (5.6.3) with the sets $[y_i \le f < y_{i+1}]$.

The sums (5.6.2) and (5.6.3) are known as *Lebesgue's sums of the first and the second kind*, respectively.

In the sums of the first kind the ordinate set Y (or rather Y_0) is split, according to the division (D), into strips $Y_{y_i, y_{i+1}}$ "parallel" to the space E of the variables of integration. The volume of each such strip is approximated to by the product of its height $y_{i+1}-y_i$ by a mean width $\mu[f>\eta_i]$.

In the sums of the second type the set E (or rather E_0) is split into the subsets $[y_i<f\leq y_{i+1}]$, and the ordinate set is split into parts accordingly. The volume of each such part is approximated to by the product of its width $\mu[y_i<f\leq y_{i+1}]$ by a mean height η_i. Thus the sums of the second type are very similar to Riemann's sums (compare Figs. 10 and 11, for a function $y=f(x)$).

Exercise 33. Suppose that $f(\geq 0)$ has a finite upper L-integral over E. Show that

$$a\bar{\mu}(E_a)\to 0 \quad . \quad . \quad . \quad . \quad (5.6.5)$$

when either $a\to 0$ or $a\to\infty$.

Exercise 34. Suppose that $f(\geq 0)$ has a finite upper L-integral over E, and that $\bar{\mu}[f>0]<\infty$. Show that

$$\left.\begin{aligned}&\sum_0^\infty(y_{i+1}-y_i)\underline{\mu}[f>\eta_i]\to\underline{\int_E} f\,dP,\\&\sum_0^\infty(y_{i+1}-y_i)\bar{\mu}[f>\eta_i]\to\bar{\int_E} f\,dP,\end{aligned}\right\} \quad . \quad . \quad (5.6.6)$$

when $\Lambda(D)\to 0$.

Deduce that, if $f(\geq 0)$ and E are bounded, then

$$\underline{\int_E} f\,dP=\int_{R\,0}^{\infty}\underline{\mu}(E_y)\,dy,\quad \bar{\int_E} f\,dP=\int_{R\,0}^{\infty}\bar{\mu}(E_y)\,dy, \quad . \quad (5.6.7)$$

where the subscript R indicates R-integrals.

The formula (5.6.3) can be extended to general functions. We consider a division

$$(D)\ \ldots<y_{-i}<\ldots<y_{-1}<0=y_0<y_1<\ldots<y_i<\ldots$$

where $y_i\downarrow-\infty$ as $i\downarrow-\infty$, and $y_i\uparrow\infty$ as $i\uparrow\infty$. The *maximal length* $\Lambda(D)$ of (D) is again defined as $\sup\,(y_{i+1}-y_i)$.

THEOREM 84. *If* f *is* L-*integrable over a set* E *of finite measure, then*

$$\sum_{-\infty}^{\infty} \eta_i \mu[y_i < f \leq y_{i+1}] \to \int_E f dP \quad . \quad . \quad (5.6.8)$$

whenever $\Lambda(D) \to 0$.

PROOF. Since E, and f in it, are measurable, all the occurring sets are measurable. Also the terms with $i \geq 0$ involve only the function f_+, the remaining terms only f_-. The result now follows on applying (5.6.3) to f_+ and $-f_-$. For the latter function we employ as division the points $|y_{-l}| < |y_{-(l+1)}|$, where $l+1 = -i > 0$; and we use the sets $[|y_{-l}| \leq -f_- < |y_{-(l+1)}|] = [y_i < f \leq y_{i+1}]$, $i < 0$.

The formula (5.6.8) is often used as an analytical definition of the L-integral. It has the advantage that it applies to general functions at once.

5.7. Fubini's theorem. Let $\mathsf{E} = \mathsf{E}_n$ be the n-dimensional space, and let $1 \leq p < n$. We denote by E_p any p-dimensional subspace of E obtained by equating to zero certain $n-p$ of the variables; and E_{n-p} is the "supplementary" space of $n-p$ dimensions, obtained by equating to zero the other p variables. We also write $P = (x_1, x_2, \ldots x_n) = (R, S)$, where R is a point of E_p and S is the "complementary" point of E_{n-p}. There are $\binom{n}{p}$ possible choices of E_p, when p is given. Thus we may choose as E_p the space of all points $R = (x_1, x_2, \ldots x_p, 0, \ldots 0)$, and as E_{n-p} the space of all points $S = (0, 0, \ldots 0, x_{p+1}, \ldots x_n)$.

Now let $f(P)$ be L-integrable over a set E in E. By extending the function we may always assume that E is the whole space E. The importance of the decomposition formula

$$\int_{\mathsf{E}} f(P) dP = \int_{\mathsf{E}_p} h(R) dP \quad \text{where} \quad h(R) = \int_{\mathsf{E}_{n-p}} f(R, S) dS, \; (5.7.1)$$

for the actual integration of a function of n variables is familiar. It reduces the integration to two successive integrations of functions of less than n variables. For fixed p, there are $\binom{n}{p}$ different forms of (5.7.1). On choosing $p=n-1$ and repeating the process, the integration is reduced to n successive integrations of functions of one variable only.

Let E be a set in E. The set of all points R of E_p, for which some $P=(R, S)$ belongs to E, may be called the "*projection*" of E on E_p; we denote it by $E_{(p)}$. Similarly, $E_{(n-p)}$ is the projection of E on the supplementary space E_{n-p}.

Conversely, $E_{n-p}(R)$ denotes, for fixed R of E_p, the set of all points $P=(R, S)$ which belong to E. It is a "*section* of E through R" and is "parallel" to E_{n-p}. If R is outside $E_{(p)}$, then this section is empty.

If, for instance, $n=3$ and $p=1$, we have plane sections parallel to a certain co-ordinate plane; if $p=2$, the sections are linear and parallel to a certain co-ordinate axis.

Now consider the characteristic function $\chi(P)$ of a measurable set E. Then $h(R)$, if it exists, is the $(n-p)$-dimensional measure $\mu(R)$ of the section $E_{n-p}(R)$; and (5.7.1) becomes

$$m(E)=\int_{E_{(p)}} \mu(R)dR. \quad . \quad . \quad . \quad (5.7.2)$$

That is to say: the n-dimensional measure of E is obtained by integrating, over the projection $E_{(p)}$ of E on E_p, the $(n-p)$-dimensional measures $\mu(R)$ of the sections parallel to E_{n-p}. In this form the decomposition formula appears geometrically "evident". However, we should observe that neither the existence of $\mu(R)$ nor that of its integral over $E_{(p)}$ is obvious.

The following theorems are due to *G. Fubini* (1907). It will be convenient to prove them for functions $f(x, y)$ of two variables only. The proof extends in an obvious way to the general case. The formula (5.7.1) takes the form

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = \int_{-\infty}^{\infty} h(x)dx = \int_{-\infty}^{\infty} g(y)dy, \quad . \quad (5.7.3)$$

where

$$h(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad g(y) = \int_{-\infty}^{\infty} f(x,y)dx. \quad . \quad (5.7.4)$$

We shall also denote the linear subspace of the x by E_x, and that of the y by E_y. The sections are linear : E_x is a section of E, through a fixed x, parallel to the y-axis ; and E_y is a section through y parallel to the x-axis.

First, we turn to (5.7.2).

THEOREM 85. *Let* E *be a measurable set. Then almost all sections* $\mathrm{E}_{n-p}(\mathrm{R})$ *are measurable. Their* (n − p)-*dimensional measure* $\mu(\mathrm{R})$ *has an* L-*integral over* $\mathrm{E}_{(p)}$, *and* (5.7.2) *holds.*

PROOF. We take E as a measurable set in the (x, y)-plane. We want to prove that almost all sections E_x and E_y have linear measures $\mu(x)$ and $\nu(y)$, respectively ; and that

$$m(E) = \int_{-\infty}^{\infty} \mu(x)dx = \int_{-\infty}^{\infty} \nu(y)dy. \quad . \quad . \quad (5.7.5)$$

It suffices to consider the sections E_x.

(i) If I is the interval $\langle a \leq x \leq b, \alpha \leq y \leq \beta \rangle$, then $\mu(x) = \beta - \alpha$ when x is in $\langle a, b \rangle$; and $\mu(x) = 0$ otherwise. Also

$$m(I) = (b-a)(\beta-a) = \int_a^b (\beta - a)dx = \int_{-\infty}^{\infty} \mu(x)dx.$$

The proof is similarly elementary in the case of an interval sum S or an open interval (I).

(ii) If O is an open set, then all O_x are (possibly empty) linear open sets, so that $\mu(x)$ exists for all x.

Now $O = \sum_1^{\infty} J_k$, where the J_k are mutually separate intervals, and $O_x = \sum_1^{\infty} J_{kx}$; if $S_p = \sum_1^{p} J_k$, then $S_{px} = \sum_1^{p} J_{kx}$. Hence

$\mu_p(x) \uparrow \mu(x)$ as $p \uparrow \infty$, where $\mu_p(x)$ is the linear measure of S_{p_x}. It follows, by (i) and Theorem 70, that

$$m(O) = \lim m(S_p) = \lim \int_{-\infty}^{\infty} \mu_p(x)dx = \int_{-\infty}^{\infty} \mu(x)dx.$$

(iii) If F is a bounded closed set, then all F_x are (possibly empty) linear bounded closed sets, so that $\mu(x)$ exists. Let $F \subset (I)$ where (I) is an open interval. Then

$$(I) - F = O, \quad (I)_x - F_x = O_x,$$

where O and O_x are open. Hence, by (i) and (ii),

$$m(F) = m((I)) - m(O) = \int_{-\infty}^{\infty} [\mu((I)_x) - \mu(O_x)]dx = \int_{-\infty}^{\infty} \mu(x)dx.$$

(iv) Let E be a bounded measurable set. By (3.4.7) and (3.5.10), there exists an ascending sequence of bounded closed sets F_n, and a descending sequence of open sets O_n, such that $m(F_n) \uparrow m(E)$ and $m(O_n) \downarrow m(E)$. If $\mu(F_{n_x}) = \mu_n(x)$, $\mu(O_{n_x}) = \mu_n^*(x)$, then, by (ii), (iii), and Theorem 70,

$$\left.\begin{aligned} m(E) &= \lim \int_{-\infty}^{\infty} \mu_n(x)dx = \int_{-\infty}^{\infty} \lim \mu_n(x)dx, \\ m(E) &= \lim \int_{-\infty}^{\infty} \mu_n^*(x)dx = \int_{-\infty}^{\infty} \lim \mu_n^*(x)dx. \end{aligned}\right\} \quad \text{(a)}$$

Hence

$$\int_{-\infty}^{\infty} [\lim \mu_n^*(x) - \lim \mu_n(x)]dx = 0. \quad . \quad . \quad \text{(b)}$$

Now $F_{n_x} \subset E_x \subset O_{n_x}$ and hence $\mu_n^*(x) \geq \mu_n(x)$, so that the integrand in (b) is non-negative. By Theorem 62, this integrand is zero p.p. This implies that E_x is measurable p.p., and that $\mu(x)$ equals p.p. either of the two limits. Substituting in (a), we obtain

$$m(E) = \int_{-\infty}^{\infty} \mu(x)dx.$$

(v) If E is unbounded, we consider, as usual, the sequence of bounded measurable sets $E^{(k)}$ of (2.5.2). Here $E^{(k)} \uparrow E$, $E_x^{(k)} \uparrow E_x$, and $m(E^{(k)}) \uparrow m(E)$. For every fixed k the linear measure $\mu_k(x)$ of $E_x^{(k)}$ exists p.p. The sum of the exceptional sets, for all k, has measure zero, and so $\mu_k(x)$ exists p.p., simultaneously for all k. It follows, by Theorem 45, that $\mu(x)$ exists and that $\mu_k(x) \uparrow \mu(x)$, p.p. Finally, applying (5.7.5) to $E^{(k)}$ and letting $k \uparrow \infty$, the proof is completed, by Theorem 70.

COROLLARY. *Suppose that two measurable* n-*dimensional sets* E *and* E* *have the same* p-*dimensional projection* $E_{(p)} = E_{(p)}^*$; *and that their supplementary sections* $E_{n-p}(R)$ *and* $E_{n-p}^*(R)$ *have the same* (n − p)-*dimensional measure* $\mu(R)$ *almost everywhere. Then* $m(E) = m(E^*)$.

This corollary supplies a new proof for the addition formula (5.5.6). In fact, if f and g are non-negative, and if Y_1, Y_2, Y_3 are the ordinate sets of f, g, and $f+g$, respectively, then the corollary is applicable to $Y_3 - Y_1$ and Y_2 (see Fig. 8). It follows that $m(Y_3 - Y_1) = m(Y_2)$, or $m(Y_3) = m(Y_1) + m(Y_3 - Y_1)$ $= m(Y_1) + m(Y_2)$: the desired formula. The case of general functions is then easily dealt with.†

It should, however, be noted that (5.5.6) was used above in (b); but only in order to conclude that $\mu(x)$ exists p.p. In our case all the sections required are linear y-intervals $\langle f < y \leq f+g \rangle$ and $\langle 0 \leq y \leq g \rangle$, which have certainly measure (length).

We turn now to *Fubini's* general theorem.

THEOREM 86. *Suppose that* f(P) *is* L-*integrable over the space* **E**.‡ *Then* h(R), *in* (5.7.1), *exists for almost all* R *and is integrable over* E_p; *and* (5.7.1) *holds*.

If f *is non-negative and has an* L-*integral, then* h(R) *exists p.p. and has an* L-*integral*; *and* (5.7.1) *holds*.

PROOF. We prove the theorem for a non-negative *f(x,y)*. In the general case, $f = f_+ - (-f_-)$, and we apply (5.5.6).

† As in the proof of Theorem 79 (iii).
‡ If f is integrable over E, then f^* is integrable over E.

(i) First, suppose that f takes only a finite number of values in the plane. Let E_k be the set where $f \equiv c_k$, say; and let $1 \leq k \leq m$. The sets E_k are measurable, by (5.4.1). Also

$$f = c_1\chi_1 + c_2\chi_2 + \dots + c_m\chi_m$$

where χ_k is the characteristic function of E_k. Hence, by Theorem 85 and (5.5.6),

$$\sum_1^m c_k h_k(x) = \sum_1^m c_k \int_{-\infty}^{\infty} \chi_k(x,y)dy = \int_{-\infty}^{\infty} f(x,y)dy = h(x)$$

exists for almost all x. Similarly,

$$\iint_{E_2} f dxdy = \sum_1^m c_k \iint_{E_2} \chi_k dxdy = \sum_1^m c_k \int_{-\infty}^{\infty} h_k(x)dx = \int_{-\infty}^{\infty} h(x)dx,$$

which proves (5.7.3) in this case.

(ii) Consider the functions

$$f_p = \begin{cases} \dfrac{i}{p} & \text{in } \left[\dfrac{i}{p} \leq f < \dfrac{i+1}{p}\right] \\ p & \text{in } [p \leq f], \end{cases}$$

where $0 \leq i < p^2$. Now, f_p is measurable and takes only a finite number of values. Hence the corresponding function $h_p(x)$ exists p.p. and has an integral over $(-\infty, \infty)$; and

$$\iint_{E_2} f_p dxdy = \int_{-\infty}^{\infty} h_p(x)dx, \quad . \quad . \quad (a)$$

by what we have proved in (i). Also $f_p \uparrow f$ as $p \uparrow \infty$. Hence $h_p(x) \uparrow h(x)$ p.p., by Theorem 70, so that $h(x)$ exists p.p. Finally,

$$\int_{-\infty}^{\infty} h_p(x)dx \uparrow \int_{-\infty}^{\infty} h(x)dx, \quad \iint_{E_2} f_p dxdy \uparrow \iint_{E_2} f dxdy;$$

and we obtain (5.7.3) from (a).

Exercise 35. Let $f(x,y) = (y^2 - x^2)[x^2 + y^2]^{-2}$ when $x^2 + y^2 \neq 0$. Show that

$$\int_0^1 \left(\int_0^1 f(x,y)dy \right) dx \neq \int_0^1 \left(\int_0^1 f(x,y)dx \right) dy.$$

Show also that $f(x,y)$ is measurable in the square $\langle 0 \leq x \leq 1\,;\ 0 \leq y \leq 1 \rangle$. Why does (5.7.3) not hold?

5.8. Relation between *R*-integral and *L*-integral. If the ordinate set of a non-negative function has content, then, by Theorem 40, it has also measure, and the two are equal. Applying this to f_+ and $-f_-$, we obtain at once the following result.

THEOREM 87. *If* f *has an* R-*integral over* E, *then it has also an* L-*integral, and the two integrals have the same value.*

The converse of this is false: the characteristic function of the rational set R, for instance, is L-integrable but not R-integrable. The Lebesgue integral is, therefore, more comprehensive than the classical integral of Riemann.

In Theorem 77 we have a complete characterisation of the class of functions f which are L-integrable over a given set E: in particular, all such functions are measurable in E. Hence, if f is R-integrable, it must certainly be measurable in E, by Theorem 87. A complete characterisation is given by the following theorem.

THEOREM 88. *A non-negative function* f(P) *has an* R-*integral over* E *if, and only if, its extension* f*(P) *is continuous p.p. in* E.†

The same property is necessary and sufficient for a general function to be R-*integrable over* E, *provided that the* R-*integral of* |f|, *which then exists, is finite.*

PROOF. We may assume that E is the whole space E: we need only consider f^*. We may also assume that $f \geq 0$. For, if f is continuous p.p., then so are f_+ and $-f_-$; and conversely.

(i) We assume, first, that f is bounded. Let J be a closed interval containing P as interior point; and let

† Such a function is measurable (Exercise 30).

$$m_J(P)=\inf f, \quad M_J(P)=\sup f,$$

where the bounds are taken with respect to all points Q of J. If J descends to P, through some sequence, then $m_J(P) \uparrow \underline{f}(P)$ and $M_J(P) \downarrow \bar{f}(P)$, say. These functions, clearly, do not depend on the choice of the sequence employed. Also $\underline{f}(P) \leq f(P) \leq \bar{f}(P)$; and $\underline{f}(P)=\bar{f}(P)$ if, and only if, f is continuous at P.

Next, let I be a closed interval, and let D be a Riemann division of it. The sum (4.4.7) is (with the notations there)

$$\Delta(D)=\sum_{1}^{m}(M_k-m_k)\mid J_k \mid=\sum_{1}^{m}\int_{J_k}(M_k-m_k)dP=\int_I g_D dP,$$

where $g_D=M_k-m_k$ inside J_k: the integrals are L-integrals, say, and we need no definition of g_D on the frontiers of the J_k since these have measure zero.

Now consider a given sequence of divisions D with $L(D) \to 0$. For any point P which is not on the frontier of one of the enumerably many intervals involved, we have, plainly, $g_D \to \bar{f}-\underline{f}$. The exceptional set is of measure zero. Also the convergence is bounded since f is bounded. Hence $\bar{f}-\underline{f}$ is L-integrable, by Theorem 71; and

$$\Delta(D)=\int_I g_D dP \to \int_I (\bar{f}-\underline{f})dP,$$

as $L(D) \to 0$.

By Theorem 57, the bounded function f is R-integrable over I if, and only if, $\Delta(D) \to 0$. This is equivalent to the integral of $\bar{f}-\underline{f}$ being zero; and this, in turn, to $\bar{f}=\underline{f}$ p.p. in I, by Theorem 62; or to f being continuous p.p. in I. This proves the theorem in the case of a function bounded in an interval I.

(ii) If f is unbounded in I, we consider the functions $f_{[n]}$ of (4.2.4). If f is continuous p.p. [continuity at a point P, where $f(P)=\infty$, having the obvious meaning], then, clearly, each $f_{[n]}$ is continuous p.p. and hence R-integrable, by (i). This implies the existence of the R-integral of f.

Conversely, if this integral exists, then all $f_{[n]}$ are R-integrable and, thus, $\bar{f}_{[n]} = \underline{f}_{[n]}$ p.p., by (i). Also $\bar{f}_{[n]} \uparrow \bar{f}$ and $\underline{f}_{[n]} \uparrow \underline{f}$. Hence, by Fatou's Lemma (5.3.1.),

$$0 \leq \int_I (\bar{f} - \underline{f}) dP \leq \underline{\lim} \int_I (\bar{f}_{[n]} - \underline{f}_{[n]}) dP = 0,$$

so that $\bar{f} = \underline{f}$ p.p., and f is continuous p.p.

(iii) In the general case the property of f having an R-integral over E is equivalent to having it over every interval I; and the property of f being continuous p.p. in E is equivalent to it being continuous p.p. in every interval I. This completes the proof.

Theorem 88 permits to decide whether a theorem, proved for the L-integral, remains valid for the R-integral. Take Theorem 78. If f and g are R-integrable (and hence are also L-integrable), then f^* and g^* are continuous p.p. Under the conditions of the theorem, the same will, plainly, hold for $f+g$, fg, f/g, and for their absolute values. The R-integrals of the latter will, therefore, exist; and they will be equal to the corresponding L-integrals, which, by Theorem 78, are finite. Hence this theorem holds also for the R-integral.

It now follows that the addition formulae (5.5.6) also hold for the R-integrals. We have thus obtained a new proof for Theorem 59, this time in its general form.

Next, consider Theorem 71. We cannot conclude, if $f_k \to f$ in E and if the f_k^* are continuous p.p., that then f^* is continuous p.p. (see the first example after (4.5.1)).

Hence the analogue of Theorem 71 *for the* R*-integral will only hold under the additional assumption that the limit function* f *be* R*-integrable over* E.

With this restriction it was first proved, directly, by *C. Arzelà* (1885) who also assumed *bounded* convergence $f_k \to f$ in an interval I. The restriction that one has to prove the integrability of f separately, makes this test often impracticable. For the Lebesgue integral this difficulty disappears.

Solutions to Exercises

Ex. 24. Proof as for Ex. 21.

Ex. 25. Bounded convergence over a set of infinite measure is not dominated.

Ex. 26. If $a \geq 0$, then $[f^* > a] = [f > a]$; if $a < 0$, then $[f^* > a] = [f > a] + {}_cE$; and ${}_cE$ is measurable if E is measurable.

Ex. 27. The set $[a < f < b] = [f > a] - [f \geq b]$ for instance, is measurable, since $[f > a]$ and $[f \geq b]$ are measurable.

Ex. 28. (i) If $b \downarrow a$, then $[f \geq b] \uparrow [f > a]$. Hence $[f > a]$ is measurable. In the cases (ii) and (iii) consider the complementary sets. In case (iv) take a sequence $r_k \downarrow a$, when $[f > r_k] \uparrow [f > a]$.

Ex. 29. If $a > 0$, then $[f^* \geq a] = [f \geq a]$ is closed; if $a \leq 0$, then $[f^* \geq a] = {}_cF + [f \geq a]$.

Ex. 30. The set $E_a = [f > a]$ consists of points P where f is continuous, and of a set Z of measure zero. If f is continuous at P, then with P there belongs a neighbourhood $K(P)$ to E_a. The sum of all these $K(P)$ is an open set O, and $E_a = O + Z$. Hence E_a is measurable.

Ex. 31. The functions $\underline{\Phi} = \underline{\lim}\, f_k$ and $\bar{\Phi} = \overline{\lim}\, f_k$ are measurable, by Theorem 74. The set $C^* = [\underline{\Phi} = \bar{\Phi}]$ is measurable, by (5.4.4); and $D = [\,|\bar{\Phi}| = \infty\,]$ is measurable, by (5.4.1). Hence $C = C^* - C^* \,.\, D$ is measurable.

Ex. 32. For P in E, and t and $t + h$ in (a,b), we have

$$f(P, t+h) - f(P, t) = h \frac{\partial}{\partial t} f(P,t) \Big|_{t=\tau},$$

where τ is a mean-value between t and $t + h$. Hence

$$\frac{f(P, t+h) - f(P, t)}{h} \to \frac{\partial}{\partial t} f(P, t)$$

boundedly in E as $h \to 0$. It follows, by Theorem 71, that

$$\frac{1}{h}\left[\int_E f(P, t+h)dP - \int_E f(P, t)dP\right] = \int_E \frac{f(P, t+h) - f(P, t)}{h}dP$$

$$\rightarrow \int_E \frac{\partial}{\partial t}f(P, t)dP.$$

Ex. 33. By (5.5.3), $a\bar{\mu}(E_a) \leq \bar{m}(Y_{0,a}) = \bar{m}(Y_0) - \bar{m}(Y_a)$, since $Y_{0,a}$ and Y_a are separated by $y=a$.

If $a \downarrow 0$, then $Y_a \uparrow Y_0$ and $\bar{m}(Y_a) \uparrow \bar{m}(Y_0)$, by (3.8.7). Hence $a\bar{\mu}(E_a) \rightarrow 0$. If $a \uparrow \infty$, the proof is similar to that of (5.6.4) [$\mu(E_\infty) = 0$, if the upper integral of f is finite].

Ex. 34. The proof of (5.6.6) is based on (5.5.3) and (5.5.4), and corresponds to that of (5.6.2).

If f is bounded, then the sets $[f > \eta_i]$ are empty for large y_i, and the sums (5.6.6) become Riemann sums for the functions $\underline{\mu}(E_y)$ and $\bar{\mu}(E_y)$, respectively. Formulae (5.6.7) now follow from (4.4.12).

Ex. 35. First, $f(x,y) = \dfrac{\partial^2}{\partial x \partial y}$ arc tan y/x. Elementary integration shows that the value of the first integral is $-\pi/4$; and that of the second is $\pi/4$.

Also f is continuous in $(0 < x < 1;\ 0 < y < 1)$ and, hence, is measurable in its closure [the frontier has measure zero]. The integral of $|f|$ is infinite, so that f is not integrable over the square and Fubini's theorem is not applicable.

CHAPTER VI

INTEGRATION AND DIFFERENTIATION

6.1. The problem. In this chapter we consider functions $y=f(x)$ of one variable only. Let us suppose, first, that $f(x)$ is continuous in a closed interval $I=\langle a, b\rangle$. The function

$$F(x)=\int_a^x f(t)dt \qquad . \quad . \quad . \quad (6.1.1)$$

is then continuous and differentiable in I; and

$$F'(x)=f(x), \qquad . \quad . \quad . \quad . \quad (6.1.2)$$

the derivative being "one-sided" at the end-points of I. All this is familiar. In fact, on this fundamental property depends almost all actual integration.

Any solution of the differential equation $y'=f(x)$ is called an *indefinite integral*, or a *primitive function*, of $f(x)$. The above $F(x)$ is such a primitive function; and *any* primitive function of f is of the form $\Phi(x)=F(x)+c$, where c is the so-called *constant of integration*.

For, if $g=\Phi-F$, then $g'\equiv 0$. Hence $g(x_2)-g(x_1)=(x_2-x_1)g'(\xi)=0$, where ξ is a mean-value between x_1 and x_2: g is constant.

It follows that

$$F(x)=\int_a^x f(t)dt=\Phi(x)-\Phi(a), \qquad . \quad . \quad (6.1.3)$$

where $\Phi(x)$ is an *arbitrary* indefinite integral.

It will be convenient to call the special integral $F(x)$ *the* indefinite integral of f.

Conversely, if $f'(x)$ is continuous in I, then $f(x)$ is a primitive function of $f'(x)$; and we have, by (6.1.3),

$$f(x)-f(a)=\int_a^x f'(t)dt. \quad . \quad . \quad . \quad (6.1.4)$$

The two formulae (6.1.2) and (6.1.4) express the familiar fact that differentiation and integration are operations inverse to each other.

The main problem which we shall discuss in this chapter is, how far the fundamental formulae (6.1.2) and (6.1.4) remain valid for general (not necessarily continuous) integrable functions f and f', respectively.

6.2. Elementary properties of the indefinite integral. We suppose, from now on, that $f(x)$ is L-integrable over the interval $I=\langle a, b\rangle$.† The indefinite integral $F(x)$, which is defined by (6.1.1), exists throughout I, by Theorem 66. Also $F(a)=0$.

First, $F(x)$ *is continuous in* I. For,†

$$F(x+h)-F(x)=\int_x^{x+h} f(t)dt \to 0$$

as $h\to 0$, by Theorem 72.

Next, $F'(x_0)=f(x_0)$ *whenever* f *is continuous at* x_0.‡ For, if a positive ϵ is given, then

$$\left|\frac{F(x_0+h)-F(x_0)}{h}-f(x_0)\right|=\left|\frac{1}{h}\int_{x_0}^{x_0+h}(f(x)-f(x_0))dx\right|$$

$$\leq\frac{1}{|h|}\left|\int_{x_0}^{x_0+h}|f(x)-f(x_0)|\,dx\right|\leq\frac{1}{|h|}\left|\int_{x_0}^{x_0+h}\epsilon dx\right|=\epsilon,$$

if $|h|\leq H(\epsilon)$, since $f(x)-f(x_0)\to 0$ as $x\to x_0$.

† x and $x+h$ are in I. We write $\int_a^b=-\int_b^a$ if $a>b$ (here, in the case $h<0$).

‡ One-sided differentiation at the end-points of I.

Let $\Sigma = \Sigma J_k$ be an interval set, a (finite or infinite) sum of mutually separate closed intervals $J_k = \langle a_k, b_k \rangle$. If $\Sigma = I$, then

$$\Sigma | F(b_k) - F(a_k) | = \Sigma | \int_{a_k}^{b_k} f dx | \leq \int_a^b | f | dx < \infty. \quad (6.2.1)$$

If $\Sigma \subset I$, then, by Theorem 72, for given $\epsilon (>0)$,

$$\Sigma | F(b_k) - F(a_k) | \leq \int_\Sigma | f | dx < \epsilon, \quad . \quad (6.2.2)$$

whenever $\mu(\Sigma) = \Sigma(b_k - a_k) < \delta(\epsilon)$.

Finally, we note that

$$F(x) = \int_a^x f_+ dx - \int_a^x (-f_-) dx = F^+(x) - F^-(x), \quad . \quad (6.2.3)$$

so that $F(x)$ appears as difference of two increasing (non-decreasing †) non-negative integrals.

6.3. The second mean-value theorem. The following result is known as the second mean-value theorem of the integral calculus.

THEOREM 89. *Suppose that* g(x) *is monotone and bounded in the interval* I = ⟨a, b⟩ ; *and that* f(x) *is* L-*integrable over* I. *Then there exists a* ξ *in* I *such that*

$$\int_a^b fg dx = g(a) \int_a^\xi f dx + g(b) \int_\xi^b f dx. \quad . \quad . \quad (6.3.1)$$

PROOF. We may assume that g decreases in I. Then g is R-integrable [§ 4.4] ; and, *a fortiori*, it is L-integrable, by Theorem 87. Again, fg is L-integrable, by Theorem 78.

(i) First, suppose that $g \geq 0$ and $f \geq -c$ in I ; and that g decreases to $g(b) = 0$.

We divide I into m equal parts by the points

$$x_k = a + (b-a)km^{-1}, \quad 0 \leq k \leq m.$$

† We use throughout the word increasing (and, similarly, decreasing) in this sense.

We then have

$$\int_a^b (f+c)g\,dx = \sum_1^m \int_{x_{k-1}}^{x_k} (f+c)g\,dx \leq \sum_1^m g(x_{k-1}) \int_{x_{k-1}}^{x_k} (f+c)\,dx$$

$$= \sum_1^m g(x_{k-1}) \int_{x_{k-1}}^{x_k} f\,dx + c\sum_1^m (x_k - x_{k-1}) g(x_{k-1}). \quad . \qquad \text{(a)}$$

On summing by parts, the first sum on the right becomes

$$\sum_1^m g(x_{k-1})(F(x_k) - F(x_{k-1})) = \sum_1^m F(x_k)(g(x_{k-1}) - g(x_k))$$

$$\leq \text{Max } F(x) \sum_1^m (g(x_{k-1}) - g(x_k)) = g(a) \text{ Max } F(x),$$

since $F(0)=0$, $g(b)=0$, and $g(x_{k-1}) \geq g(x_k)$. The last sum in (a) tends to the R- (and L-) integral of g over I as $m \to \infty$, by Theorem 58. Hence, in the limit,

$$\int_a^b (f+c)g\,dx \leq g(a) \text{ Max } F(x) + c\int_a^b g\,dx,$$

$$\int_a^b fg\,dx \leq g(a) \text{ Max } F(x). \quad . \quad . \quad . \quad . \qquad \text{(b)}$$

(ii) Now, let f be a general function, L-integrable over I; and let g decrease to zero, as before. We write, with $c > 0$,

$${}_cf(x) = \text{Max } (f(x), -c), \quad {}_cF(x) = \int_a^x {}_cf\,dx.$$

Here ${}_cf$ is measurable, by (5.4.5); and it is integrable, since $|{}_cf| \leq |f|$. By (b),

$$\int_a^b {}_cfg\,dx \leq g(a) \text{ Max } {}_cF(x). \quad . \quad . \qquad \text{(c)}$$

Now ${}_cf \downarrow f$ dominatedly as $c \uparrow \infty$, since $|{}_cf| \leq |f|$. Hence ${}_cF(x) \downarrow F(x)$, by Theorem 71, so that, given ϵ (>0),

$${}_cF(b) < F(b) + \epsilon,$$

when $c > C(\epsilon)$. Also

$$\int_x^b {}_c f dt \geq \int_x^b f dt, \quad {}_cF(b) - {}_cF(x) \geq F(b) - F(x).$$

It follows that

$${}_cF(x) \leq F(x) + ({}_cF(b) - F(b)) < \text{Max } F(x) + \epsilon,$$

and thus $\text{Max } {}_cF(x) < \text{Max } F(x) + \epsilon$, when $c > C(\epsilon)$. We obtain, therefore, (b) again from (c), as $c \to \infty$. For, ${}_cfg \downarrow gf$ dominatedly, since $| {}_cfg | \leq | fg |$.

Applying (b) to $-f$, we find that

$$\int_a^b fg dx \geq g(a) \text{ Min } F(x). \quad . \quad . \quad \text{(d)}$$

Now (b) and (d) imply

$$\int_a^b fg dx = g(a) \int_a^\xi f dx \quad . \quad . \quad . \quad (6.3.2)$$

for a suitable ξ in I, since $F(x)$ is continuous. This is (6.3.1), since we have assumed $g(b) = 0$.

Finally, the general formula (6.3.1) follows on applying (6.3.2) to $g(x) - g(b)$.

It should be noted that the formula (6.3.2) holds whenever g decreases and is non-negative. For then $g(b)$ can be changed to zero without altering the value of the integral of fg.

If f is R-integrable (and hence is L-integrable), then so is fg. The second mean-value theorem holds, therefore, also for the R-integral. A direct and simpler proof for this is, of course, possible.

6.4. Functions of bounded variation. A function $g(x)$, defined (finitely) in the interval $I = \langle a, b \rangle$, is said to be *of bounded variation* (is b.v.) in I, if

$$\Sigma | g(b_k) - g(a_k) | \leq A < \infty \quad . \quad . \quad (6.4.1)$$

for any division $I=\Sigma J_k$ of I into a finite number of mutually separate intervals $J_k=\langle a_k, b_k\rangle$; the constant A is to be the same for all divisions.

By (6.2.1), *the indefinite integral* F(x) *is b.v. in* I.

If g is b.v. in I, it is bounded there ; it is also b.v. in any subinterval. If g_1 and g_2 are b.v. in I, then g_1+g_2, g_1-g_2, and g_1g_2 are b.v. ; so is g_1/g_2 provided that $|g_2|\geq d>0$. We leave the simple proofs to the reader. For the product, for instance, the result follows from

$$|g_1(x)g_2(x)-g_1(y)g_2(y)|\leq|g_1(x)||g_2(x)-g_2(y)| \\ +|g_2(y)||g_1(x)-g_1(y)| \quad (6.4.2)$$

and from the boundedness of g_1 and g_2.

If g is monotone increasing then it is, clearly, b.v. ; so is the difference of two such functions. The converse is also true.

THEOREM 90. *If* g(x) *is of bounded variation in* I, *then it is of the form*

$$g(x)=g_1(x)-g_2(x), \quad . \quad . \quad . \quad (6.4.3)$$

where $g_1(x)$ *and* $g_2(x)$ *increase in* I.

PROOF. For every x in I, the function g is b.v. in $\langle a, x\rangle$. Now

$$g(x)-g(a)=\Sigma(g(b_k)-g(a_k)) \\ =\Sigma_+(g(b_k)-g(a_k))-\Sigma_-|g(b_k)-g(a_k)|,$$

where the sum Σ is extended over all intervals J_k of a division of $\langle a, x\rangle$, and Σ_+ and Σ_- are extended over those J_k, for which $g(b_k)-g(a_k)$ is non-negative, or negative, respectively. Clearly, if Σ_+ happens to be "large" for a division, then Σ_- must also be correspondingly large. Hence, if we put

$$P(x)=\sup\Sigma_+, \quad N(x)=\sup\Sigma_-, \quad . \quad . \quad (6.4.4)$$

where the upper bounds are taken with respect to all possible divisions of $\langle a, x\rangle$, we have in

$$g(x)=g(a)+P(x)-N(x) \quad . \quad . \quad . \quad (6.4.5)$$

a representation of the desired kind: $P(x)$ and $N(x)$ are, plainly, non-negative increasing functions of x in I. These functions are called *the positive, or negative, variation* of g in $\langle a, x\rangle$. Their sum

$$V(x) = P(x) + N(x) \quad . \quad . \quad . \quad (6.4.6)$$

is called the *total variation* of g in $\langle a, x\rangle$. The number $V(b)$ is the smallest possible value of A in (6.4.1).

It is also easy to see that $P(x)$ and $N(x)$ are continuous at each point x where g is continuous.

In the case of the indefinite integral $F(x)$, the formula (6.2.3) gives a representation (6.4.3). In fact, it is not difficult to prove that $F^+(x)$ and $F^-(x)$, in (6.2.3), are just the positive and negative variations of $F(x)$. The total variation is the indefinite integral of $|f|$.

The function $g(x)$ is said to be *absolutely continuous* (a.c.) in I if, for every given positive ϵ,

$$\Sigma|\, g(b_k) - g(a_k)\, | < \epsilon, \quad . \quad . \quad . \quad (6.4.7)$$

whenever the mutually separate intervals $J_k = \langle a_k, b_k\rangle$ are contained in I and their total length $\Sigma(b_k - a_k) < \delta(\epsilon)$.

If g is a.c. in I, then it is, *a fortiori*, b.v. in I. It is also continuous in I; and it is a.c. in any subinterval. All this is easy to prove. Also the positive and negative variations (6.4.4) of an a.c. function are themselves a.c.

If g_1 and g_2 are a.c., then $g_1 + g_2$, $g_1 - g_2$, and $g_1 g_2$ are a.c.; the quotient g_1/g_2 is a.c. if $|\, g_2\,| \geq d > 0$. The proofs are again elementary.

A continuous function need not be b.v. (and thus not a.c.), as the example of $g(x) = x \sin 1/x$ $(x \neq 0)$, $g(0) = 0$ (in the interval $\langle -1, 1\rangle$) shows. There exist also continuous monotone (and hence b.v.) functions which are not a.c. (see § 6.6).

We now restate (6.2.2):

Theorem 91. *The indefinite integral* F(x) *of a function* f(x), L-*integrable over* I, *is absolutely continuous in* I.

We also note that, in (6.2.3), $F(x)$ is represented as the difference of its two variations which are a.c.

Suppose again that $g(x)$ is b.v. in $I = \langle a, b \rangle$. We say that $h \to +0$ when $h \to 0$ through positive values and we wish to prove that the one-sided limits

$$g(\xi - 0) = \lim_{h \to +0} g(\xi - h), \quad g(\xi + 0) = \lim_{h \to +0} g(\xi + h) \quad (6.4.8)$$

exist throughout the interior of I, and that $g(a+0)$ and $g(b-0)$ exist. We may assume that g increases and is bounded in I. Let ξ be interior to I, or let $\xi = a$. We consider first a sequence of positive numbers H_k such that $H_k \downarrow 0$ and that $\xi + H_k \leq b$. Then $g(\xi + H_k) \downarrow \gamma$, say. We have to show that the value of γ does not depend on the choice of our sequence.

Consider an arbitrary sequence of numbers $h_k \to +0$; we may assume that $h_k \leq H_1$. To each h_i corresponds an H_{k_i} such that $0 < H_{k_i+1} < h_i \leq H_{k_i}$, so that $g(\xi + H_{k_i+1}) \leq g(\xi + h_i) \leq g(\xi + H_{k_i})$. Hence $g(\xi + h_i) \to \gamma$. This proves the existence of $g(\xi + 0)$; the proof for $g(\xi - 0)$ is similar.

Clearly, if $g(x)$ increases,

$$g(\xi - 0) \leq g(\xi) \leq g(\xi + 0), \quad . \quad . \quad (6.4.9)$$

and g will be continuous at ξ if, and only if, $g(\xi - 0) = g(\xi + 0)$.

Theorem 92. *If* g(x) *is of bounded variation in* I, *then it is continuous in* I *except, perhaps, in an enumerable subset: it is thus continuous p.p. in* I.

Proof. Again we may assume that g increases. If ξ is a point of discontinuity, then there is a rational number r such that

$$g(\xi - 0) < r < g(\xi + 0).$$

To different points ξ correspond different numbers r, since $\xi_1 < \xi_2$ implies $g(\xi_1 + 0) \leq g(\xi_2 - 0)$. The rational numbers being enumerable, our proposition is proved.

The four numbers

$$D_{-}(\xi,g) = \varliminf_{h\to+0} \frac{g(\xi-h)-g(\xi)}{-h},\ D^{-}(\xi,g) = \varlimsup_{h\to+0} \frac{g(\xi-h)-g(\xi)}{-h},$$

$$D_{+}(\xi,g) = \varliminf_{h\to+0} \frac{g(\xi+h)-g(\xi)}{h},\ D^{+}(\xi,g) = \varlimsup_{h\to+0} \frac{g(\xi+h)-g(\xi)}{h}. \quad (6.4.10)$$

always † exist, finite or infinite. They are called the *lower and upper derivatives* from the left, or from the right: in short, they are the four *derived numbers* of g at ξ. Clearly,

$$D_{-}(\xi) \leq D^{-}(\xi), \quad D_{+}(\xi) \leq D^{+}(\xi). \quad (6.4.11)$$

Also

$$D_{+}(\xi,g_1+g_2) \geq D_{+}(\xi,g_1) + D_{+}(\xi,g_2),$$
$$D^{+}(\xi,g_1+g_2) \leq D^{+}(\xi,g_1) + D^{+}(\xi,g_2); \quad (6.4.12)$$

and there are similar inequalities for the left-hand derivatives. All these derivatives are measurable, by (5.4.6).

If g increases, then all four derived numbers are non-negative. The ordinary derivative $g'(\xi)$ will exist if, and only if, all these derived numbers are equal and finite at ξ. It is, however, convenient to denote by $g'(\xi)$ the common value even if it is infinite.

The following theorem is fundamental for the theory of the indefinite integral. It is due to *Lebesgue*, and its delicate proof is based on Vitali's covering theorem (Theorem 46).

THEOREM 93. *If* g(x) *is of bounded variation in* I, *then it is differentiable p.p. in* I. *The derivative* g′(x) *is* L-*integrable over* I, *and hence is finite p.p.*

If g(x) *increases then*

$$\int_a^b g'(x)dx \leq g(b) - g(a) \quad (6.4.13)$$

PROOF. We may assume that g increases and is bounded in I.

† At the end-points of I only two exist.

(i) First, we prove that the four derived numbers of g are equal p.p. in I. Given two positive rational numbers $r < s$, consider the set $E = E(r,s)$, measurable by (5.4.3), of all points ξ in I for which

$$D_+(\xi) < r < s < D^-(\xi). \quad . \quad . \quad . \quad \text{(a)}$$

We shall prove that

$$\mu(E(r,s)) = 0. \quad . \quad . \quad . \quad (6.4.14)$$

Suppose that, to the contrary, $\mu(E) = \mu > 0$. According to the left-hand side inequality in (a), there exists for each ξ of E an infinity of intervals $\langle \xi, \xi + h \rangle$ with $h \to +0$, such that

$$g(\xi + h) - g(\xi) < hr. \quad . \quad . \quad . \quad . \quad \text{(b)}$$

By the corollary of Vitali's Theorem 46 we can find a finite sum S of such h-intervals

$$J_i = \langle \xi_i, \xi_i + h_i \rangle \qquad (i = 1, 2, \ldots m), \quad \xi_i \subset E,$$

say, which do not overlap and such that, if a positive ϵ (< 1) is given,

$$\sum_1^m h_i < (1 + \epsilon)\mu, \quad \mu(E^*) = \mu(E \,.\, S) > (1 - \epsilon)\mu. \quad . \quad \text{(c)}$$

On the other hand, since $E^* \subset E$ and according to the right-hand inequality in (a), there exists, for each ξ^* of E^*, an infinity of intervals $\langle \xi^* - k, \xi^* \rangle$ with $k \to +0$, such that

$$\frac{g(\xi^* - k) - g(\xi^*)}{-k} > s, \quad g(\xi^*) - g(\xi^* - k) > ks. \quad . \quad \text{(d)}$$

We may also assume that all these k-intervals are contained in S. For, $E^* \subset S$ and hence ξ^* is in some of the above intervals J_i. If it is not the left end-point of a J_i, we can choose k so small that $\xi^* - k$ is also in J_i: the m left end-points of the J_i can be omitted from E^* without affecting the estimate (c) for $\mu(E^*)$.

Again using the corollary of Vitali's theorem we find a finite sum $S^* \subset S$ of such k-intervals,

$$J_l^* = \langle \xi_l^* - k_l, \xi_l^* \rangle \qquad (l = 1, 2, \ldots . m^*), \quad \xi_l^* \subset E^*,$$

say, which do not overlap and such that

$$|S^*| = \sum_1^{m^*} k_l \geq \mu(E^* . S^*) > (1-\epsilon)\mu(E^*) > (1-\epsilon)^2\mu, \quad \text{(e)}$$

by (c). Now, we have

$$\sum_1^{m^*}(g(\xi_l^*) - g(\xi_l^* - k_l)) \leq \sum_1^{m}(g(\xi_i + h_i) - g(\xi_i)), \quad . \quad \text{(f)}$$

since $S^* \subset S$ and g increases. From this follows, by (b) and (d),

$$s\sum_1^{m^*} k_l < r\sum_1^{m} h_i \, ;$$

and, applying (e) and (c),

$$s(1-\epsilon)^2\mu < r(1+\epsilon)\mu, \quad s(1-\epsilon)^2 < r(1+\epsilon).$$

On letting $\epsilon \to 0$, we obtain $s \leq r$: a contradiction. This proves (6.4.14).

Next, let E_1 be the set of all points ξ in I for which $D_+(\xi) < D^-(\xi)$. Clearly, any such ξ must belong to some set $E(r,s)$, so that

$$E_1 \subset \Sigma E(r,s),$$

where the sum is extended over the enumerable set of all pairs of positive rational numbers $r < s$. Hence $\mu(E_1) = 0$, by (6.4.14). This is to say, we have proved that

$$D^-(\xi) \leq D_+(\xi) \quad . \quad . \quad . \quad (6.4.15)$$

p.p. in I. In a similar way it is proved that

$$D_-(\xi) \geq D^+(\xi) \quad . \quad . \quad . \quad (6.4.16)$$

p.p. in I. It now follows, from (6.4.11), that all the four derived numbers coincide p.p. in I, so that $g'(\xi)$ exists for almost all ξ in I, finite or infinite.

(ii) It remains to show that $g'(\xi)$ is L-integrable and that (6.4.13) holds. We extend the definition of g by putting $g(x)=g(a)$ if $x<a$, and $g(x)=g(b)$ if $x>b$. Now $g(x)$, as an increasing function, is measurable, and so is $g(x+h)$. The function

$$D_+(x)=\varliminf_{h\to+0}\frac{g(x+h)-g(x)}{h}$$

is also measurable, by (5.4.6). Also $D_+(x)\geq 0$,† since g increases. Hence the integral of D_+ over I exists, finite or infinite, by Theorem 76. Again $D_+(x)=g'(x)$ whenever $g'(x)$ exists, that is p.p. in I. Hence

$$\int_a^b g'dx=\int_a^b D_+dx\leq\varliminf_{h\to+0}\frac{1}{h}\int_a^b(g(x+h)-g(x))dx, \qquad \text{(g)}$$

by Fatou's Lemma (5.3.1).

Now, if $h>0$, the ordinate set of $g(x+h)$ over $\langle a, b\rangle$ is congruent to the ordinate set of g over $\langle a+h, b+h\rangle$. It follows that

$$\frac{1}{h}\int_a^b(g(x+h)-g(x))dx=\frac{1}{h}\Big[\int_{a+h}^{b+h}gdx-\int_a^b gdx\Big]$$

$$=\frac{1}{h}\Big[\int_b^{b+h}gdx-\int_a^{a+h}gdx\Big]=g(b)-\frac{1}{h}\int_a^{a+h}gdx\leq g(b)-g(a).$$

Hence (6.4.13) follows from (g), and g' is integrable over I. This completes the proof of the theorem.

6.5. The derivative of the indefinite integral. The indefinite integral $F(x)$ of $f(x)$ is b.v. (and even a.c.). Hence, by Theorem 93, its derivative $F'(x)$ exists finitely p.p. in I. But we do not know yet whether, or where, $F'(x)=f(x)$, as in (6.1.2). We need the following theorem.

THEOREM 94. *Let* f(x) *and* g(x) *be* L*-integrable over* I, *and let* F(x) *and* G(x) *be their indefinite integrals, respectively.*

If $F(x)\equiv G(x)$, *then* $f(x)=g(x)$ *p.p. in* I.

† $D_+(x)=g'(x)=0$ outside $\langle a, b\rangle$.

Proof. If $h = f - g$, then

$$H(x) = \int_a^x h\,dt = F(x) - G(x) \equiv 0.$$

Also

$$\int_\alpha^\beta h\,dt = H(\beta) - H(\alpha) = 0 \qquad . \quad . \quad . \quad \text{(a)}$$

for all $\alpha < \beta$ in I.

Let E be the (measurable) subset of I where $h \geq 0$. It is, clearly, sufficient to prove that $h = 0$ p.p. in E. Now, if $0 < \eta < 1$, we can find, by (3.5.10), a closed subset F of E such that

$$\eta\mu(E) \leq \mu(F) \leq \mu(E). \qquad . \quad . \quad . \quad \text{(b)}$$

The complement $I - F = I.{}_cF$ is an interval set since ${}_cF$ is open. It follows that

$$\int_{I-F} h\,dx = 0, \qquad \int_F h\,dx = \int_a^b h\,dx - \int_{I-F} h\,dx = 0,$$

by (a) and Theorem 67. This implies $h = 0$ p.p. in F, by Theorem 62. Hence, and by (b), $h > 0$ in a part of E of at most outer measure $(1 - \eta)\mu(E)$. On letting $\eta \to 1$, we see that $h = 0$ p.p. in E.

We can now answer the question concerning (6.1.2).

Theorem 95. *If* f(x) *is* L-*integrable over* I, *and* F(x) *is its indefinite integral, then*

$$F'(x) = f(x) \qquad . \quad . \quad . \quad . \quad (6.5.1)$$

p.p. in I.

Proof. We may assume that $f \geq 0$, when $F(x)$ increases and $F'(x) \geq 0$ p.p. First, we prove the formula

$$\int_a^x F'(t)\,dt = F(x). \qquad . \quad . \quad . \quad (6.5.2)$$

(i) Suppose that f is bounded in I: that $0 \leq f \leq M$, say. Then

$$0 \leq F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$$

p.p. and boundedly in I. For,

$$|F(x+h) - F(x)| = \left| \int_x^{x+h} f dt \right| \leq |h| M.$$

Hence, by Theorem 71, if $a \leq x < b$,

$$\int_a^x F' dt = \lim_{h \to +0} \int_a^x \frac{F(t+h) - F(t)}{h} dt = \lim \frac{1}{h} \left[\int_{a+h}^{x+h} F dt - \int_a^x F dt \right]$$

$$= \lim \frac{1}{h} \left[\int_x^{x+h} F dt - \int_a^{a+h} F dt \right] = F(x) - F(a) = F(x),$$

since $F(x)$ is continuous (§ 6.2.).

(ii) If f is unbounded, we use Fatou's Lemma (5.3.1), and obtain, as above,

$$\int_a^x F' dt \leq \lim_{h \to +0} \int_a^x \frac{F(t+h) - F(t)}{h} dt = F(x). \quad . \quad . \quad \text{(a)}$$

Now consider the bounded integrable function $f_{[a]} = \text{Min}(f, a)$ and its indefinite integral $F_a(x)$. Then $F(x) - F_a(x)$, the indefinite integral of the non-negative function $f - f_{[a]}$, increases in I, so that $F'(x) \geq F_a'(x)$ p.p. in I. Also (6.5.2) holds for $F_a(x)$. Hence

$$F_a(x) = \int_a^x F_a' dt \leq \int_a^x F' dt. \quad . \quad . \quad . \quad \text{(b)}$$

Again, $f_{[a]} \uparrow f$ and $F_a(x) \uparrow F(x)$ as $a \to \infty$, by Theorem 70. Hence we obtain from (b), in the limit,

$$F(x) \leq \int_a^x F' dt. \quad . \quad . \quad . \quad . \quad \text{(c)}$$

This, together with (a), proves (6.5.2).

(iii) It is now easy to complete the proof of the theorem. By (6.5.2), $F(x)$ is the indefinite integral of its derivative

$F'(x)$. Since it is also the indefinite integral of $f(x)$, Theorem 94 shows that $F'(x)=f(x)$ p.p. in I.

For the R-integral the proof of Theorem 95 is much simpler. For, if f is R-integrable, it is continuous p.p. in I, by Theorem 88. Hence (6.5.1) holds p.p.

6.6. Integration of a derivative. The formula (6.1.4) expresses the fact that $f(x)$ is an indefinite integral of its derivative. However, this formula is not always true, even if the integral of f' exists: a necessary condition is that $f(x)$ should be a.c. in I (Theorem 91).

The following is an example of a function $f(x)$, continuous in $I=\langle 0, 1\rangle$ and increasing from $f(0)=0$ to $f(1)=1$, yet such that $f'(x)=0$ p.p. in $\langle 0, 1\rangle$: hence (6.1.4) does not hold (and $f(x)$ is not a.c.).

Consider the sequence of all rational numbers $0<r_k<1$, $k\geq 1$, in some order of enumeration; and let

$$g(0)=0, \quad x=g(y)=\sum_{r_k<y} 2^{-k} \qquad . \qquad . \quad (6.6.1)$$

This function of y increases (strictly) in $I=\langle 0, 1\rangle$ from $g(0)=0$ to $g(1)=1$. Also

$$g(r_k-0)=g(r_k) \quad (=x_k, \text{ say}), \qquad g(r_k+0)=g(r_k)+2^{-k},$$

so that g is discontinuous at $y=r_k$: the "jump" is 2^{-k}. Otherwise $g(y)$ is continuous in $\langle 0, 1\rangle$, as is seen on approximating to y through rational numbers.

The values x taken by g form a subset E of I, the complement of which (with respect to I) are the half-open intervals $i_k=(x_k, x_k+2^{-k}\rangle$. Since these do not overlap and have total length 1, the set E has measure zero.

Now define the inverse function of g by

$$f(x)=\begin{cases} r_k \text{ when } x \text{ belongs to } i_k, \\ y \text{ where } x=g(y), \text{ otherwise.} \end{cases} \qquad . \quad (6.6.2)$$

Then $f(x)$ is continuous and increases from $f(0)=0$ to $f(1)=1$. Also it is constant in the interiors of all the i_k, so that $f'(x)=0$ p.p. in $\langle 0, 1\rangle$.

Next, we wish to show that (6.1.4) holds when $f(x)$ is a.c. We need the following result.

THEOREM 96. *Suppose that* f(x) *is absolutely continuous, and that* $f'(x)=0$ *p.p. in* I. *Then* f(x) *is constant.*

PROOF. It is sufficient to show that $f(a)=f(b)$. For, $f(x)$ satisfies the conditions of the theorem in every sub-interval $\langle a, x\rangle$.

Let E be the set of all points in the interior (I) of I for which $f'(x)=0$. We have $\mu(E)=b-a$. Given $\epsilon>0$, and a point ξ of E, all intervals $\langle\xi, \xi+h\rangle$ with sufficiently small $h>0$, will have the property that (i) they are contained in I, and that (ii)

$$|f(\xi+h)-f(\xi)|<\epsilon h. \quad . \quad . \quad . \quad \text{(a)}$$

On the other hand, by the corollary of Vitali's theorem (Theorem 46), we can find a finite sum S of such h-intervals so that they do not overlap and that

$$\mu(E-S\,.\,E)<\delta\mu(E)=\delta(b-a),$$

where $0<\delta<1$. Hence, if S^* is the finite sum of complementary open intervals in (I),

$$\mu(S^*)\le\mu(I-E\,.\,S)\le\mu(I-E)+\mu(E-E\,.\,S)<\delta(b-a), \quad \text{(b)}$$

since $\mu(I-E)=0$. The intervals of S and S^* together give a finite division of I into intervals $\langle x_i, x_{i+1}\rangle$, say. Now

$$|f(b)-f(a)|\le\Sigma|f(x_{i+1})-f(x_i)|+\Sigma^*|f(x_{i+1})-f(x_i)|, \quad \text{(c)}$$

where Σ and Σ^* are extended over S and S^*, respectively. Here

$$\Sigma<\epsilon\Sigma(x_{i+1}-x_i)<\epsilon(b-a)$$

by (a); and

$$\Sigma^*<\epsilon(b-a)$$

if $\delta<\delta(\epsilon)$. For, $\mu(S^*)$ becomes small for small δ, by (b); and so does Σ^* because f is a.c. It now follows from (c) that

$$|f(b)-f(a)|<2\epsilon(b-a)$$

if $\delta<\delta(\epsilon)$. As $\epsilon\to0$, we obtain $f(b)=f(a)$.

We can now prove the main theorem.

THEOREM 97. *In order that a function* f(x) *be an indefinite* L-*integral of its derivative in* I; *that is, in order that*

$$\int_a^x f'(t)dt = f(x) - f(a) \quad . \quad . \quad . \quad (6.6.3)$$

for all x *in* I, *it is necessary and sufficient that* f(x) *be absolutely continuous in* I.

PROOF. The condition is necessary, since the left-hand side of (6.6.3) is a.c., by Theorem 91.

Conversely, if f is a.c. in I, then f' exists p.p. and is L-integrable over I, by Theorem 93. Let $g(x)$ be the indefinite integral of f'. Then $g' = f'$ p.p. in I, by Theorem 95. Also g is a.c. Hence $h = f - g$ is a.c., and $h' = 0$ p.p. in I. It now follows from Theorem 96 that $f - g$ is constant so that f is an indefinite integral of f'.

This theorem answers the question about the validity of (6.6.3) completely. We may, however, ask whether or not the mere existence of f' and its integrability over I is sufficient; that is, whether it does not imply that f is a.c. The example of the function (6.6.2) shows that this is not usually true. We cannot allow the non-existence of f' even for a set of measure zero. On the other hand, if f' exists and is finite for all x, and if it is integrable over I, then we can prove our case.† We first assume that f' is bounded.

THEOREM 98. *Suppose that* f′(x) *exists and is bounded for all* x *in* I.‡ *Then* f(x) *is an indefinite* L-*integral of* f′(x).

PROOF. First, f is continuous since it is differentiable. Let $|f'| \leq M$. Then, by the mean-value theorem of the differential calculus,

$$\Sigma |f(b_k) - f(a_k)| = \Sigma (b_k - a_k) |f'(\xi_k)| \leq M \Sigma (b_k - a_k),$$

where $a_k < \xi_k < b_k$. This becomes as small as we please if

† An *enumerable* exceptional set is permissible (see the book by *Kestelman*).

‡ At the end-points one-sided derivatives are assumed.

$\Sigma(b_k - a_k)$ is small enough. Hence f is a.c. in I, and Theorem 97 can be applied.

The theorem is not true for the R-*integral*: another conspicuous advantage of Lebesgue's definition over the classical one.

The following example was given by *V. Volterra* (1881). The function $\phi(x) = x^2 \sin 1/x$ $(x \neq 0)$, $\phi(0) = 0$, is continuous for all x. Also

$$\phi'(x) = 2x \sin 1/x - \cos 1/x \ (x \neq 0) \ ; \ \phi'(0) = \lim_{x \to 0} \phi(x)/x = 0.$$

Hence $\phi'(x)$ is discontinuous at $x = 0$. Clearly, $|\phi'(x)| \leq 3$. Also ϕ' has an infinity of zeros other than $x = 0$: they are the roots of the equation $\tan 1/x = 1/2x$, and they have $x = 0$ as limit.

Now consider a perfect and nowhere dense subset E of $I = \langle 0, 1 \rangle$, of positive measure, such as obtained, in § 1.15, by removing from I a certain sequence of non-overlapping open intervals $i_k = (a_k, b_k)$ without common end-points.

First, we define $f(x)$ in the open complement $O = \Sigma i_k$ of E. Let ξ_k be the greatest zero of $\phi'(x - a_k)$ less than $\frac{1}{2}(a_k + b_k)$. We put

$$f(x) = \begin{cases} \phi(x - a_k) \\ \phi(\xi_k - a_k) \end{cases} \text{when} \quad \begin{matrix} a_k < x \leq \xi_k \\ \xi_k < x \leq \frac{1}{2}(a_k + b_k), \end{matrix} \qquad (6.6.4)$$

$$f(x) = f(b_k - x) \quad \text{when} \quad \tfrac{1}{2}(a_k + b_k) < x < b_k.$$

Clearly, $f'(x)$ is continuous in O.

If we now put $f(x) \equiv 0$ in E, then f is continuous throughout I. Also $f'(x) \equiv 0$ in E. For, let x belong to E, and let $h > 0$, say. Then $f(x+h) - f(x) = 0$, if $x + h$ also belongs to E. Otherwise $x + h$ must belong to some i_k, so that $x \leq a_k < x + h < b_k$. In this case

$$|f(x+h) - f(x)| = |f(x+h)| \leq (x + h - a_k)^2 \leq h^2.$$

Hence, in any case, $h^{-1}(f(x+h) - f(x)) \to 0$ as $h \to 0$. It follows that $f'(x)$ exists throughout I; also $|f'(x)| \leq 3$.

Now f' is discontinuous at the end-points of the intervals i_k and at the limiting points of these end-points: that is, in E. Since E has positive measure, f' cannot be R-integrable over I, by Theorem 88.

The proof of our final theorem is much more delicate than that of Theorem 98. We require three lemmas.

LEMMA 1. *Suppose that* f(x) *is continuous, and that* $D_+(x, f) \geq 0$ *in* I. *Then* f(x) *increases in* I.

PROOF. Let $a \leq x_1 < x_2 \leq b$, and suppose that

$$f(x_2) - f(x_1) < 0.$$

Then

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = -c, \quad . \quad . \quad . \quad . \quad \text{(a)}$$

say, where $c > 0$. The function

$$g(x) = \frac{f(x) - f(x_1)}{x - x_1}$$

is continuous in $(x_1, b\rangle$. Also $g(x) > -\frac{1}{2}c$ for all $x > x_1$ sufficiently near to x_1, because $D_+(x_1) \geq 0$. In view of (a) we can, therefore, find a greatest ξ with $x_1 < \xi < x_2$ such that $g(\xi) = -\frac{1}{2}c$. Hence $g(x) < -\frac{1}{2}c$ in $(\xi, x_2\rangle$, so that there

$$\frac{f(x) - f(\xi)}{x - \xi} = \frac{f(x) - f(x_1)}{x - \xi} + \frac{f(x_1) - f(\xi)}{x - \xi}$$

$$= \frac{1}{x - \xi}[(x - x_1)g(x) + (x_1 - \xi)g(\xi)] < \frac{-c}{2(x - \xi)}[(x - x_1) + (x_1 - \xi)]$$
$$= -\tfrac{1}{2}c.$$

On letting $x \to \xi + 0$, we obtain $D_+(\xi) \leq -\frac{1}{2}c$: a contradiction.

LEMMA 2. *Let* E, *of measure zero, be a subset of* I. *Then, given a positive* ϵ, *there exists a continuous increasing function* $\gamma(x) = \gamma(x, \epsilon)$ *in* I, *such that* $\gamma(a) = 0$, $0 \leq \gamma(x) \leq \epsilon$, *and that* $D_+(x, \gamma) = \infty$ *in* E.

PROOF. By (3.4.7) there exists an open set $O_k \supset E$ such that $|O_k| < 4^{-k}\epsilon$, k being a positive integer. We put

$$\gamma_k(x) = \begin{cases} 2^k \text{ if } x \text{ belongs to } O_k, \\ 0 \quad \text{otherwise}, \end{cases} \qquad \Gamma_k(x) = \int_a^x \gamma_k(t)dt.$$

Clearly, $\Gamma_k(a)=0$ and $\Gamma_k(x)$ increases. Also

$$0\leq \Gamma_k(x)\leq 2^k|\, O_k\,|<2^{-k}\epsilon.$$

Further $\Gamma_k'(x)=\gamma_k(x)=2^k$ in O_k, and hence in E. Now

$$\gamma(x)=\sum_1^{\infty}\Gamma_k(x)$$

is uniformly convergent in I and therefore continuous. Also $\gamma(x)$ increases, $\gamma(a)=0$, and $0\leq\gamma(x)<\epsilon\Sigma 2^{-k}=\epsilon$. Finally, by (6.4.12), if x belongs to E,

$$D_+(x,\gamma)\geq D_+(x,\ \sum_1^m\Gamma_k)+D_+(x,\ \sum_{m+1}^{\infty}\Gamma_k)\geq D_+(x,\sum_1^m\Gamma_k)$$

$$\geq\sum_1^m D_+(x,\ \Gamma_k)=\sum_1^m\Gamma_k'(x)=\sum_1^m 2^k,$$

so that $D_+(x,\gamma)=\infty$ in E.

Lemma 3. *Suppose that* f(x) *is continuous in* I, *that* $\mathrm{D}_+(\mathrm{x},\mathrm{f})\geq 0$ *p.p., and that* $\mathrm{D}_+(\mathrm{x},\mathrm{f})\neq-\infty$ *in* I. *Then* f(x) *increases in* I.

Proof. Let E be the set, of measure zero, where $D_+(x)<0$; and let $\gamma(x)=\gamma(x,\epsilon)$ be the corresponding function of Lemma 2. Then $f(x)+\gamma(x)$ is continuous, and

$$D_+(x,f+\gamma)\geq D_+(x,f)+D_+(x,\gamma)\geq 0$$

throughout I, since $D_+(x,f)$ is finite in E. Hence $f+\gamma$ increases in I, by Lemma 1. It follows, if $a\leq x_1<x_2\leq b$, that

$$f(x_2)+\gamma(x_2)\geq f(x_1)+\gamma(x_1)\geq f(x_1),\ f(x_2)-f(x_1)\geq-\gamma(x_2)\geq-\epsilon.$$

Since $\epsilon(>0)$ is arbitrary, we finally find that $f(x_2)\geq f(x_1)$.

Theorem 99. *Suppose that* f′(x) *exists and is finite for all* x *in* I,† *and that it is* L*-integrable over* I. *Then* f(x) *is an indefinite* L*-integral of* f′(x).

† See footnotes on p. 148.

PROOF. The function $g_n(x) = \text{Max}\,(f'(x), -n)$ is measurable, by (5.4.5), and hence is integrable over I, since $|g_n| \leq |f'|$. Also $g_n(x) \downarrow f'(x)$ dominatedly as $n \to \infty$. Hence

$$G_n(x) = \int_a^x g_n dt \downarrow \int_a^x f' dt. \quad . \quad . \quad . \quad \text{(a)}$$

If x is in $\langle a, b)$, then for small positive h,

$$\frac{G_n(x+h) - G_n(x)}{h} = \frac{1}{h} \int_x^{x+h} g_n dt \geq -n.$$

Since f' is finite, it follows that in $\langle a, b)$

$$D_+(G_n - f) \geq D_+(G_n) + D_+(-f) \geq -n - f' \neq -\infty.$$

Also $G_n - f$ is continuous in $\langle a, b\rangle$, and $g_n \geq f'$, so that

$$(G_n - f)' = g_n - f' \geq 0$$

p.p. in I, by Theorem 95. By Lemma 3, the function $G_n - f$ increases in I. Hence, in I,

$$G_n(x) - f(x) \geq G_n(a) - f(a) = -f(a), \quad f(x) - f(a) \leq G_n(x).$$

As $n \to \infty$, we obtain, by (a),

$$f(x) - f(a) \leq \int_a^x f' dt. \quad . \quad . \quad . \quad \text{(b)}$$

The opposite inequality is also true, as is seen on applying (*b*) to $-f$. This proves the theorem.

The condition that f' be integrable is indispensable: if $y = x^2 \sin(^1/_{x^2})$ for $x \neq 0$, $y(0) = 0$, then y' exists and is finite for all x [$y'(0) = 0$], but $|y'|$, and hence y', is not integrable in $\langle 0, 1\rangle$. Again there exist differentiable functions for which $f'(x)$ is finite p.p. (and is infinite otherwise), f' is integrable, and yet f is not an indefinite L-integral of f'. This (slight) deficiency of the L-integral has led to more comprehensive definitions of the integral (*Denjoy* (1912), *Perron* (1914)). They include, as special cases, the L-integral as well as Cauchy integrals like that in (4.1.11). These integrals are not, however, absolute integrals.

6.7. Integration by parts and rule of substitution. To complete the account of the "calculus" for the L-integral, we have still to establish the familiar rules of integration by parts and for the substitution of a new variable.

THEOREM 100. *Suppose that* f(x) *is* L-*integrable over* ⟨a, b⟩, *and that* g(x) *is absolutely continuous in* ⟨a, b⟩. *Then*

$$\int_a^b fg\,dx = \Phi(b)g(b) - \Phi(a)g(a) - \int_a^b \Phi g' dx, \quad . \quad (6.7.1)$$

where $\Phi(x)$ *is some indefinite integral of* f(x).

PROOF. Since f is integrable, and g is bounded, fg is integrable. Also g' exists p.p. and is integrable, by Theorem 93. Hence $\Phi g'$ is integrable, since Φ is bounded. Both integrals in (6.7.1) therefore exist.

Next the function $h = \Phi g$ is, as product of two a.c. functions, a.c. itself. Hence, by Theorem 97,

$$\int_a^b h' dt = h(b) - h(a).$$

Since $h' = fg + \Phi g'$ p.p., this is (6.7.1).

We turn now to the substitution rule. We shall meet, in addition to the variable x, a new independent variable t. The derivative of a function $h(t)$, say, will be denoted by $\dot{h}(t)$.

THEOREM 101. *Suppose that* f(x) *is bounded and* L-*integrable over* ⟨a, b⟩ : *that* $\Phi(t)$ *is absolutely continuous in* ⟨α, β⟩; *and that* $\Phi(\alpha) = \mathrm{a}$, $\Phi(\beta) = \mathrm{b}$, *and* $\mathrm{a} \le \Phi(\mathrm{t}) \le \mathrm{b}$. *Then* $\mathrm{f}(\Phi(\mathrm{t}))\dot{\Phi}(\mathrm{t})$ *is* L-*integrable over* ⟨α, β⟩, *and*

$$\int_a^b f(x)dx = \int_\alpha^\beta f(\phi(t))\dot{\phi}(t)dt. \quad . \quad . \quad (6.7.2).$$

PROOF. We proceed in several steps.

(i) If $F(x)$ is the indefinite integral of f, then $\Phi(t) = F(\phi(t))$ is defined in ⟨α, β⟩. We wish to prove that $\Phi(t)$ is a.c.

Consider a finite number of mutually separate intervals $\langle \alpha_k, \beta_k \rangle$ contained in $\langle \alpha, \beta \rangle$. If $|f| \leq M$, then, given an ϵ (>0),

$$\Sigma |\Phi(\beta_k) - \Phi(\alpha_k)| \leq \Sigma \left| \int_{\phi(\alpha_k)}^{\phi(\beta_k)} |f(x)| \, dx \right|$$

$$\leq M\Sigma |\phi(\beta_k) - \phi(\alpha_k)| < \epsilon,$$

whenever $\Sigma(\beta_k - \alpha_k) < \delta(\epsilon)$. For, $\phi(t)$ is a.c. Hence $\Phi(t)$ is also a.c.

(ii) We suppose, next, that $f(x)$ is continuous.

Let $h > 0$, and let t and $t+h$ be in $\langle \alpha, \beta \rangle$. Then, if $k \neq 0$,

$$\frac{\Phi(t+h) - \Phi(t)}{h} = \frac{F(x+k) - F(x)}{k} \cdot \frac{\phi(t+h) - \phi(t)}{h}, \qquad \text{(a)}$$

where $\phi(t) = x$ and $\phi(t+h) = x+k$. Also $k \to 0$ as $h \to 0$.

Now, $\Phi(t)$ and $\phi(t)$ being a.c., their derivatives $\dot{\Phi}(t)$ and $\dot{\phi}(t)$ exist and are finite p.p. in $\langle \alpha, \beta \rangle$. Also $F'(x) = f(x)$ for all x, since f is continuous.† Hence we obtain from (a), as $h \to 0$, that

$$\dot{\Phi}(t) = f(x)\dot{\phi}(t) = f(\phi(t))\dot{\phi}(t) \quad . \quad . \quad . \quad \text{(b)}$$

for almost all t. Hence, by Theorem 97 and since $\Phi(\alpha) = 0$,

$$\int_a^b f dx = \int_a^{\phi(\beta)} f dx = \Phi(\beta) = \int_\alpha^\beta \dot{\Phi}(t) dt = \int_\alpha^\beta f(\phi(t))\dot{\phi}(t) dt,$$

as desired.

(iii) To prove the theorem in the general case, we may assume that $f \geq 0$. Let Y be the measurable and bounded ordinate set of f. By (3.4.1), there exists a decreasing sequence of bounded interval sets $\Sigma_k \supset Y$ over $\langle a, b \rangle$, such that

$$|\Sigma_k| \downarrow m(Y) = \int_a^b f dx. \quad . \quad . \quad . \quad . \quad \text{(c)}$$

† If f is not continuous, $F'(x) = f(x)$ holds only for *almost all* x. But it is not obvious (though true if $\dot{\phi}(t) \neq 0$) that this implies $F'(x) = f(\phi(t))$ for *almost all* t. Here lies the difficulty of the proof in the general case.

If $\Sigma_k = \lim_{m\to\infty} \sum_{i=1}^{m} J_i^{(k)} = \lim S_k^{(m)}$, then, by (5.7.5),†

$$|\Sigma_k| = \int_a^b \mu_k(x)dx, \quad |S_k^{(m)}| = \int_a^b \mu_k^{(m)}(x)dx,$$

where $\mu_k(x)$ and $\mu_k^{(m)}(x)$ are the bounded lengths of the x-sections of Σ_k and $S_k^{(m)}$, respectively. Also $\mu_k^{(m)}(x) \uparrow \mu_k(x)$ as $m \to \infty$. Now $\mu_k^{(m)}(x)$ is a step function: it is constant in the intervals of a finite division of $\langle a, b\rangle$. We can, evidently, approximate to it by a continuous polygon-function $p_k^{(m)}(x)$, say, so that also $p_k^{(m)}(x) \to \mu_k(x)$ as $m \to \infty$.

Again,

$$\int_a^b p_k^{(m)}(x)dx = \int_\alpha^\beta p_k^{(m)}(\phi(t))\dot{\phi}(t)dt, \qquad \text{(d)}$$

by what we have proved in (ii). As $m \to \infty$, we obtain

$$|\Sigma_k| = \int_a^b \mu_k(x)dx = \int_\alpha^\beta \mu_k(\phi(t))\dot{\phi}(t)dt, \qquad \text{(e)}$$

by Theorem 71 (Lebesgue's Test), since $|p_k^{(m)}(x)| \le A$, say, in $\langle a, b\rangle$ and $|p_k^{(m)}(\phi(t))\dot{\phi}(t)| \le A|\dot{\phi}(t)|$ in $\langle\alpha, \beta\rangle$.

Next, $\mu_k(x) \downarrow \bar{f}(x)$, say, boundedly as $k \to \infty$. Hence, by (c) and the above argument,

$$m(Y) = \int_a^b f(x)dx = \int_\alpha^\beta \bar{f}(\phi(t))\dot{\phi}(t)dt. \qquad \text{(f)}$$

Also $\bar{f}(x) \ge f(x)$, since $\Sigma_k \supset Y$ and so $\mu_k(x) \ge f(x)$.

Now, let $0 \le f(x) \le M$, say. Applying (f) to $M - f(x)$, we find a function $M - \underline{f}(x)$, say, such that

$$\int_a^b (M-f)dx = \int_\alpha^\beta [M - \underline{f}(\phi(t))]\dot{\phi}(t)dt = M(b-a) - \int_\alpha^\beta \underline{f}(\phi(t))\dot{\phi}(t)dt,$$

$$\int_a^b f(x)dx = \int_\alpha^\beta \underline{f}(\phi(t))\dot{\phi}(t)dt. \qquad \text{(g)}$$

Also $M - \underline{f}(x) \ge M - f(x)$, or $f(x) \ge \underline{f}(x)$.

The condition that $a \le \phi(t) \le b$ in $\langle\alpha, \beta\rangle$ has not been used

† In the simple case of an interval set.

except for defining the occurring compound functions of t. Hence (f) and (g) can be applied to any interval $\langle a, x\rangle$ where $x=\phi(t)$: clearly, $\bar{f}(x)$ and $\underline{f}(x)$ can be taken the same for all these intervals. It follows that

$$\int_\alpha^t \underline{f}(\phi(\tau))\dot{\phi}(\tau)d\tau = \int_\alpha^t \bar{f}(\phi(\tau))\dot{\phi}(\tau)d\tau \quad . \quad . \quad \text{(h)}$$

for every t in $\langle \alpha, \beta\rangle$; and this implies, by Theorem 95, that $\underline{f}(\phi(t))\dot{\phi}(t)=\bar{f}(\phi(t))\dot{\phi}(t)$ for almost all t in $\langle\alpha, \beta\rangle$. For these t they equal $f(\phi(t))\dot{\phi}(t)$, since $\underline{f}\leq f\leq \bar{f}$. Hence (f) (or (g)), proves our theorem.

The condition, in Theorem 101, that $f(x)$ be bounded in $\langle a, b\rangle$, cannot usually be omitted. If, for instance,

$$f(x)=\frac{1}{\sqrt{x}} \text{ in } \langle 0, 1\rangle, \quad \phi(t)=t^2 \sin^2\frac{\pi}{2t} \text{ in } \langle 0, 1\rangle, \quad . \quad (6.7.3)$$

then the conditions of the theorem are otherwise satisfied, but $f(\phi(t))\dot{\phi}(t)$ is not L-integrable over $\langle 0, 1\rangle$.

If, however, $\phi(t)$ *increases in* $\langle\alpha, \beta\rangle$, *then* $f(x)$ *need not be bounded* : formula (6.7.2) holds with the proviso that we put $f(\phi(t))\dot{\phi}(t)=0$ whenever $\dot{\phi}(t)=0$, whether $f(\phi(t))$ be finite or infinite.

For, we may assume that $f\geq 0$. If $f_{[k]}=\text{Min }(f, k)$ then

$$\int_a^b f_{[k]}dx=\int_\alpha^\beta f_{[k]}(\phi(t))\dot{\phi}(t)dt,$$

by Theorem 101. Also $f_{[k]}\uparrow f$ in $\langle a, b\rangle$, and $f_{[k]}(\phi(t))\dot{\phi}(t) \uparrow f(\phi(t))\dot{\phi}(t)$ p.p. in $\langle\alpha, \beta\rangle$ as $k\to\infty$, since $\dot{\phi}(t)\geq 0$. Hence (6.7.2) follows, by Theorem 70.

FURTHER LITERATURE ON THE INTEGRAL

1. H. Lebesgue, *Leçons sur l'intégration et la recherche des fonctions primitives*, Paris, 1904.
2. Ch. J. de la Vallée Poussin, *Intégrales de Lebesgue, fonctions d'ensembles, classes de Baire*, Paris, 1916.
3. C. Carathéodory, *Vorlesungen über reelle Functionen*, Leipzig, 1918.
4. E. Kamke, *Das Lebesguesche Integral*, Leipzig–Berlin, 1925.
5. L. C. Young, *The Theory of Integration*, Cambridge, 1927.
6. St. Saks, *Théorie de l'intégral*, Warszawa, 1933; 2nd edition, *Theory of the Integral*, Warszawa–Lwow, 1937.
7. H. Kestelman, *Modern Theories of Integration*, Oxford, 1937.
8. E. C. Titchmarsh, *The Theory of Functions*, 2nd edition, Oxford, 1939.
9. E. J. McShane, *Integration*, Princeton, 1944.
10. J. C. Burkill, *The Lebesgue Integral*, Cambridge, 1951.

INDEX

(The numbers indicate paragraphs)