Original copy

TOPOLOGICAL SPACES

including a treatment of
MULTI-VALUED FUNCTIONS
VECTOR SPACES AND CONVEXITY

CLAUDE BERGE

Maître de Recherches au Central National
de la Recherche Scientifique

Translated by
E. M. PATTERSON

OLIVER & BOYD
EDINBURGH AND LONDON

OLIVER AND BOYD LTD

Tweeddale Court
Edinburgh 1

39a Welbeck Street
London, W.1

This is an English translation of *Espaces Topologiques, fonctions multivoques*, by Claude Berge, published by Dunod, Paris 1959

First English Edition 1963

Printed in Great Britain by
Robert Cunningham & Sons Ltd, Alva, Scotland

PREFACE

The Theory of Sets, which scarcely existed about fifty years ago, now plays an important role in many mathematical disciplines: measure theory, functions of real or complex variables, linear operations, etc. To apply it to these different domains, it is not necessary to have an axiomatic approach: a set is simply a collection of objects and the arguments which we shall use will be independent of the nature of these objects.

In Set Topology,[1] with which we are concerned in this book, we study sets in topological spaces and topological vector spaces; whenever these sets are collections of n-tuples or classes of functions, we recover well-known results of classical analysis.

But the role of topology does not stop there; the majority of text-books seem to ignore certain problems posed by the calculus of probabilities, the decision functions of statistics, linear programming, cybernetics, economics; thus, in order to provide a topological tool which is of equal interest to the student of pure mathematics and the student of applied mathematics, we have felt it desirable to include a systematic development of the properties of *multi-valued functions*. This term is generic; it indicates that we are not solely concerned with 'single-valued' functions. In fact, convention forces us to use different terms, following the preoccupations of different authors: we speak of a *multi-valued mapping*[2] whenever we study properties concerned with linearity or continuity; we speak of a *binary relation* whenever we study certain structural properties (order, equivalence, etc.); we speak of an *oriented graph* whenever we study combinatorial properties. In the first case we can say that we have a topological theory, in the second an algebraic theory and in the third a combinatorial theory. The coexistence of these different terminologies might appear to be unfortunate, but their use is standard whenever the points of view and the methods differ (we also follow the usual conventions regarding the words *equivalence* and *partition*).

The reader working without supervision is advised to omit on the first reading the paragraphs marked with an asterisk (these are generally in order to demonstrate a point of detail rigorously). It is hoped that the examples, which are taken from very different domains, will render the exposition more concrete.

We wish to express our thanks to Professor A. Lichnerowicz, who en-

(1) 'Set' topology or 'general' topology is so named in order to distinguish it from combinatorial topology, which is not considered in the present work.

(2) The following terms are also used: multiform functions, multivalent mappings. In this book we shall simply call them mappings.

couraged us to undertake this work and who kindly read the manuscript; equally we wish to thank Dr M. Kervaire, whose suggestions have been valuable for each one of our chapters.

TRANSLATOR'S PREFACE

Every effort has been made to preserve the spirit of the French edition in the present translation. The main alterations which have been made have been incorporated at the author's own suggestion. These consist of various amendments and the addition of new material, mainly in Chapter VIII.

I am indebted to Dr Berge and to the publishers for their close co-operation; also to Dr A. J. White for his helpful comments.

LIST OF SYMBOLS

ε, η are used throughout to denote strictly positive numbers.

Chapter I, § 1:

$a \in A$	a belongs to A
$a \notin A$	a does not belong to A
$A \subset B$	A is contained in B
$A \supset B$	A contains B
$A \subset\subset B$	A is strictly contained in B
$A \not\supset B$	A does not contain B
$A = B$	A is equal to B
$A \neq B$	A is not equal to B
$\varnothing$	the empty set
$\{a/a \text{ satisfies } (L)\}$	the set of elements a which satisfy (L)
$\mathbf{R}, \hat{\mathbf{R}}, \mathbf{N}, \mathbf{R}_r^+, \mathbf{R}_r$	the set of real numbers, the complete set of real numbers, the set of positive integers, the set of rational numbers $\geqq 0$, the set of rational numbers
$[\lambda, \mu], [\lambda, \mu[,]\lambda, \mu],]\lambda, \mu[$	closed interval (or segment), half-closed intervals, open interval
$\Rightarrow$	implies that
$\Leftrightarrow$	is equivalent to
$(\forall_A a)$:	for all a in A, we have
$(\exists_A a)$:	there exists an a in A such that
$a \to b$	a corresponds to b

§ 2:

$A \cup B$	the union of A and B
$A \cap B$	the intersection of A and B
$A - B$	A minus B
$-B$	the complement of B

§ 3:

$\mathscr{A} \subset B$	$\mathscr{A}$ is contained in B
$\mathscr{A} \vdash B$	$\mathscr{A}$ is partially contained in B
$\mathscr{A} \succ \mathscr{B}$	$\mathscr{A}$ is finer than $\mathscr{B}$ (in the interior sense)
$\mathscr{A} \approx \mathscr{B}$	$\mathscr{A}$ is equivalent to $\mathscr{B}$ (in the interior sense)
$\mathscr{A} \vdash \mathscr{B}$	$\mathscr{A}$ is finer than $\mathscr{B}$ (in the exterior sense)
$\mathscr{A} \simeq \mathscr{B}$	$\mathscr{A}$ is equivalent to $\mathscr{B}$ (in the exterior sense)

§ 4:

$\prod_{i\in I} A_i = A_i \times A_j \times \cdots$	the Cartesian product of the A_i
$\sum_{i\in I} A_i = A_i + A_j + \cdots$	the Cartesian sum of the A_i
$\bigcup_{i\in I} A_i = A_i \cup A_j \cup \cdots$	the union of the A_i
$\bigcap_{i\in I} A_i = A_i \cap A_j \cap \cdots$	the intersection of the A_i

§ 7:

$A \to \bar{A}$	closure operation
$A \to \mathring{A}$	interior operation

§ 9:

$\overline{\mathrm{Lim}}_{n=\infty} (A_n)$, $\underline{\mathrm{Lim}}_{n=\infty} (A_n)$	superior and inferior principal limits of the A_n

Chapter II:

$\hat{\Gamma}$	transitive closure of Γ
Γ^-, Γ^+	inverses (lower, upper) of Γ
$\Gamma_1 \cap \Gamma_2, \Gamma_1 \times \Gamma_2, \Gamma_1 \cdot \Gamma_2$	operations on Γ_1 and Γ_2: intersection, Cartesian product, composition product.

Chapter III:

$\mathscr{P}(A)$	family of subsets of A
$\aleph_0, \aleph_1, \ldots$	aleph nought, aleph one, ... (cardinal numbers)
$\omega_1, \omega_2, \ldots$	omega one, omega two, ... (ordinal numbers)
$\geqq, >, =, <, \leqq, \equiv$	greater than, strictly greater than, equal to, strictly less than, less than, identical to
$o(A)$	cardinal number of A
$\max_{a\in A} a$ (or $\max A$)	maximum of A
$\min_{a\in A} a$ (or $\min A$)	minimum of A
$\sup_{a\in A} a$ (or $\sup A$)	supremum of A
$\inf_{a\in A} a$ (or $\inf A$)	infimum of A

Chapter IV:

$\mathscr{G}, \mathscr{F}$	family of open sets, family of closed sets
$\bar{A}$	closure (topological) of A
$\mathring{A}$	interior (topological) of A
Fr A	frontier of A
$U(x)$, $V(x)$	open neighbourhoods of x
$M(x)$, $N(x)$	neighbourhoods of x
$\mathscr{V}(x)$	the family of open neighbourhoods of x
$\mathscr{N}(x)$	the family of neighbourhoods of x
$\mathscr{B}(x)$	a fundamental base of neighbourhoods of x
$(x_i) = (x(i) \mid i \in I; \mathscr{B})$	filtered family
(x_n)	sequence
$(x_i) \vdash (y_j)$	(x_i) is a sub-sequence of (y_j)

Chapter V:

$d(x, y)$	distance from x to y
$B_\lambda(x_0)$	ball of centre x_0 and radius λ
$S_\lambda(x_0)$	sphere of centre x_0 and radius λ
$\mathscr{G}_d$	family of open sets with respect to the metric d
$C_K(a)$	connected component containing the element a in K

Chapter VII:

0	zero, or neutral element (the same symbol for the zero scalar and the neutral element of a vector space)
$+$	sum
E_f^α	plane
H_f^α	closed half-space
H'^α_f	open half-space
$\mathbf{P}_n$	subset of $\mathbf{R}^n$ defined on page 142.
k$[A]$, lin $[A]$, s$[A]$, $[A]$	closures of the set A: conical, linear, spatial, convex
kc$[A]$	conical-convex closure of A
$\ddot{C}$	profile of the convex set C

Chapter VIII:

$\langle a, x\rangle$	scalar product of a and x
$\bar{c}[A]$	closed-convex closure of A
$]a_1, a_2, \ldots, a_n[, [a_1, a_2, \ldots, a_n]$	convex polyhedra generated by $a_1, a_2, \ldots, a_n$ (open and closed)
$\mathbf{A}$	matrix with coefficients a_j^i
$\otimes$	complete product of matrices

Chapter IX:

L_p	space defined on page 231
$\mathscr{L}_p$	space defined on page 232
$\mathscr{D}$	space defined on page 250
$\mathscr{C}$	space defined on page 253
$\mathscr{V}$	space defined on page 253
$(f_n) \rightharpoonup g$	(f_n) converges weakly to g

CONTENTS

CHAPTER I

FAMILIES OF SETS

§ 1. Sets: general notations

A **set** A is a collection of objects of any kind (for example points in a plane, real numbers, functions) which are called the **elements** (or **points**) of A; in general sets are denoted by capital Latin letters and elements are denoted by small Latin letters.

In certain cases a set can be determined by means of a list or, more generally, by means of a property of its elements; for example, the set of positive rational numbers,[1] which we denote by $\mathbf{R}_r^+$, is the collection of positive numbers x which have the following property: x is the quotient of an integer p by an integer q, where q is not zero.

If a is an element of the set A, we write $a \in A$ or $A \ni a$; if A and B are two sets and

$$a \in A \text{ implies that } a \in B,$$

we write $A \subset B$ (A is **contained** in B) and we also say that A is a **subset** of B or that B **contains** A and write $B \supset A$. If $A \subset B$ and B is not contained in A, we write $A \subset\subset B$ (A is *strictly contained* in B); if $A \subset B$ and $B \subset A$, we write $A = B$ (A is *equal* to B). In this case $a \in A$ is equivalent to $a \in B$.

We write $a \notin A$ if it is not true that $a \in A$; $B \not\supset A$ if it is not true that $B \supset A$; $A \neq B$ if it is not true that $A = B$, etc. A subset of A is said to be **empty** if it contains no elements; this set is denoted by $\emptyset$. A subset of A is said to be **full** if it is equal to the set A. The set consisting of the elements $a_1, a_2, a_3, \ldots, a_n$ is denoted by $\{a_1, a_2, a_3, \ldots, a_n\}$. More generally, the set of elements of A which satisfy a property (L) is denoted by

$$\{a \mid a \in A,\ a \text{ satisfies } (L)\}.$$

EXAMPLES. The set of positive integers $\{0, 1, 2, \ldots, n, \ldots\}$ is denoted by $\mathbf{N}$ and the set of real numbers, positive or negative, is denoted by $\mathbf{R}$. We can represent $\mathbf{R}$ by means of the points on an oriented line with an origin O, such that to the number λ there corresponds the point P whose abscissa $OP = \lambda$ (Cantor's postulate); we also call $\mathbf{R}$ the **Euclidean** (or **oriented**, or **real**) line. The set of positive rational numbers can be denoted by

$$\mathbf{R}_r^+ = \{x \mid x = p / q,\ p \in \mathbf{N}, q \in \mathbf{N}, q \neq 0\}.$$

Clearly

$$\mathbf{N} \subset \mathbf{R}_r^+ \subset \mathbf{R}.$$

(1) It is assumed that the reader is familiar with the concept of real number.

In many problems it is convenient to adjoin to $\mathbf{R}$ two elements, denoted by $+\infty$ and $-\infty$, which satisfy

$$+\infty = (-\infty)(-\infty) = (+\infty)(+\infty) = \lambda+(+\infty) = \lambda-(-\infty),$$
$$-\infty = (-\infty)(+\infty) = (+\infty)(-\infty) = \lambda+(-\infty) = \lambda-(+\infty),$$
$$-\infty<\lambda<+\infty,$$
$$+(-\infty) = -(+\infty) = -\infty; \quad \frac{\lambda}{-\infty} = \frac{\lambda}{+\infty} = 0,$$

for each $\lambda \in \mathbf{R}$ and also

$$(+\infty)\lambda = \lambda(+\infty) = (-\lambda)(-\infty) = (-\infty)(-\lambda) = +\infty,$$

for each real number $\lambda>0$. The set $\mathbf{R}$ together with the elements $+\infty$ and $-\infty$ is denoted by $\hat{\mathbf{R}}$ and is called the **augmented oriented line** or the **augmented real number system.**

The following sets are called **intervals**:

$$[\lambda, \mu] = \{x \,/\, x \in \hat{\mathbf{R}},\ x \geqq \lambda,\ x \leqq \mu\} \quad \textbf{(closed interval, or segment)},$$
$$]\lambda, \mu[= \{x \,/\, x \in \hat{\mathbf{R}},\ x > \lambda,\ x < \mu\} \quad \textbf{(open interval)},$$
$$]\lambda, \mu] = \{x \,/\, x \in \hat{\mathbf{R}},\ x > \lambda,\ x \leqq \mu\} \quad \textbf{(right half-closed interval)},$$
$$[\lambda, \mu[= \{x \,/\, x \in \hat{\mathbf{R}},\ x \geqq \lambda,\ x < \mu\} \quad \textbf{(left half-closed interval)}.$$

Thus the set of real positive numbers can be denoted by $[0, +\infty[$ and the set of real strictly positive numbers can be denoted by $]0, +\infty[$.

We use certain standard abbreviatory symbols, which are virtually indispensable in simplifying the writing of formulae. For example '$a \in A \Rightarrow b \in B$' is read '$a \in A$ implies that $b \in B$' and '$a \in A \Leftrightarrow b \in B$' is read '$a \in A$ is equivalent to $b \in B$'. We use the symbols $\forall$ and $\exists$; $(\forall_A a)$ is read 'for each element a in A' and $(\exists_A a)$ is read 'there exists in A an element a such that'. For example, 'to each element x, there corresponds an element y such that $f(x, y, z) = 0$ for all z' is written:

$$(\forall x)(\exists y)(\forall z) : f(x, y, z) = 0.$$

We note that the opposite statement can be obtained by a very simple rule, which enables us to write it down automatically as follows:

$$(\exists x)(\forall y)(\exists z) : f(x, y, z) \neq 0.$$

In formulae such as these, the symbols $(\exists x)$, $(\forall y)$, $(\exists z)$ (which are called **quantifiers**) cannot, in general, be displaced. The reader is advised to practise handling these symbols so as not to lose time in understanding the arguments which are to follow.

A **one-one correspondence** between two sets A and B is a rule in which there is associated with each element a of A an element b of B, this being denoted by $a \rightarrow b$, such that for each $b \in B$ there exists one and only one $a \in A$ for which $a \rightarrow b$. In such a correspondence, the elements of A and B correspond in pairs as indicated in figure 1.

For example, if A is the set of positive integers and B is the set of even positive integers, the correspondence $n \to 2n$ is one-one.

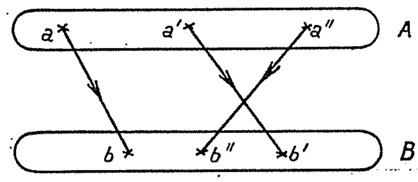


FIG. 1

In order to study a set A we often illustrate it on paper by means of a set of points which we suppose to be in one-one correspondence with A. This enables us to represent diagrammatically certain *operations* on sets.

The **Cartesian sum** A_1+A_2 of two sets A_1 and A_2 is defined to be the set formed by the pairs $(1, a_1)$, where $a_1 \in A_1$ and $(2, a_2)$, where $a_2 \in A_2$; we can represent it as indicated in figure 2 on the left-hand side. The **Cartesian product** $A_1 \times A_2$ of two sets A_1 and A_2 is defined to be the set formed by the pairs (a_1, a_2), where $a_1 \in A_1$ and $a_2 \in A_2$; we can represent it as indicated in figure 2 on the right-hand side.

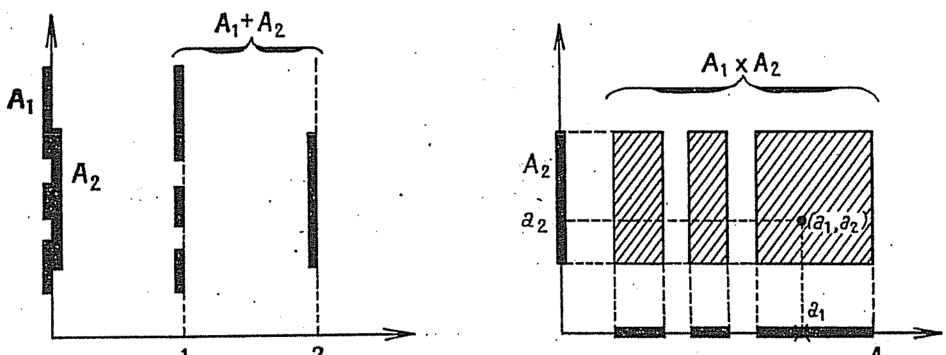


FIG. 2

EXAMPLE 1. The product $\mathbf{R} \times \mathbf{R} = \{(x_1, x_2) / x_1 \in \mathbf{R}, x_2 \in \mathbf{R}\}$, which can be represented by the points of a plane (determined by two oriented rectangular axes), is called the **Euclidean plane**, and is denoted by $\mathbf{R}^2$.

EXAMPLE 2. The product of two plane circles C_1 and C_2, which can be represented by the points of a two-dimensional torus, is called the **Euclidean torus.**

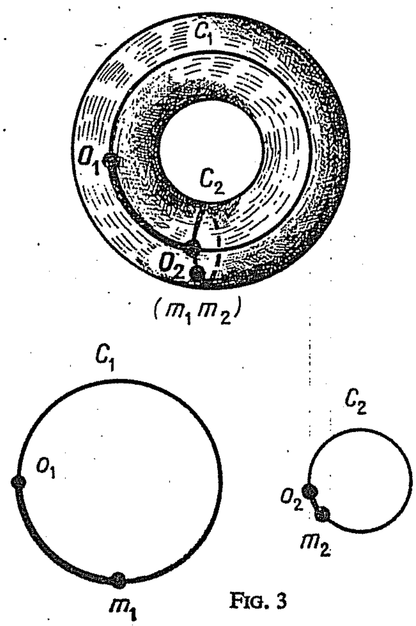


FIG. 3

§ 2. Elementary operations on sets

If A and B are two sets, their **union** $A \cup B$ is defined to be the set of elements which belong to at least one of the sets A and B; their **intersection** $A \cap B$ is defined to be the set of elements which belong to both A and B; the **difference** $A-B$ is defined to be the set of elements which belong to A and not to B.

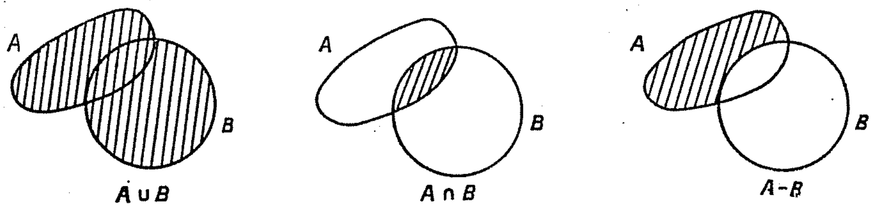


FIG. 4

If $A \cap B = \varnothing$, we say that the sets A and B are **disjoint**; if $A \cap B \neq \varnothing$, we say that the sets A and B **intersect**. If $B \subset X$, the difference $X-B$ is called the **complement** of B relative to X and is denoted by $-B$ or $\complement_X B$. The following properties are easily verified:

(1) $-\varnothing = X; \quad -X = \varnothing$,
(2) $-(-A) = A$,
(3) $A \supset B \Rightarrow -A \subset -B$,
(4) $A \cap (-A) = \varnothing; \quad A \cup (-A) = X$.

Theorem. *The operations* $\cup$ *and* $\cap$ *satisfy the following conditions*:

(1) $A \cup B = B \cup A$ (commutativity of $\cup$),
(1′) $A \cap B = B \cap A$ (commutativity of $\cap$),
(2) $A \cup (B \cup C) = (A \cup B) \cup C$ (associativity of $\cup$),
(2′) $A \cap (B \cap C) = (A \cap B) \cap C$ (associativity of $\cap$),
(3) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (distributivity of $\cup$ with respect to $\cap$),
(3′) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (distributivity of $\cap$ with respect to $\cup$),
(4) $A \cap (A \cup B) = A$ (absorption of $\cup$ by $\cap$),
(4′) $A \cup (A \cap B) = A$ (absorption of $\cap$ by $\cup$),
(5) $A \cup A = A$ (idempotence with respect to $\cup$),
(5′) $A \cap A = A$ (idempotence with respect to $\cap$).

These formulae can easily be proved; for example, (3) can be proved as follows:

$$x \in A \cap (B \cup C) \Leftrightarrow \left\{\begin{matrix} x \in A & \\ x \in B & \text{or} \quad x \in C \end{matrix}\right\}$$
$$\Leftrightarrow \{x \in A \cap B \quad \text{or} \quad x \in A \cap C\} \Leftrightarrow x \in (A \cap B) \cup (A \cap C)$$

and the other formulae can be proved by similar means.

REMARK 1. Because of formulae (2) and (2′), we can write

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

for simplicity.

REMARK 2. We see at once that from each formula such as (1) we can deduce a formula such as (1′) by making the following changes:

$$\begin{array}{lll} \emptyset & \text{for } X & \text{(and vice versa),} \\ A & \text{for } -A & \text{(and vice versa),} \\ \cup & \text{for } \cap & \text{(and vice versa),} \\ \supset & \text{for } \subset & \text{(and vice versa).} \end{array}$$

This remark is valid for every identity which involves only sets, the operations $\cup$, $\cap$, $-$ and the relations $\supset$ and $\subset$; we say that pairs of formulae such as (1) and (1′) are *duals* of one another.

If $A \subset X$, the **characteristic function** of A relative to X is a numerical function ϕ_A, defined on the elements of X by

$$\begin{array}{ll} \phi_A(x) = 1 & \text{if } x \in A, \\ \phi_A(x) = 0 & \text{if } x \in X-A. \end{array}$$

The following relations are easily proved:

$$\begin{aligned} \phi_{A \cup B}(x) &= \max\{\phi_A(x), \phi_B(x)\} = \phi_A(x)+\phi_B(x)-\phi_A(x)\,\phi_B(x), \\ \phi_{A \cap B}(x) &= \min\{\phi_A(x), \phi_B(x)\} = \phi_A(x)\,\phi_B(x), \\ \phi_{-A}(x) &= 1-\phi_A(x). \end{aligned}$$

§ 3. Families of sets

A set I and a correspondence $i \to a_i$ in which there corresponds to each i in I an element a_i of a set A, is called a **family of elements** in A and is denoted by $(a_i \mid i \in I)$; I is called the **index set**. In the case in which $I = \{1, 2, \ldots, n\}$ we have a set called an *n-tuple* and if $n=2$, this n-tuple is called a *pair*. In the case in which I is the set of strictly positive integers, we obtain a **sequence**

$$(a_1, a_2, a_3, \ldots) = (a_n).$$

A set I and a correspondence in which there corresponds to each element i of I a subset A_i of A is called a **family of sets** in A and is denoted by $\mathscr{A} = (A_i \mid i \in I)$.

Any collection $\{A, B, C, \ldots\}$ of distinct sets can be considered as a family of sets; to do this it is sufficient to take the property which characterises a set A (or even the set A itself) as the index for A. A family of sets of this type is called a **collective family**. In a collective family all the sets are different; in general, however, it may happen that two sets A_i and A_j of a family $\mathscr{A} = (A_i \mid i \in I)$ are equal although $i \neq j$. It is important not to forget that the family is defined by the A_i *and* the correspondence $i \to A_i$; the notion of 'family of sets' is more general than that of 'collection of sets'.

Let $\mathscr{A} = (A_i \mid i \in I)$ be a family of non-empty sets; we shall assume that there exists a family $(a_i \mid i \in I)$ such that $a_i \in A_i$ for each i (axiom of choice; see Chapter III).

EXAMPLE. If i denotes an interior point in a circle C and A_i the set of points on the radius through i, then the A_i form a family of sets; this is not a collective family.

Let $\mathscr{A} = (A_i \,/\, i \in I)$ and $\mathscr{B} = (B_j \,/\, j \in J)$ be two families and let C be a single set. We say that $\mathscr{A}$ is **partially contained** in the set C if $A_i \subset C$ for at least one $i \in I$; we write this $\mathscr{A} \vdash C$. We say that $\mathscr{A}$ is **finer than $\mathscr{B}$ in the exterior sense** if $\mathscr{A}$ is partially contained in every set $B_j \in \mathscr{B}$ and we write this $\mathscr{A} \vdash \mathscr{B}$. If $\mathscr{A} \vdash \mathscr{B}$ and $\mathscr{B} \vdash \mathscr{A}$, we say that $\mathscr{A}$ is **equivalent in the exterior sense** to $\mathscr{B}$ and we write $\mathscr{A} \simeq \mathscr{B}$. On the other hand, we say that $\mathscr{A}$ is **finer than $\mathscr{B}$ in the interior sense** if

$$(\forall_I\, i)\,(\exists_J\, j):\qquad A_i \subset B_j,$$

and we write $\mathscr{A} \succ \mathscr{B}$. If $\mathscr{A} \succ \mathscr{B}$ and $\mathscr{B} \succ \mathscr{A}$, we say that $\mathscr{A}$ is **equivalent in the interior sense** to $\mathscr{B}$ and we write this $\mathscr{A} \approx \mathscr{B}$.

Theorem 1. *If $\mathscr{A} = (A_i \,/\, i \in I)$ and $\mathscr{B} = (B_j \,/\, j \in J)$ are two families of sets, the family $\mathscr{A} \bigcap \mathscr{B} = (A_i \cap B_j \,/\, (i,j) \in I \times J)$ is finer than $\mathscr{A}$ and finer than $\mathscr{B}$, in both the interior and exterior senses; moreover*

$$\mathscr{C} \succ \mathscr{A},\ \mathscr{C} \succ \mathscr{A}\ \Rightarrow\ \mathscr{C} \succ \mathscr{A} \bigcap \mathscr{B}.$$

Proof. We see at once that $\mathscr{A} \bigcap \mathscr{B}$ is finer than $\mathscr{A}$, for

$$(\forall\, i)\ (\exists\, (i,j)) : A_i \cap B_j \subset A_i,$$
$$(\forall\, (i,j))\ (\exists\, i) : A_i \cap B_j \subset A_i,$$

and a similar argument applies to $\mathscr{B}$. Moreover, if $\mathscr{C} = (C_k \,/\, k \in K)$ is a family which is finer than both $\mathscr{A}$ and $\mathscr{B}$ in the interior sense, then we have

$$(\forall\, k)\,(\exists\, (i,j)) : C_k \subset A_i \cap B_j.$$

Theorem 2. *If $\mathscr{A} = (A_i \,/\, i \in I)$ and $\mathscr{B} = (B_j \,/\, j \in J)$ are two families of sets, the family $\mathscr{A} \bigcup \mathscr{B} = (A_i \cup B_j \,/\, (i,j) \in I \times J)$ is less fine than $\mathscr{A}$ and less fine than $\mathscr{B}$ in both the interior and exterior senses; moreover*

$$\mathscr{A} \vdash \mathscr{C},\ \mathscr{B} \vdash \mathscr{C}\ \Rightarrow\ \mathscr{A} \bigcup \mathscr{B} \vdash \mathscr{C}.$$

Proof. We see at once that $\mathscr{A}$ is finer than $\mathscr{A} \bigcup \mathscr{B}$, for

$$(\forall\, (i,j))\ (\exists\, i) : A_i \subset A_i \cup B_j,$$
$$(\forall\, i)\ (\exists\, (i,j)) : A_i \subset A_i \cup B_j$$

and a similar argument applies to $\mathscr{B}$. Moreover if both $\mathscr{A}$ and $\mathscr{B}$ are finer in the exterior sense than $\mathscr{C} = (C_k \,/\, k \in K)$ then we have

$$(\forall k)\ (\exists\,(i,j)) : A_i \cup B_j \subset C_k,$$

whence

$$\mathscr{A} \bigcup \mathscr{B} \vdash \mathscr{C}.$$

For collective families (and for no others), $\mathscr{A} \cup \mathscr{B}$ denotes the collection of sets which are in either $\mathscr{A}$ or $\mathscr{B}$; likewise $\mathscr{A} \cap \mathscr{B}$ denotes the collection of sets which are in both $\mathscr{A}$ and $\mathscr{B}$. The larger symbols $\bigcup$ and $\bigcap$ used in the above theorems are thus necessary to avoid confusion.

§ 4. Operations in a family of sets

The operations $\cup$ and $\cap$ defined in § 2 for two sets A and B extend easily to a family of sets $(A_i \,/\, i \in I)$. The **union** of the A_i, which is written $\bigcup_{i\in I} A_i$, is defined to be the set of elements which belong to at least one A_i. The **intersection** of the A_i, which is written $\bigcap_{i\in I} A_i$, is defined to be the set of elements which belong to all the A_i. The **product** of the A_i, which we write $\prod_{i\in I} A_i$, is the set of families of elements $\xi = (a_i \,/\, i \in I)$ such that $a_i \in A_i$ for each i. The **sum** of the A_i, which we write $\sum_{i\in I} A_i$, is the set of pairs (i, a) such that $a \in A_i$ and $i \in I$. We note that if the family consists of only two sets A and B, then these operations reduce to those defined in § 2.

EXAMPLE. The product

$$\mathbf{R}\times\mathbf{R}\times\ldots\times\mathbf{R} = \mathbf{R}^n = \{(x_1, x_2, \ldots x_n) \,/\, x_1, x_2, \ldots, x_n \in \mathbf{R}\},$$

which is the set of n-tuples of real numbers, is called **Euclidean space of n dimensions.** By using three rectangular axes, $\mathbf{R}^3$ can be identified with the ordinary space of elementary geometry.

Theorem 1. *For two families* $\mathscr{A} = (A_i \,/\, i \in I)$ *and* $\mathscr{B} = (B_j \,/\, j \in J)$ *we have*

$$(1)\quad -\bigcup_{i\in I} A_i = \bigcap_{i\in I} (-A_i),$$

$$(1')\quad -\bigcap_{i\in I} A_i = \bigcup_{i\in I} (-A_i),$$

$$(2)\quad \Big(\bigcup_{i\in I} A_i\Big) \cap \Big(\bigcup_{j\in J} B_j\Big) = \bigcup_{\substack{i\in I\\ j\in J}} (A_i \cap B_j),$$

$$(2')\quad \Big(\bigcap_{i\in I} A_i\Big) \cup \Big(\bigcap_{j\in J} B_j\Big) = \bigcap_{\substack{i\in I\\ j\in J}} (A_i \cup B_j).$$

The verification of these formulae is immediate and is left as an exercise to the reader.

Theorem 2. *If two families* $\mathscr{A} = (A_i \,/\, i \in I)$ *and* $\mathscr{B} = (B_i \,/\, i \in I)$ *have the same index set, then*

$$(1)\ (\forall i) : A_i \neq \emptyset \ \Leftrightarrow\ \prod A_i \neq \emptyset,$$

$$(2)\ \left\{\begin{matrix}(\forall i) : A_i \neq \emptyset \\ \prod\limits_{i\in I} A_i \subset \prod\limits_{i\in I} B_i\end{matrix}\right\} \ \Leftrightarrow\ (\forall i) : \emptyset \subset\subset A_i \subset B_i,$$

$$(3)\ \prod_{i\in I} A_i \cap \prod_{i\in I} B_i = \prod_{i\in I} (A_i \cap B_i),$$

$$(4)\ \prod_{i\in I} A_i \cup \prod_{i\in I} B_i \subset \prod_{i\in I} (A_i \cup B_i).$$

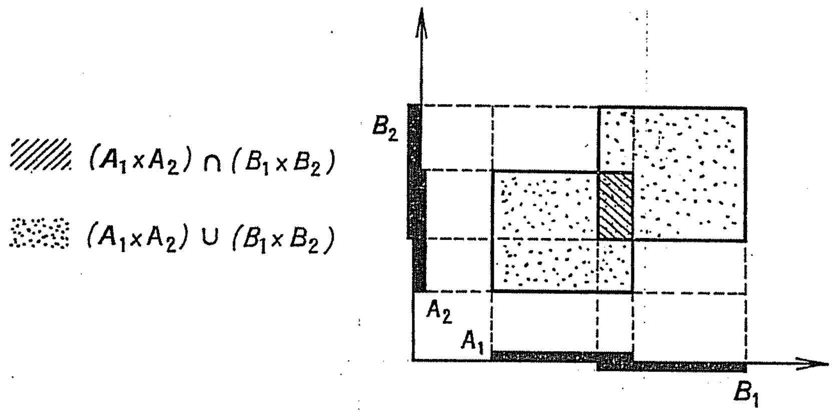


FIG. 5

These formulae can be proved immediately if we use the axiom of choice (§ 3).

§ 5. Partitions

We shall now study some important families of sets. A family $\mathscr{A} = (A_i \,/\, i \in I)$ is called a **partition** of the set A if

$$(1)\quad (\forall i) : A_i \neq \emptyset,\ A_i \subset A,$$
$$(2)\quad i \neq j \ \Rightarrow\ A_i \cap A_j = \emptyset,$$
$$(3)\quad \bigcup_{i\in I} A_i = A.$$

In other words, every point of A belongs to one and only one A_i.

A family $\mathscr{B} = (B_j \,/\, j \in J)$ is called a **sub-partition** of the partition $\mathscr{A}$ if it is a partition and if $\mathscr{B} \succ \mathscr{A}$; because of the latter condition and the fact that $\mathscr{A}$ is a partition we have

$$B_j \cap A_i \neq \emptyset \ \Rightarrow\ B_j \subset A_i.$$

Let $\mathscr{A}$ and $\mathscr{B}$ be partitions and let K be the subset of $I \times J$ defined by

$$(i,j) \in K \ \Leftrightarrow\ A_i \cap B_j \neq \emptyset.$$

We then write:

$$\mathscr{A} \sqcap \mathscr{B} = (A_i \cap B_j \,/\, (i,j) \in K).$$

Theorem 1. *$\mathscr{A} \bigcap \mathscr{B}$ is a sub-partition of both $\mathscr{A}$ and $\mathscr{B}$; every sub-partition of both $\mathscr{A}$ and $\mathscr{B}$ is a sub-partition of $\mathscr{A} \bigcap \mathscr{B}$.*

Proof. $\mathscr{A} \bigcap \mathscr{B}$ is a partition, because

(1) $A_i \cap B_j \neq \emptyset$, since $(i,j) \in K$,
(2) $(i,j) \neq (i',j')$ implies that
$$(A_i \cap B_j) \cap (A_{i'} \cap B_{j'}) = (A_i \cap A_{i'}) \cap (B_j \cap B_{j'}) = \emptyset,$$
(3) $\bigcup_{(i,j) \in K} (A_i \cap B_j) = \bigcup_{i \in I} A_i \cap \bigcup_{j \in J} B_j = A \cap A = A.$

To prove the other part of the theorem, we observe that

$$\mathscr{C} \succ \mathscr{A},\ \mathscr{C} \succ \mathscr{B} \quad \Rightarrow \quad \mathscr{C} \succ \mathscr{A} \bigcap \mathscr{B}.$$

Theorem 2. *If $\mathscr{A} = (A_i \mid i \in I)$ is a partition of A and $\mathscr{B} = (B_j \mid j \in J)$ is a partition of B, the family $\mathscr{A} \, \Pi \, \mathscr{B} = (A_i \times B_j \mid (i,j) \in I \times J)$ is a partition of the set $A \times B$.*

Proof. Using Theorem 2 of § 4, we have

(1) $A_i \times B_j \neq \emptyset; \qquad A_i \times B_j \subset A \times B,$

(2) $(i,j) \neq (i',j') \Rightarrow \left\{ \begin{matrix} A_i \cap A_{i'} = \emptyset \\ \text{or} \\ B_j \cap B_{j'} = \emptyset \end{matrix} \right\} \Rightarrow \begin{aligned} &(A_i \times B_j) \cap (A_{i'} \times B_{j'}) \\ &= (A_i \cap A_{i'}) \times (B_j \cap B_{j'}) = \emptyset, \end{aligned}$

(3) $\bigcup_{(i,j) \in I \times J} A_i \times B_j = \bigcup_{i \in I} (A_i \times \bigcup_{j \in J} B_j) = \bigcup_{i \in I} A_i \times B = A \times B.$

The idea of a partition can easily be generalised. A family $\mathscr{A} = (A_i \mid i \in I)$ is called a **covering** of A if

(1) $(\forall i) : A_i \neq \emptyset,$
(2) $\bigcup_{i \in I} A_i = A.$

In other words, each element of A belongs to at least one A_i. The *order* of a covering $\mathscr{A}$ is the greatest integer m for which there exist $m+1$ sets of $\mathscr{A}$ having a non-empty intersection; a partition is a covering of order 0. If $\mathscr{A}$ is a covering of A and $\mathscr{B}$ is a family such that $\mathscr{A} \succ \mathscr{B}$, it is clear that $\mathscr{B}$ is also a covering of A and we say that $\mathscr{A}$ is a **sub-covering** of $\mathscr{B}$. There are analogues for coverings of the above theorems for partitions.

§ 6. Filter bases

A family $\mathscr{B} = (B_i \mid i \in I)$ is called a **filter base**[1] (or just a **base**) if

(1) $(\forall i) : B_i \neq \emptyset,$
(2) $(\forall i)(\forall j)(\exists k) : B_k \subset B_i \cap B_j.$

[1] The idea of a filter base is due to H. Cartan (*C.R. Acad. Sc.*, 1937, vol. 205, p. 595).

A family $\mathscr{B}' = (B_j' \,/\, j \in J)$ is called a **sub-base** of $\mathscr{B}$ if it is a base and if $\mathscr{B}' \vdash \mathscr{B}$; in other words

$$(\forall\, i)\,(\exists\, j) : B_j' \subset B_i.$$

EXAMPLE. Let N be the set of integers. If

$$S_m = \{n \,/\, n \in \mathrm{N};\ n \geqq m\}$$

denotes the section beginning at m, the family $\mathscr{S} = (S_m \,/\, m \in \mathrm{N})$ is a filter base in N. This is called the **Fréchet base.** The base $\mathscr{S}' = (S_{2p} \,/\, p \in \mathrm{N})$ is a base such that $\mathscr{S}' \vdash \mathscr{S}$ and $\mathscr{S} \vdash \mathscr{S}'$; thus we have $\mathscr{S}' \simeq \mathscr{S}$.

Theorem 1. *If $\mathscr{B}$ is a filter base and if $\mathscr{C} = (C_j \,/\, j \in J)$ is any family such that $\mathscr{C} \simeq \mathscr{B}$, then $\mathscr{C}$ is also a filter base.*

Proof. Since $\mathscr{B} \vdash \mathscr{C}$, we have $C_j \neq \varnothing$; moreover, for each pair of indices j and j' there exist indices i and i' such that

$$B_i \subset C_j; \quad B_{i'} \subset C_{j'}.$$

In $\mathscr{B}$, there exists a set B such that

$$B \subset B_i \cap B_{i'} \subset C_j \cap C_{j'}.$$

Since $\mathscr{C} \vdash \mathscr{B}$, there exists a set C belonging to $\mathscr{C}$ and contained in B. Therefore $C \subset C_j \cap C_{j'}$ and the theorem is proved.

Theorem 2. *If $\mathscr{A} = (A_i \,/\, i \in I)$ and $\mathscr{B} = (B_j \,/\, j \in J)$ are two filter bases, the family $\mathscr{A} \bigcup \mathscr{B} = (A_i \cup B_j \,/\, (i,j) \in I \times J)$ is a filter base.*

Proof. Clearly $A_i \cup B_j \neq \varnothing$. Let $A_i \cup B_j$ and $A_{i'} \cup B_{j'}$ be two sets in $\mathscr{A} \bigcup \mathscr{B}$. There exist sets A and B in $\mathscr{A}$ and $\mathscr{B}$ respectively such that

$$A \subset A_i \cap A_{i'},$$
$$B \subset B_j \cap B_{j'},$$

and we therefore have

$$A \cup B \subset (A_i \cap A_{i'}) \cup (B_j \cap B_{j'}) =$$
$$= (A_i \cup B_j) \cap (A_{i'} \cup B_{j'}) \cap (A_i \cup B_{j'}) \cap (A_{i'} \cup B_j) \subset (A_i \cup B_j) \cap (A_{i'} \cup B_{j'}).$$

Theorem 3. *If $\mathscr{A}$ and $\mathscr{B}$ are two bases such that $A_i \cap B_j \neq \varnothing$ for each i in I and each j in J, the family $\mathscr{A} \bigcap \mathscr{B} = (A_i \cap B_j \,/\, (i,j) \in I \times J)$ is a base.*

Proof. Let $A_i \cap B_j$ and $A_{i'} \cap B_{j'}$ be two sets in $\mathscr{A} \bigcap \mathscr{B}$; then there exist sets A and B in $\mathscr{A}$ and $\mathscr{B}$ respectively such that

$$A \subset A_i \cap A_{i'}, \qquad B \subset B_j \cap B_{j'}.$$

We therefore have

$$A \cap B \subset (A_i \cap A_{i'}) \cap (B_j \cap B_{j'}) = (A_i \cap B_j) \cap (A_{i'} \cap B_{j'}).$$

COROLLARY. *If $\mathscr{B} = (B_i \mid i \in I)$ is a base such that $B_i \cap A \neq \varnothing$ for each i, the family $\mathscr{C} = (B_i \cap A \mid i \in I)$ is a finer filter base than $\mathscr{B}$.*

This follows at once, for if we put $\mathscr{A} = \{A\}$, then we have

$$\mathscr{C} = \mathscr{A} \textstyle\bigcap \mathscr{B}.$$

Theorem 4. *If $\mathscr{A} = (A_i \mid i \in I)$ and $\mathscr{B} = (B_j \mid j \in J)$ are filter bases in the sets A and B, the family $\mathscr{A} \,\Pi\, \mathscr{B} = (A_i \times B_j \mid (i, j) \in I \times J)$ is a filter base in the set $A \times B$.*

Proof. Clearly $A_i \times B_j \neq \varnothing$. Given two sets $A_i \times B_j$ and $A_{i'} \times B_{j'}$ there exist sets A and B in $\mathscr{A}$ and $\mathscr{B}$ respectively such that

$$A \subset A_i \cap A_{i'},$$
$$B \subset B_j \cap B_{j'}.$$

Hence, by Theorem 2 of § 4,

$$A \times B \subset (A_i \cap A_{i'}) \times (B_j \cap B_{j'}) = (A_i \times B_j) \cap (A_{i'} \times B_{j'}).$$

Let $\mathscr{A} = (A_i \mid i \in I)$ be any family of sets. The **grill** $\ddot{\mathscr{A}}$ is defined to be the collection of sets which meet all the A_i. The grill $\ddot{\mathscr{B}}$ of a filter base $\mathscr{B}$ possesses certain interesting properties; in particular it is clear that each set in $\mathscr{B}$ belongs to $\ddot{\mathscr{B}}$.

Theorem 5. *If $\mathscr{B}$ is a filter base, the following properties are equivalent:*

(1) *for any set A, $\mathscr{B} \vdash A$ or $\mathscr{B} \vdash (-A)$,*
(2) *$\mathscr{B} \simeq \ddot{\mathscr{B}}$,*
(3) *for each base $\mathscr{B}'$ such that $\mathscr{B}' \vdash \mathscr{B}$, we have $\mathscr{B}' \simeq \mathscr{B}$.*

If a filter base satisfies one of these properties, we say that it is an **ultra-filter base.**

Proof of Theorem 5. We first show that (1) implies (2). Since $\mathscr{B} \subset \ddot{\mathscr{B}}$, we also have $\ddot{\mathscr{B}} \vdash \mathscr{B}$. If $\mathscr{B}$ is not partially contained in $\ddot{\mathscr{B}}$, there exists a set H in $\ddot{\mathscr{B}}$ which does not contain any B_i and so, by (1), $\mathscr{B} \vdash (-H)$. Therefore there exists a set B in $\mathscr{B}$ which does not meet H, which contradicts $H \in \ddot{\mathscr{B}}$.

We next show that (2) implies (3). If a filter base $\mathscr{B}' = (B'_j \mid j \in J)$ is such that $\mathscr{B}' \vdash \mathscr{B}$, then each B'_j meets each B_i, so that $\mathscr{B}' \subset \ddot{\mathscr{B}}$. Since $\mathscr{B} \vdash \ddot{\mathscr{B}}$, we have

$$(\forall j)\,(\exists i) : B_i \subset B'_j.$$

We therefore have $\mathscr{B} \vdash \mathscr{B}'$, whence $\mathscr{B} \simeq \mathscr{B}'$.

Finally we show that (3) implies (1). If $\mathscr{B}$ is not partially contained in A, we have

$$(\forall i) : B_i \cap (-A) \neq \varnothing.$$

Then $\mathscr{C} = (B_i \cap (-A) \,/\, i \in I)$ is a filter base, by the corollary to Theorem 3. Since $\mathscr{C} \vdash \mathscr{B}$, we have $\mathscr{C} \simeq \mathscr{B}$ and so $\mathscr{B} \vdash \mathscr{C}$, whence

$$(\forall\, i)\,(\exists\, j) : B_j \subset B_i \cap (-A).$$

Therefore

$$\mathscr{B} \vdash (-A).$$

EXAMPLE. Let X be a set and let x_0 be one of its elements; the collective family of all the subsets of X which contain x_0 is an ultra-filter base, for

$$x_0 \in A \quad \Rightarrow \quad \mathscr{B} \vdash A,$$
$$x_0 \notin A \quad \Rightarrow \quad \mathscr{B} \vdash (-A).$$

Theorem 6. *If $\mathscr{B}$ is a filter base, there exists an ultra-filter base $\mathscr{B}_0$ such that $\mathscr{B}_0 \vdash \mathscr{B}$.*

This result will be proved later, as an application of the theory of ordering relations (Chapter III, § 6).

§ 7. Closure operations in a set

If with each subset A of a set X we associate a subset $\bar{A}$ of X such that

(1) $\bar{A} \supset A$,
(2) $A \supset B \quad \Rightarrow \quad \bar{A} \supset \bar{B}$,
(3) $\overline{(\bar{A})} = \bar{A}$,
(4) $\bar{\varnothing} = \varnothing$,

then we call the correspondence $A \to \bar{A}$ a **closure operation.**(1)

In a similar manner, a correspondence $A \to \mathring{A}$ is called an **interior operation** if

(1) $\mathring{A} \subset A$,
(2) $A \supset B \quad \Rightarrow \quad \mathring{A} \supset \mathring{B}$,
(3) $(\mathring{\mathring{A}}) = \mathring{A}$,
(4) $\mathring{X} = X$.

The sets F such that $F = \bar{F}$ are said to be **closed** and the sets G such that $G = \mathring{G}$ are said to be **open**, with respect to the appropriate operations in each case.

EXAMPLE. If X denotes the real line $\mathbf{R}$ and $\bar{A}$ the smallest segment $[a, b]$ which contains all the points of the set A, then the correspondence $A \to \bar{A}$ is a closure operation.

Theorem 1. *A one-one correspondence can be established between closure operations and interior operations by means of the formulae:*

$$\mathring{A} = -(\overline{-A})\;; \quad \bar{A} = -(\overset{\circ}{-A})$$

(1) The idea of closure is due to E. H. Moore (*Introduction to a form of general analysis*, New Haven, 1910, p. 53).

Proof. The correspondence $A \to \mathring{A} = -(\overline{-A})$ is clearly an interior operation, for

(1) $\mathring{A} = -(\overline{-A}) \subset -(-A) = A$,
(2) $A \subset B \Rightarrow -A \supset -B \Rightarrow (\overline{-A}) \supset (\overline{-B}) \Rightarrow \mathring{A} \subset \mathring{B}$,
(3) $\mathring{\mathring{A}} = -(\overline{\overline{-A}}) = \mathring{A}$,
(4) $\mathring{X} = -\overline{\varnothing} = X$.

In a similar manner, we can prove that the correspondence $A \to \bar{A} = -(\mathring{-A})$ is a closure operation. It is easily shown that the correspondence thus set up between the two kinds of operations is one-one.

REMARK. The complement of an open set is a closed set (and vice versa), for $-F = -\bar{F} = (\mathring{-F})$; in what follows, we shall study only properties of closure operations, those of interior operations being deducible immediately on appealing to duality.

Theorem 2. *The intersection of a family $\mathscr{F} = (F_i \mid i \in I)$ of closed sets is a closed set.*

Proof. Writing $F = \bigcap_{i \in I} F_i$, we have, by condition (2) for a closure operation,

$$\bar{F} \subset \bar{F}_i = F_i$$

for each i. Therefore $\bar{F} \subset F$; hence, by condition (1), $\bar{F} = F$.

Theorem 3. *The closure $\bar{A}$ of A is the intersection of all the closed sets which contain A.*

Proof. Let E be the intersection of all the closed sets which contain A; by the preceding theorem, E is closed. Since $\bar{A}$ is a closed set which contains A, we have $E \subset \bar{A}$. But we also have $E = \bar{E} \supset \bar{A}$, by condition (2), whence $E = \bar{A}$.

Theorem 4. *If $A \to u[A]$ and $A \to v[A]$ are two closure operations and if $vu[A] = v[u[A]]$ is closed in the u-sense, the correspondence $A \to vu[A]$ is a closure operation in which the set corresponding to A is the intersection of all the sets closed in both the u- and v-senses and containing A; moreover $vu[A] \supset uv[A]$.*

Proof. $A \to vu[A]$ is a closure operation, since

(1) $vu[A] \supset v[A] \supset A$,
(2) $A \supset B \Rightarrow u[A] \supset u[B] \Rightarrow vu[A] \supset vu[B]$,
(3) $vuvu[A] = vvu[A] = vu[A]$.

If a set F is closed in the u-sense and in the v-sense, we have

$$F \supset A \Rightarrow u[F] = F \supset u[A] \Rightarrow v[F] = F \supset vu[A].$$

Since $vu[A]$ is closed in the u-sense and the v-sense, it is therefore the intersection of all the sets closed in both senses which contain A.

For the last part, we have

$$A \subset vu[A] \Rightarrow v[A] \subset vu[A] \Rightarrow uv[A] \subset vu[A].$$

EXAMPLE. We discuss certain closure operations in the space $\mathbf{R}^3$, the familiar space of elementary three-dimensional geometry.

A set A is said to be **starred** or **star-like** if $a \in A$ and $\lambda \in [0, 1]$ imply that $\lambda a \in A$. A set A' is said to be **haloed** if $a \in A'$ and $\lambda \in [1, +\infty[$ imply that $\lambda a \in A'$. A set A'' is said to be a **cone** if it is simultaneously starred and haloed.

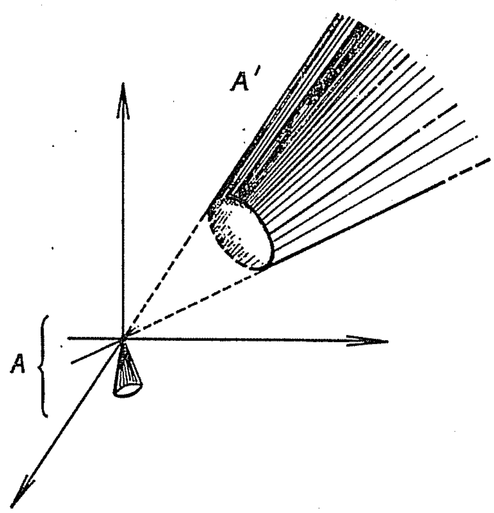


FIG. 6

We define the **starred closure** of a set B to be

$$e[B] = \{x \,/\, x = \lambda b;\ b \in B,\ \lambda \in [0, 1]\},$$

the **haloed closure** to be

$$a[B] = \{x \,/\, x = \lambda b;\ b \in B,\ \lambda \in [1, +\infty[\},$$

and the **conical closure** to be

$$c[B] = \{x \,/\, x = \lambda b;\ b \in B;\ \lambda \in [0, +\infty[\}.$$

The set $ae[B]$ is a cone of vertex 0 and so is starred; we have

$$ae[B] = ea[B] = c[B].$$

§ 8.* Lattices of sets

We say that a family $\mathscr{T}$ is a **lattice**[1] with respect to the operations $\cup$ and $\cap$ if

$$A \in \mathscr{T},\ B \in \mathscr{T} \quad \Rightarrow \quad A \cup B \in \mathscr{T},\ A \cap B \in \mathscr{T}.$$

We say that a lattice $\mathscr{T} = (A_i \,/\, i \in I)$ is **complete** with respect to the operation $\cup$ if

$$J \subset I \quad \Rightarrow \quad \bigcup_{i \in J} A_i \in \mathscr{T}.$$

In a similar manner we can define completeness of a lattice with respect to the operation $\cap$. For sets in a lattice we have the following properties:

$$\left.\begin{array}{ll}(1) & A \cup B = B \cup A \\ (1') & A \cap B = B \cap A\end{array}\right\} \quad \text{(commutativity)},$$

$$\left.\begin{array}{ll}(2) & A \cup (B \cup C) = (A \cup B) \cup C \\ (2') & A \cap (B \cap C) = (A \cap B) \cap C\end{array}\right\} \quad \text{(associativity)},$$

$$\left.\begin{array}{ll}(3) & A \cup (A \cap B) = A \\ (3') & A \cap (A \cup B) = A\end{array}\right\} \quad \text{(absorption)},$$

$$\left.\begin{array}{ll}(4) & A \cup A = A \\ (4') & A \cap A = A\end{array}\right\} \quad \text{(idempotence)}.$$

More generally let $\mathscr{T} = (A_i \,/\, i \in I)$ be a family of sets; then an **operation** $\wedge$ in $\mathscr{T}$ is a law in which there corresponds to each ordered pair of sets A_i and A_j a third set of $\mathscr{T}$, which we write

$$A_k = A_i \wedge A_j.$$

Given two operations $\vee$ and $\wedge$ in $\mathscr{T}$, we say that $\mathscr{T}$ is a lattice with respect to these two operations if we have the above four properties: commutativity, associativity, absorption and idempotence, with $\vee$ and $\wedge$ in place of $\cup$ and $\cap$.

An interesting example of a lattice is obtained from a closure operation; we prove the following result.

Theorem. *Given a closure operation, the family $\mathscr{F} = (F_i \,/\, i \in I)$ of closed sets is a lattice with respect to the operations $\vee$ and $\wedge$ defined by*

$$F_i \vee F_j = \overline{F_i \cup F_j}; \quad F_i \wedge F_j = F_i \cap F_j.$$

Proof. Evidently we have $F_i \vee F_j \in \mathscr{F}$, $F_i \wedge F_j \in \mathscr{F}$; we shall prove the four properties of a lattice for one of these operations, namely $\vee$, the proof for the other being similar.

(1) Commutativity:

$$F_i \vee F_j = \overline{F_j \cup F_i} = F_j \vee F_i.$$

[1] The idea of a lattice is due to O. Ore and Garrett Birkhoff.

(2) Associativity: we put

$$A = F_i \vee (F_j \vee F_k) = \overline{F_i \cup \overline{(F_j \cup F_k)}},$$
$$B = (F_i \vee F_j) \vee F_k = \overline{\overline{(F_i \cup F_j)} \cup F_k}$$

and prove that $A = B$. A is a closed set which contains F_i and $F_j \cup F_k$; therefore $A \supset \overline{F_i \cup F_j}$ and so

$$A \supset \overline{\overline{(F_i \cup F_j)} \cup F_k} = B.$$

In a similar manner we can show that $A \subset B$, and so we have $A = B$.

(3) Absorption: write

$$A = F_i \wedge (F_i \vee F_j) = F_i \cap \overline{(F_i \cup F_j)}.$$

Then $A \subset F_i$ and $A \supset F_i$, whence $A = F_i$.

(4) Idempotence: we have

$$F_i \vee F_i = \overline{F_i \cup F_i} = \overline{F}_i = F_i.$$

Various kinds of lattices play an important role in analysis. A lattice $\mathscr{T}$ is said to be **distributive** if any three sets A, B, C satisfy

$$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C).$$

A lattice $\mathscr{T}$ is said to be **modular** if

$$A \wedge [B \vee (A \wedge C)] = (A \wedge B) \vee (A \wedge C).$$

We say that a closure operation is distributive if the lattice of closed sets is distributive and that it is modular if the lattice of closed sets is modular.

We remark that if a lattice is distributive, it is also modular, for, because of the distributivity, we have

$$A \wedge [B \vee (A \wedge C)] = (A \wedge B) \vee [A \wedge (A \wedge C)] = (A \wedge B) \vee (A \wedge C).$$

EXAMPLE. Let X be the three-dimensional space $\mathbf{R}^3$. We call the following sets **linear (affine) varieties**: the empty set $\varnothing$, each set $\{a\}$ consisting of a single point a, each line D, each plane P and the whole space $\mathbf{R}^3$.

It is clear that the intersection of a set of linear varieties is a linear variety. If A is a set and we denote by $\bar{A}$ the intersection of all the linear varieties which contain A, then the correspondence $A \to \bar{A}$ is clearly a closure operation, for

(1) $\bar{A} \supset A$,

(2) $A \supset B \quad \Rightarrow \quad \bar{A} \supset \bar{B}$,

(3) $\overline{(\bar{A})} = \bar{A}$,

(4) $\bar{\varnothing} = \varnothing$.

This is not a distributive closure operation, as figure 7 shows: P is a plane, D is a straight line, $A = \{a\}$ is a point not on the plane or the line; $P \wedge (A \vee D)$ is a straight line, although

$$(P \wedge A) \vee (P \wedge D) = \emptyset \vee (P \wedge D) = P \wedge D$$

is a point.

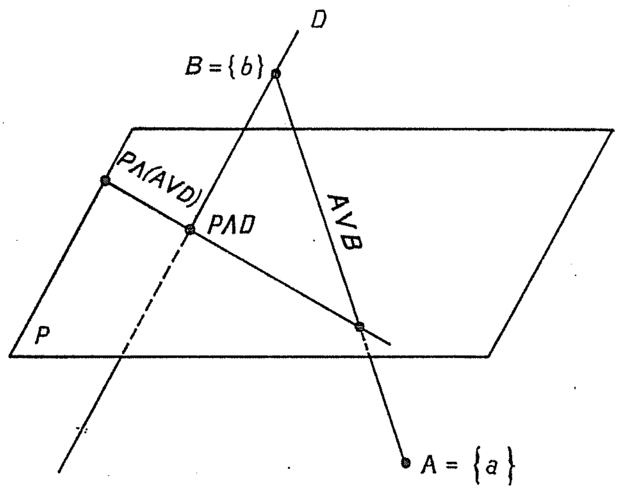


FIG. 7

On the other hand, we shall show that this closure operation is modular. Suppose that

$$a \in A \wedge [B \vee (A \wedge C)].$$

Then, in particular, $a \in B \vee (A \wedge C)$, which shows that either

(1) $a \in B$, or

(2) $a \in A \cap C$, or

(3) there exist points $b \in B$ and $c \in A \cap C$ such that a lies on the line joining b and c. We now prove that

$$a \in (A \wedge B) \vee (A \wedge C) = \overline{(A \cap B) \cup (A \cap C)}.$$

In case (1), this result holds because $a \in A \cap B$. In case (2) it holds because $a \in A \cap C$. In case (3), we can argue as follows: since A is closed and contains a and c, then it contains the whole line bc; therefore $b \in A \cap B$ and $c \in A \cap C$, which imply that $a \in \overline{(A \cap B) \cup (A \cap C)}$.

Conversely, suppose that $a \in \overline{(A \cap B) \cup (A \cap C)}$. Then it is clear that $a \in A$ (since A is closed); also

$$a \in \overline{(A \cap B) \cup (A \cap C)} \subset \overline{B \cup (A \cap C)}$$

and hence

$$a \in A \cap \overline{[B \cup (A \cap C)]} = A \wedge [B \vee (A \wedge C)].$$

Thus we have proved that

$$A \wedge [B \vee (A \wedge C)] = (A \wedge B) \vee (A \wedge C)$$

as required.

REMARK 1. We can also express the fact that a lattice is distributive by means of the relation

$$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C).$$

For if the lattice is distributive, we have

$$(A \vee B) \wedge (A \vee C) = [(A \vee B) \wedge A] \vee [(A \vee B) \wedge C].$$

Because of the absorption property, the right-hand side is equal to $A \vee [(A \vee B) \wedge C]$ and therefore to $A \vee [(A \wedge C) \vee (B \wedge C)]$, since the lattice is distributive. By the associative property, this is the same as $[A \vee (A \wedge C)] \vee (B \wedge C)$ which, again using absorption, is equal to $A \vee (B \wedge C)$.

Conversely, if the condition is satisfied, then we can prove that the lattice is distributive by interchanging $\vee$ and $\wedge$ in the above argument.

REMARK 2. We can also express the fact that a lattice is modular by means of the relation

$$A \vee [B \wedge (A \vee C)] = (A \vee B) \wedge (A \vee C).$$

In fact we have, using the commutativity and absorption properties,

$$(A \vee B) \wedge (A \vee C) = (A \vee C) \wedge [B \vee ((A \vee C) \wedge A)],$$

which is equal to

$$[(A \vee C) \wedge B] \vee [(A \vee C) \wedge A] = [(A \vee C) \wedge B] \vee A,$$

since the lattice is modular. To prove the converse, it is sufficient to interchange $\vee$ and $\wedge$ in the above argument.

§ 9. Principal limits of a family of sets

In this section we introduce the idea of **principal limits** of a family of sets. We first consider a sequence of sets:

$$(A_n) = (A_1, A_2, \ldots)$$

We define the **superior (principal) limit** of the sequence (A_n) to be the set of points which belong to an infinity of the A_i; we write this $\overline{\operatorname{Lim}}_{n=\infty} (A_n)$. We define the **inferior (principal) limit** of the sequence (A_n) to be the set of points which belong to all the A_i from a certain value of the suffix onwards; we write this $\underline{\operatorname{Lim}}_{n=\infty} (A_n)$. Clearly

$$\underline{\operatorname{Lim}}_{n=\infty} (A_n) \subset \overline{\operatorname{Lim}}_{n=\infty} (A_n).$$

If $\underline{\mathrm{Lim}}_{n=\infty} (A_n) = \overline{\mathrm{Lim}}_{n=\infty} (A_n)$ we say that this set is the **(principal) limit** of the sequence (A_n) and we denote it by $\mathrm{Lim}_{n=\infty} (A_n)$.

Not every sequence of sets admits a limit; on the other hand, in the two following cases the sequences do admit limits (which can be $\emptyset$):

(1) $A_1 \supset A_2 \supset A_3 \supset \ldots$ (decreasing sequence),
(2) $A_1 \subset A_2 \subset A_3 \subset \ldots$ (increasing sequence).

The above ideas can be expressed differently in terms of the sections $S_p = (n \,/\, n \in \mathrm{N}, n \geqq p)$; we have

$$a \in \overline{\mathrm{Lim}}_{n=\infty} (A_n) \Leftrightarrow (\forall\, p)\ (\exists_{S_p} n) : a \in A_n,$$
$$a \in \underline{\mathrm{Lim}}_{n=\infty} (A_n) \Leftrightarrow (\exists\, p)\ (\forall_{\,S_p} n) : a \in A_n.$$

More generally, let $(A_i) = (A_i \,/\, i \in I)$ be a family of sets and let $\mathscr{B} = (S_p \,/\, p \in P)$ be a filter base on I; then the set of elements a which satisfy

$$(\forall_{\mathscr{B}}\, S)\ (\exists_S\, i) : a \in A_i$$

is called the **superior limit of** (A_i) **with respect to** $\mathscr{B}$. The set of elements a which satisfy

$$(\exists_{\mathscr{B}}\, S)\ (\forall_S\, i) : a \in A_i$$

is called the **inferior limit of** (A_i) **with respect to** $\mathscr{B}$. These sets are denoted by $\overline{\mathrm{Lim}}_{\mathscr{B}} (A_i)$ and $\underline{\mathrm{Lim}}_{\mathscr{B}} (A_i)$.

Theorem. *The superior and inferior limits of a family satisfy*

$$\overline{\mathrm{Lim}}_{\mathscr{B}} (A_i) = \bigcap_{S \in \mathscr{B}} \Big(\bigcup_{i \in S} A_i\Big),$$
$$\underline{\mathrm{Lim}}_{\mathscr{B}} (A_i) = \bigcup_{S \in \mathscr{B}} \Big(\bigcap_{i \in S} A_i\Big).$$

Proof. If $a \in \overline{\mathrm{Lim}}_{\mathscr{B}} (A_i)$, then, for each set S of $\mathscr{B}$, we have

$$a \in \bigcup_{i \in S} A_i$$

and conversely. If $a \in \underline{\mathrm{Lim}}_{\mathscr{B}} (A_i)$, then, for some set S of $\mathscr{B}$, we have

$$a \in \bigcap_{i \in S} A_i$$

and conversely.

In the case in which (A_n) is a sequence of sets and the base $\mathscr{B}$ is the family of sections S_p, the above formulae become

$$\overline{\mathrm{Lim}}_{n=\infty} (A_n) = \bigcap_{p=1}^{\infty} (A_p \cup A_{p+1} \cup \ldots),$$

$$\underline{\mathrm{Lim}}_{n=\infty} (A_n) = \bigcup_{p=1}^{\infty} (A_p \cap A_{p+1} \cap \ldots).$$

CHAPTER II

MAPPINGS OF ONE SET INTO ANOTHER

§ 1. Single-valued, semi-single-valued and multi-valued mappings

Let X and Y be two sets. If with each element x of X we associate a subset $\Gamma(x)$ of Y, we say that the correspondence $x \to \Gamma(x)$ is a **mapping** of X into Y; the set $\Gamma(x)$ is called the **image** of x under the mapping Γ. Where no confusion is possible we shall denote this set indifferently by Γx or $\Gamma(x)$. The set $X^* = \{x \,/\, x \in X, \Gamma x \neq \emptyset\}$ is called the **domain** (or set of definition) of Γ and $Y^* = \bigcup_{x \in X} \Gamma x$ is called the **range** (or set of values) of Γ; we also say that Γ is **defined** on X^* and that it is a mapping of X **onto** Y^*.

If the mapping Γ of X into Y is such that the set Γx always consists of a single element, we say that Γ is a **single-valued function** or a **single-valued mapping** of X into Y. Single-valued mappings will usually be denoted by small Greek letters; general or **multi-valued** mappings will be denoted by capital Greek letters.

EXAMPLE. If $x \to \sigma x$ is a single-valued mapping of a set X into the line $\mathbf{R} =]-\infty, +\infty[$, we say that σ is a **numerical function** defined on X; if σ is a single-valued mapping of X into the complete line $\hat{\mathbf{R}} = [-\infty, +\infty]$, we say that it is a **generalised numerical function**; if σ is a single-valued mapping of X into the set $\mathbf{C}$ of complex numbers, we say that it is a **complex function.** In these cases, we usually denote the image of x by $\phi(x)$ or $f(x)$.

A mapping is called **semi-single-valued** if

$$\Gamma x \cap \Gamma x' \neq \emptyset \quad \Rightarrow \quad \Gamma x = \Gamma x'.$$

Clearly a single-valued mapping is also semi-single-valued.

EXAMPLE. Let X be the three-dimensional space $\mathbf{R}^3$ and let O be a fixed point in X; given $x \in X$, let Γx be the set of points other than O on the straight line joining x to O. This determines a semi-single-valued mapping Γ, whose domain is $\mathbf{R}^3 - \{O\}$.

A mapping is called **injective** if

$$x \neq x' \quad \Rightarrow \quad \Gamma x \cap \Gamma x' = \emptyset.$$

An injective mapping is evidently semi-single-valued.

If Γ is a mapping of X into Y and A is a non-empty subset of X, we write

$$\Gamma A = \bigcup_{x \in A} \Gamma x.$$

If $A = \emptyset$, we write $\Gamma\emptyset = \emptyset$. The set ΓA is called the **image** of A under the mapping Γ; if we compare x to a source of light, making a shadow Γx on a screen, then ΓA is the shadow produced by a set A of sources of light.

If $\mathscr{A} = (A_i \mid i \in I)$ is a family of sets, we write

$$\Gamma\mathscr{A} = (\Gamma A_i \mid i \in I)$$

and call $\Gamma\mathscr{A}$ the image of $\mathscr{A}$ under the mapping Γ.

Theorem 1. *$A \subset B$ implies that $\Gamma A \subset \Gamma B$.*

Proof. If $y \in \Gamma A$, then $y \in \Gamma x$ for some $x \in A$. This implies that $y \in \Gamma x$ for some $x \in B$, whence $y \in \Gamma B$.

COROLLARY. *$\mathscr{A} \succ \mathscr{B}$ implies that $\Gamma\mathscr{A} \succ \Gamma\mathscr{B}$; $\mathscr{A} \vdash \mathscr{B}$ implies that $\Gamma\mathscr{A} \vdash \Gamma\mathscr{B}$.*

These properties can be deduced at once as follows:

$$(\forall\, i)\ (\exists\, j) : A_i \subset B_j \ \Rightarrow\ (\forall\, i)\ (\exists\, j) : \Gamma A_i \subset \Gamma B_j,$$
$$(\forall\, j)\ (\exists\, i) : A_i \subset B_j \ \Rightarrow\ (\forall\, j)\ (\exists\, i) : \Gamma A_i \subset \Gamma B_j.$$

Theorem 2. *We have*

$$\Gamma\Big(\bigcup_{i \in I} A_i\Big) = \bigcup_{i \in I} \Gamma A_i.$$

Proof. If $y \in \Gamma(\bigcup A_i)$, then $y \in \Gamma x$ for some $x \in \bigcup A_i$, whence, for at least one index i_0, we have $y \in \Gamma A_{i_0}$. Therefore $y \in \bigcup \Gamma A_i$. By reversing the argument, we can deduce the desired formula.

Theorem 3. *We have*

$$\Gamma\Big(\bigcap_{i \in I} A_i\Big) \subset \bigcap_{i \in I} \Gamma A_i.$$

Proof. If $y \in \Gamma(\bigcap A_i)$, then $y \in \Gamma x$ for some x such that $x \in A_i$ for all i, whence

$$y \in \bigcap_{i \in I} \Gamma A_i$$

and so the result is proved.

Theorem 4. *If $A \subset X$ and if Γ is a mapping of X with range Y, then*

$$-\Gamma A \subset \Gamma(-A).$$

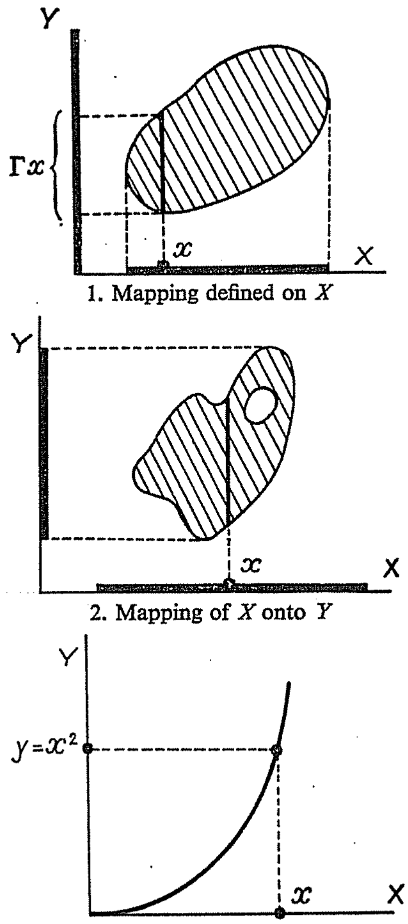
If, further, Γ is injective, then

$$-\Gamma A = \Gamma(-A).$$

Proof. If $y \in -\Gamma A$, then clearly $y \in \Gamma(-A)$, since Γ is a mapping whose range is Y. If Γ is injective, then

$$y \in \Gamma(-A) \ \Rightarrow\ (\exists_{-A}\, x) : y \in \Gamma x \ \Rightarrow\ y \notin \Gamma A \ \Rightarrow\ y \in -\Gamma A.$$

Theorem 5. *If $\mathscr{A} = (A_i \mid i \in I)$ is a partition of X and Γ is an injective mapping with domain X and range Y, the family $\Gamma\mathscr{A}$ is a partition of Y.*



1. Mapping defined on X

2. Mapping of X onto Y

3. Function $f(x) = x^2$

FIG. 8

Proof. If $\mathscr{A}$ is a covering of X and Γ is a mapping with range Y, then $\Gamma\mathscr{A}$ is a covering of Y. If $\Gamma(A_i) \cap \Gamma(A_j) \neq \emptyset$ there exist elements $z \in A_i$ and $z' \in A_j$ such that $\Gamma(z) \cap \Gamma(z') \neq \emptyset$. But Γ is injective and so $z = z'$; hence $A_i = A_j$ and so $\Gamma(A_i) = \Gamma(A_j)$.

Theorem 6. *If $\mathscr{B} = (B_i \mid i \in I)$ is a filter base in X and Γ is a mapping whose domain is X, the family $\Gamma\mathscr{B}$ is a filter base in Y.*

Proof. If Γ is a mapping whose domain is X, then, for all i, $\Gamma B_i \neq \emptyset$. Furthermore, for each (i, j) in $I \times I$, there exists a set B_k contained in $B_i \cap B_j$. Therefore, by Theorem 3,

$$\Gamma B_k \subset \Gamma(B_i \cap B_j) \subset \Gamma B_i \cap \Gamma B_j.$$

Hence $\Gamma\mathscr{B}$ is a filter base.

REMARK. If Γ is a mapping of X into Y we call the subset

$$\sum_{x \in X} \Gamma x = \{(x, y) \mid x \in X, y \in Y, y \in \Gamma x\}$$

of $X \times Y$ the **graphical representation** of Γ. If Γ is a numerical function, this reduces to the well-known concept illustrated in figure 8, part 3.

§ 2. Operations on mappings

If Γ_1 and Γ_2 are two mappings of X into Y, their **union** is the mapping $(\Gamma_1 \cup \Gamma_2)$ of X into Y defined by

$$(\Gamma_1 \cup \Gamma_2)\, x = \Gamma_1 x \cup \Gamma_2 x.$$

The **intersection** of Γ_1 and Γ_2 is the mapping $(\Gamma_1 \cap \Gamma_2)$ of X into Y defined by

$$(\Gamma_1 \cap \Gamma_2)\, x = \Gamma_1 x \cap \Gamma_2 x.$$

The **Cartesian product** of Γ_1 and Γ_2 is the mapping $(\Gamma_1 \times \Gamma_2)$ of X into $Y \times Y$ defined by

$$(\Gamma_1 \times \Gamma_2)\, x = \Gamma_1 x \times \Gamma_2 x.$$

More generally, given any operation $\vee$ on sets, we write

$$(\Gamma_1 \vee \Gamma_2)\, x = \Gamma_1 x \vee \Gamma_2 x.$$

If Γ_1 is a mapping of X into Y and Γ_2 is a mapping of Y into Z, the **composition product** of Γ_2 by Γ_1 is the mapping $(\Gamma_2 . \Gamma_1)$ of X into Z defined by

$$(\Gamma_2 \cdot \Gamma_1)\, x = \Gamma_2\,(\Gamma_1 x).$$

We have

$$(1)\quad (\Gamma_1 \cup \Gamma_2)\, A = \Gamma_1 A \cup \Gamma_2 A;$$
$$(2)\quad (\Gamma_1 \cap \Gamma_2)\, A \subset \Gamma_1 A \cap \Gamma_2 A,$$

for

$$\Big(\bigcup_x \Gamma_1 x\Big) \cap \Big(\bigcup_y \Gamma_2 y\Big) = \bigcup_{x,y} \Gamma_1 x \cap \Gamma_2 y = (\Gamma_1 \cap \Gamma_2)A \cup \bigcup_{x \neq y} (\Gamma_1 x \cap \Gamma_2 y);$$
$$(3)\quad (\Gamma_1 \times \Gamma_2)A \subset \Gamma_1 A \times \Gamma_2 A,$$

for

$$\bigcup_x \Gamma_1 x \times \bigcup_y \Gamma_2 y \supset \bigcup_{x,y} (\Gamma_1 x \times \Gamma_2 y) \supset \bigcup_x \Gamma_1 x \times \Gamma_2 x;$$
$$(4)\quad (\Gamma_2 \cdot \Gamma_1)A = \Gamma_2(\Gamma_1 A).$$

A mapping Δ is said to be **constant** if there exists a subset C of Y such that $\Delta x = C$ for all x. A constant mapping satisfies the following property:

$$(\Delta \cap \Gamma)A = \Delta A \cap \Gamma A.$$

Another important mapping is the **identity** mapping, which is the single-valued mapping I of X onto X defined by

$$Ix = \{x\}.$$

If Γ_1, Γ_2, Γ_3 are mappings of X into X, we have

$$\Gamma_3 \cdot (\Gamma_2 \cdot \Gamma_1)x = (\Gamma_3 \cdot \Gamma_2) \cdot \Gamma_1 x$$
(associativity of composition product).

Thus, if Γ is a mapping of X into X, we can write

$$\Gamma^2 x = \Gamma(\Gamma x),$$
$$\Gamma^3 x = \Gamma(\Gamma^2 x) = \Gamma^2(\Gamma x), \text{ etc.}$$

The **transitive closure** of Γ is a mapping $\hat{\Gamma}$ of X into X defined by

$$\hat{\Gamma} x = \{x\} \cup \Gamma x \cup \Gamma^2 x \cup \Gamma^3 x \cup \ldots$$

The correspondence $A \to \hat{\Gamma} A$ is a closure operation, for

$$(1)\quad \hat{\Gamma} A \supset A,$$
$$(2)\quad A \supset B \;\Rightarrow\; \hat{\Gamma} A \supset \hat{\Gamma} B,$$
$$(3)\quad \hat{\Gamma}(\hat{\Gamma} A) = \hat{\Gamma} A.$$

Theorem 1. *If Γ_1 and Γ_2 are two semi-single-valued mappings, $\Gamma_1 \cap \Gamma_2$ and $\Gamma_1 \times \Gamma_2$ are semi-single-valued mappings.*

Proof. We have

$$(\Gamma_1 \cap \Gamma_2)x \cap (\Gamma_1 \cap \Gamma_2)x' \neq \emptyset \Rightarrow \begin{Bmatrix} \Gamma_1 x \cap \Gamma_1 x' \neq \emptyset \\ \Gamma_2 x \cap \Gamma_2 x' \neq \emptyset \end{Bmatrix} \Rightarrow$$
$$\Rightarrow \begin{Bmatrix} \Gamma_1 x = \Gamma_1 x' \\ \Gamma_2 x = \Gamma_2 x' \end{Bmatrix} \Rightarrow \Gamma_1 x \cap \Gamma_2 x = \Gamma_1 x' \cap \Gamma_2 x'$$

and the proof for the Cartesian product $\Gamma_1 \times \Gamma_2$ is similar.

Theorem 2. *If one of the mappings Γ_1 and Γ_2 is injective, the mappings $\Gamma_1 \cap \Gamma_2$ and $\Gamma_1 \times \Gamma_2$ are injective.*

Proof. If a and b are two distinct points, then

$$(\Gamma_1 a \cap \Gamma_2 a) \cap (\Gamma_1 b \cap \Gamma_2 b) = (\Gamma_1 a \cap \Gamma_1 b) \cap (\Gamma_2 a \cap \Gamma_2 b) = \emptyset$$

and the proof for the Cartesian product $\Gamma_1 \times \Gamma_2$ is similar.

§ 3. Upper and lower inverses of a mapping

If Γ is a mapping of X into Y, its **lower inverse** is the mapping Γ^- of Y into X defined by

$$\Gamma^- y = \{x \,/\, x \in X,\ y \in \Gamma x\}.$$

This is a mapping whose domain is Y^* and whose range is X^*; for any non-empty subset B of Y we have

$$\Gamma^- B = \{x \,/\, x \in X,\ \Gamma x \cap B \neq \emptyset\}$$

and we also write $\Gamma^- \emptyset = \emptyset$. It is clear that the inverse of Γ^- is $(\Gamma^-)^- = \Gamma$ and that $y \in \Gamma x$ is equivalent to $x \in \Gamma^- y$.

For a single-valued mapping σ, the lower inverse is denoted by σ^{-1} and is simply called the **inverse**, because no confusion is possible.

Theorem 1. *If Γ is single-valued, Γ^- is injective; if Γ is injective Γ^- is single-valued; if Γ is semi-single-valued, Γ^- is semi-single-valued.*

Proof. If Γ is single-valued, then

$$y \neq y' \quad \Rightarrow \quad \Gamma^- y \cap \Gamma^- y' = \emptyset.$$

If Γ is injective, the set $\Gamma^- y = \{x \,/\, y \in \Gamma x\}$ has only one element and Γ^- is therefore single-valued. If Γ is semi-single-valued, we have

$$\Gamma^- y \cap \Gamma^- y' \neq \emptyset \quad \Rightarrow \quad (\exists\, x) : y \in \Gamma x,\ y' \in \Gamma x \quad \Rightarrow \quad \Gamma^- y = \Gamma^- y'.$$

Apart from the mapping Γ^-, it is sometimes useful to consider another correspondence of a similar kind, which we call the **upper inverse** of Γ; in this mapping to each subset B of Y there corresponds the set

$$\Gamma^+ B = \{x \,/\, x \in X^*, \Gamma x \subset B\}.$$

In particular, if $B = \varnothing$, we have $\Gamma^+ \varnothing = \varnothing$. We observe that for all subsets B we have $\Gamma^+ B \subset \Gamma^- B$ and that, for a single-valued mapping σ, we have $\sigma^+ B = \sigma^- B = \sigma^{-1} B$.

EXAMPLE. Let X be the set of possible positions in the game of chess; a position consists of the coordinates of the different pieces on the chess-board and the player whose move is next. The set X is then the union of three disjoint sets X_1, X_2, X_0; X_1 is the set of positions in which White can move, X_2 is the set of positions in which Black can move and X_0 is the set of positions of checkmate or stalemate, when it is not possible for either to move.

If $x \in X_1$ (resp. X_2) we shall denote by Γx the set of positions which White (resp. Black) can reach immediately after position x; this determines a mapping Γ of $X - X_0$ into X. Then $\hat{\Gamma} x$ denotes the set of positions that it will be possible to reach eventually, starting with the position x; $\Gamma^- x$ is the set of possible positions which could have occurred immediately before the position x and, if A is a subset of X, $\Gamma^+ A$ is the set of positions which can only give a position belonging to A in the following move.

If K_2 denotes the set of positions in which Black is checkmated, clearly $K_2 \subset X_0$. It is easily verified that the set of positions in which White can 'mate in two moves' is $\Gamma^- \Gamma^+ \Gamma^- K_2$.

In all the following theorems, Γ denotes a mapping whose domain is X and whose range is Y.

Theorem 2. *If B_1 and B_2 are subsets of Y, then we have*

$$(1)\ -\Gamma^+ B_1 = \Gamma^- (-B_1);\ -\Gamma^- B_2 = \Gamma^+ (-B_2),$$
$$(2)\ \Gamma^+ B_1 \cup \Gamma^+ B_2 \subset \Gamma^+ (B_1 \cup B_2),$$
$$(3)\ \Gamma^- B_1 \cup \Gamma^- B_2 = \Gamma^- (B_1 \cup B_2).$$

The proofs of these results are immediate.

We say that a family of sets $\mathscr{A}$ is **complemented** if

$$A \in \mathscr{A} \Rightarrow X - A \in \mathscr{A}.$$

Theorem 3. *The subsets P of Y such that $\Gamma^+ P = \Gamma^- P$ (called 'pure' subsets) form a complemented lattice $\mathscr{P}$.*

Proof. If $P \in \mathscr{P}$, then $Y - P \in \mathscr{P}$, for

$$\Gamma^+ (-P) = -\Gamma^- P = -\Gamma^+ P = \Gamma^- (-P).$$

Further, if P_1 and P_2 belong to $\mathscr{P}$, we have

$$\Gamma^- (P_1 \cup P_2) = \Gamma^- P_1 \cup \Gamma^- P_2 = \Gamma^+ P_1 \cup \Gamma^+ P_2 \subset \Gamma^+ (P_1 \cup P_2).$$

Since the opposite inclusion holds automatically, we have $P_1 \cup P_2 \in \mathscr{P}$. Also $P_1 \cap P_2 \in \mathscr{P}$, for

$$-(P_1 \cap P_2) = (-P_1) \cup (-P_2) \in \mathscr{P}.$$

Theorem 4. *The subsets S of X such that $\Gamma^- \Gamma S = S$ (called 'stable' subsets) form a complemented lattice $\mathscr{S}$.*

Proof. Clearly $\mathscr{S}$ is complemented; moreover, if S_1 and S_2 belong to $\mathscr{S}$, we have

$$\Gamma^-\Gamma (S_1 \cup S_2) = \Gamma^-\Gamma S_1 \cup \Gamma^-\Gamma S_2 = S_1 \cup S_2$$

and so $S_1 \cup S_2 \in \mathscr{S}$; also $S_1 \cap S_2 \in \mathscr{S}$, since

$$-(S_1 \cap S_2) = (-S_1) \cup (-S_2) \in \mathscr{S}.$$

Theorem 5. *The correspondence $A \to \Gamma^+\Gamma A$ is a closure operation (called the 'Galois closure').*

Proof. It can easily be verified that

$$\begin{aligned} &(1)\ \Gamma^+\Gamma A \supset A,\\ &(2)\ A \supset B \ \Rightarrow\ \Gamma^+ \Gamma A \supset \Gamma^+ \Gamma B,\\ &(3)\ \Gamma^+ \Gamma (\Gamma^+ \Gamma A) = \Gamma^+ (\Gamma\Gamma^+) (\Gamma A) = \Gamma^+ (\Gamma A). \end{aligned}$$

REMARK. In general, $\Gamma^-\Gamma$ is not a closure operation, as is shown by a case such as that illustrated in figure 9.

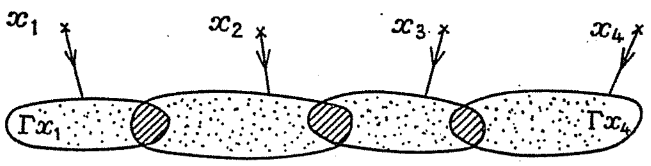


FIG. 9

In fact, if we take $A = \{x_2\}$, we have

$$\Gamma^-\Gamma A = \{x_1, x_2, x_3\},$$
$$\Gamma^- \Gamma (\Gamma^-\Gamma A) = \{x_1, x_2, x_3, x_4\}.$$

EXAMPLE. We now re-examine the concept of linear closure (defined in § 8, Chapter I). Let X be the set $\mathbf{R}^3$ and Y the set of planes in $\mathbf{R}^3$. Let Γ be the mapping of X into Y in which the point x corresponds to the set of planes not containing x. If x_1, x_2, x_3 are three non-collinear points, we denote the plane containing them by $P(x_1, x_2, x_3)$ and the straight line containing the points x_1 and x_2 by $D(x_1, x_2)$. Then $x \in P(x_1, x_2, x_3)$ is equivalent to

$$\Gamma x \subset \Gamma x_1 \cup \Gamma x_2 \cup \Gamma x_3 = \Gamma \{x_1, x_2, x_3\}.$$

Thus

$$\Gamma^+ \Gamma \{x_1, x_2, x_3\} = P(x_1, x_2, x_3)$$

and similarly

$$\Gamma^+ \Gamma \{x_1, x_2\} = D(x_1, x_2).$$

Here the Galois closure coincides with the linear closure.

Let $\mathscr{A}$ be a complemented family of sets and let $\mathscr{B}$ be a filter base. Then we say that $\mathscr{B}$ is an **ultra-filter base with respect to** $\mathscr{A}$ if, for each set A of $\mathscr{A}$, we have either $\mathscr{B} \vdash A$ or $\mathscr{B} \vdash X - A$ (if $\mathscr{A} = \mathscr{P}(X)$, the collective family of subsets of X, we recover the definition of ultra-filter base, given in Chapter I). The following result can be proved immediately:

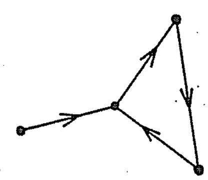
Let $\mathscr{B} = (B_i \mid i \in I)$ be an ultra-filter base in X and let Γ be a mapping of X into Y; then $\Gamma\mathscr{B}$ is an ultra-filter base with respect to the lattice $\mathscr{P}_\Gamma$ of pure subsets of Y.

In particular, if σ is a single-valued mapping of X into Y and $\mathscr{B}$ is an ultra-filter base in X, then $\sigma\mathscr{B}$ is an ultra-filter base in Y (for $\sigma^- = \sigma^+ = \sigma^{-1}$: every subset of Y is pure).

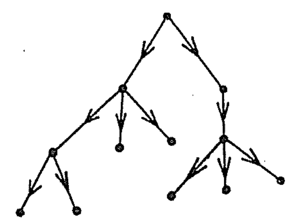
§ 4. Graphs

A pair consisting of a set X and a mapping Γ of X into itself is called a **graph** (or, more precisely, an oriented graph). If two elements x_1 and x_2 of X are such that $x_2 \in \Gamma x_1$, we say that x_1 is *linked to* x_2 *by the relation* Γ, which is represented by joining the two points by an oriented line from x_1 to x_2.

If there exists an element x of X such that $x \in \Gamma\hat{\Gamma}x$, we say that the graph (X, Γ) is **cyclic**; if Γ is injective, we say that the graph (X, Γ) is a **tree**. A **path** is a sequence of points $x_1, x_2, x_3, \ldots$ such that $x_i \in \Gamma x_{i-1}$ for $i = 2, 3, \ldots$; a **chain** is a sequence of points $x_1, x_2, x_3, \ldots$ such that



Cyclic graph



Tree

FIG. 10

$x_i \in \Gamma x_{i-1} \cup \Gamma^- x_{i-1}$ for $i = 2, 3, \ldots$ The study of combinatorial properties of chains and paths constitutes the **theory of graphs**.

CHAPTER III

ORDERED SETS

§ 1. Order and equivalence

Let A be any set; a **binary relation** in A is a proposition (L) such that, given any pair (a, b), where $a, b \in A$, we can say whether the proposition (L) is true or false. For example, if A is a set of individuals, 'a is a brother of b' and 'a is an ancestor of b' are binary relations in A. All the pairs (a, b) which satisfy (L) form a subset L of $A \times A$ and those which do not satisfy (L) form the set $-L$. If $(a, b) \in L$, we say that *a is related to b by (L)* and we often indicate this by writing $a \geqq b$. This relation is called a **pre-ordering** if

(1) $a \geqq a$ (reflexivity)
(2) $a \geqq b,\ b \geqq c \ \Rightarrow\ a \geqq c$ (transitivity).

Given a pre-ordering $\geqq$, we write $b \leqq a$ if $a \geqq b$. If $a \geqq b$ and it is not true that $b \geqq a$, we write $a > b$. If, however, $a \geqq b$ and $b \geqq a$, we write $a \equiv b$. The relation $\leqq$ is also a pre-ordering, for

(1) $a \leqq a$,
(2) $c \leqq b,\ b \leqq a \ \Rightarrow\ c \leqq a$.

The relation $\equiv$ is an **equivalence relation**: that is, it satisfies

(1) $a \equiv a$ (reflexivity),
(2) $a \equiv b,\ b \equiv c \ \Rightarrow\ a \equiv c$ (transitivity),
(3) $a \equiv b \ \Rightarrow\ b \equiv a$ (symmetry).

A relation $\geqq$ which satisfies

(1) $a \geqq a$,
(2) $a \geqq b,\ b \geqq c \ \Rightarrow\ a \geqq c$,
(3) $a \geqq b,\ b \geqq a \ \Rightarrow\ a = b$,

is called a **partial ordering** or just an **ordering** (in other words an ordering is a pre-ordering which admits equality as the corresponding equivalence).

EXAMPLES. In the set N the following determine orderings: 'p divides q', 'p is a multiple of q', 'p is greater than or equal to q', 'p is less than or equal to q'.

In a set A of individuals, 'a possesses the same parents as b' is an equivalence relation, 'a is b or an ancestor of b' is an ordering; 'a is stronger than b' is a strict ordering $(a > b)$, but 'a wins regularly in playing against b' is not an ordering, because the property of transitivity is not necessarily satisfied.

In the family of subsets of any set X, '$A \supset B$' is an ordering.

Given an equivalence relation $\equiv$ in X, a set of the form

$$S_a = \{x \,/\, x \in X, x \equiv a\}$$

is called an **equivalence class.**

Theorem. *The collective family of equivalence classes is a partition of X.*

Proof. Let S, S', etc. be the equivalence classes with respect to an equivalence relation $\equiv$ in a set X. Then

(1) $S \neq \emptyset$, since $S = S_a \ni a$,

(2) $S \neq S' \Rightarrow S \cap S' = \emptyset$, for if $S_a \cap S_b \neq \emptyset$, then there exists an element c of this intersection and we have

$$x \in S_a \iff x \equiv a \iff x \equiv c \iff x \equiv b \iff x \in S_b,$$

whence $S_a = S_b$.

(3) Clearly $X = \bigcup_{a \in X} S_a$.

REMARK. This theorem shows that every equivalence relation in X determines a partition of X. Conversely, every partition $\mathscr{A} = (A_i \,/\, i \in I)$ defines an equivalence relation; for if we write $a \equiv b$ whenever a and b belong to the same set A_i, we have

$$\begin{aligned}
&(1)\ a \equiv a,\\
&(2)\ a \equiv b,\ b \equiv c \quad \Rightarrow \quad a \equiv c,\\
&(3)\ a \equiv b \quad \Rightarrow \quad b \equiv a.
\end{aligned}$$

EXAMPLE 1. In the set $\mathbf{N}$, write $p \equiv q(3)$ if $p = q \pm 3k$, where $k \in \mathbf{N}$. This determines an equivalence relation; in arithmetic, the corresponding equivalence classes are called *integers modulo* 3.

EXAMPLE 2. In elementary geometry, we call a pair (x, y) in $\mathbf{R}^3 \times \mathbf{R}^3$ a *localised vector*, which we denote by $\overrightarrow{xy}$. We can define an equivalence relation $\equiv$ by

$$\overrightarrow{xy} \equiv \overrightarrow{x'y'} \iff x - y = x' - y'.$$

With respect to $\equiv$, an equivalence class is called a *free vector*.

Another equivalence $\simeq$ is defined by

$$\overrightarrow{xy} \simeq \overrightarrow{x'y'} \iff \overrightarrow{xy} \equiv \overrightarrow{x'y'} \text{ and } x, y, x', y' \text{ collinear.}$$

With respect to $\simeq$,an equivalence class is called a *line vector*.

§ 2. Countable infinite and continuum infinite sets

A set is said to be **finite** if it consists of a finite number of elements: that is, it has no elements, or just one element, or two elements, or three elements, etc. A set which does not consist of a finite number of elements is said to be **infinite**. Let A and B be two sets, finite or infinite. If there exists a one-one correspondence between them, we say that A and B have the same number of elements or **have the same power** and write $A \sim B$. Clearly $\sim$ is an equivalence relation, for

$$\begin{aligned}&(1)\ A \sim A,\\ &(2)\ A \sim B,\ B \sim C \quad \Rightarrow \quad A \sim C,\\ &(3)\ A \sim B \quad \Rightarrow \quad B \sim A.\end{aligned}$$

If a set A has the same power as the set $\mathbf{N} = \{0, 1, 2, \ldots\}$ of positive integers, we say that it is **denumerable**; if A has the same power as the set $[0, 1]$ of real numbers between 0 and 1 we say that it is **continuum infinite** or **has the power of the continuum.** A set which is either finite or denumerable is said to be **countable.**[1]

The set $[0, 1]$ is not denumerable. For suppose that there is a one-one correspondence between $\mathbf{N}$ and $[0, 1]$, such as

$$\begin{aligned}1 &\to a_1 = 0{\cdot}648729\ldots\\ 2 &\to a_2 = 0{\cdot}844371\ldots\\ 3 &\to a_3 = 0{\cdot}119762\ldots\\ &\ldots\ldots\ldots\ldots\ldots\ldots\end{aligned}$$

To avoid one number a_i of $[0, 1]$ being represented by two distinct decimal developments (e.g. $0{\cdot}499999\ldots$ and $0{\cdot}5000000\ldots$) we do not include those which consist only of nines from a certain point onwards. If we form a number a (such as $0{\cdot}513\ldots$) in which the first digit is different from the first digit of a_1, the second digit is different from the second digit of a_2, the third digit is different from the third digit of a_3 and so on (avoiding the use of 9) then a does not appear in the table representing the correspondence, because it differs from each a_i in at least one place. Thus the correspondence cannot after all be one-one, which gives a contradiction.

EXAMPLE 1. Consider the set of positive rational numbers

$$\mathbf{R}_r^+ = \left\{x \,/\, x = \frac{p}{q},\ p \in \mathbf{N},\ q \in \mathbf{N},\ q \neq 0\right\}$$

and form the following table

[1] (Translator's note.) The terms 'denumerable' and 'countable' are not always used in the same sense as here. The French edition uses only the word 'dénombrable', but we have felt it advantageous to use the two words so as to make it quite clear when the finite case is included.

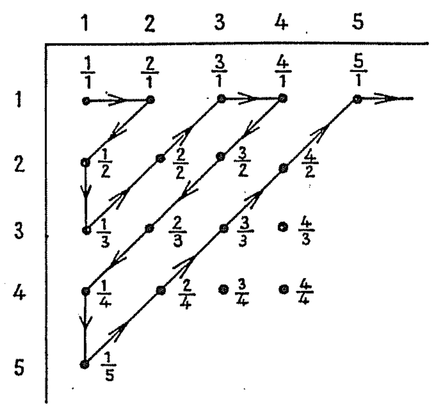


FIG. 11

If we number the elements of the different diagonals successively in the senses indicated by the arrows in the table, omitting those that have already been numbered (e.g. $\frac{2}{4} = \frac{1}{2}$), we can establish a one-one correspondence between N and $\mathbf{R}_r^+$:

$$\begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\ 0 & 1 & 2 & \frac{1}{2} & \frac{1}{3} & 3 & 4 & \dots \end{array}$$

Therefore the set $\mathbf{R}_r^+$ is denumerable.

In a similar manner, we can prove that *if two sets A and B are denumerable, then $A \times B$ is also denumerable.*

EXAMPLE 2. The positive augmented half-line $[0, +\infty]$ has the power of the continuum, for the correspondence $\lambda \to (\frac{1}{\lambda} - 1)$ is a one-one correspondence between $[0, 1]$ and $[0, +\infty]$.

The terminology introduced above is extended to families of sets; we say that a family $\mathscr{A} = (A_i \mid i \in I)$ is **finite** if the set I is finite and otherwise we say that $\mathscr{A}$ is **infinite.** We also say that $\mathscr{A}$ is **denumerable** (resp. **countable**) if the set I is **denumerable** (resp. **countable**) and is **continuum infinite** if I has the power of the continuum.

In addition to the above, we say that $\mathscr{A}$ is **locally finite at a point** $a \in A$ if the set $I_a = \{i \mid i \in I, a \in A_i\}$ is finite and that $\mathscr{A}$ is **locally finite** if it is locally finite at each point. In a similar manner, we can define families which are **locally denumerable,** or **locally countable,** or **locally continuum infinite.**

EXAMPLE 3. *The union of a denumerable family of denumerable sets is denumerable*; this can be verified exactly as in Example 1.

EXAMPLE 4. Let X be the set $\mathbf{R}$ and let $\mathscr{A} = (A_i \,/\, i \in \mathrm{I})$ be a locally denumerable family of proper intervals (that is, intervals which each have more than one point). We shall prove that $\mathscr{A}$ is denumerable.

We begin by enumerating the positive and negative rational numbers $\lambda_1, \lambda_2, \lambda_3 \ldots$ as in Example 1. We then set up a correspondence as follows: with the number λ_1 we associate the denumerable set I_1 of indices i such that $\lambda_1 \in A_i$; with λ_2 we associate the set I_2 of indices i such that $\lambda_2 \in A_i$ and so on. Since each A_i includes a rational number, we cover all indices in this way and so $I = \bigcup_{n=1} I_n$, which is the union of a denumerable family and so is denumerable (cf. Example 3).

§ 3.* Transfinite cardinal numbers

As we have already seen, the relation $\sim$, which expresses the fact that two sets have the same power, is an equivalence relation. The idea of cardinal number is introduced as a means of labelling the equivalence classes corresponding to the relation $\sim$. For example, a set A 'has cardinal number 3' if there is a one-one correspondence between A and the set consisting of the thumb, the first finger and the second finger; in fact '3' is the name which the first arithmeticians gave to the equivalence class of $\sim$ which contains the set: thumb, first finger, second finger.

Clearly we can do something similar for any finite set and in fact we extend the idea to infinite sets as well. We define the **cardinal number** $o(A)$ of a set A to be the equivalence class of $\sim$ which contains the set A. The cardinal number of a denumerable set is denoted by $\aleph_0$ (aleph nought); that of a continuum infinite set[1] is denoted by $\aleph_1$.

If two sets A and B are such that there is a one-one correspondence between A and a subset of B, we write $o(A) \leqq o(B)$. It can be proved that, given any two sets A and B, then either $o(A) \leqq o(B)$ or $o(B) \leqq o(A)$. For cardinal numbers, $\leqq$ is a pre-ordering, for

$$(1)\ o(A) \leqq o(A)$$
$$(2)\ o(A) \leqq o(B),\ o(B) \leqq o(C) \quad \Rightarrow \quad o(A) \leqq o(C).$$

The following theorem shows that $\leqq$ is, in fact, an ordering relation.

Bernstein's theorem. *If $o(A) \leqq o(B)$ and $o(B) \leqq o(A)$, then $o(A) = o(B)$.*

Proof. Let σ be a one-one correspondence between A and a subset of B and let τ be a one-one correspondence between B and a subset of A; write $A' = A - \tau B$.

[1] (Translator's note.) Many authors prefer to denote the cardinal number of a continuum infinite set by c whenever the continuum hypothesis (see page 35) is not assumed.

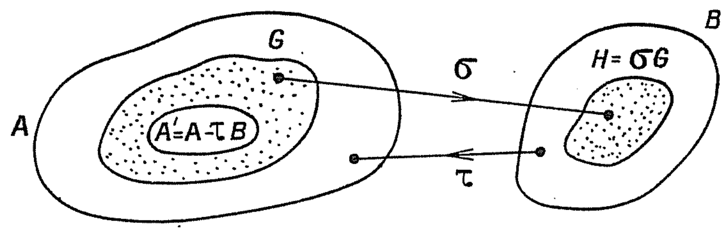


FIG. 12

Let $\mathscr{A} = (A_i \,/\, i \in I)$ denote the collective family of subsets of A which satisfy

$$(1)\ A_i \supset A',$$
$$(2)\ \tau(\sigma A_i) \subset A_i.$$

This family $\mathscr{A}$ is non-empty, for $A \in \mathscr{A}$. If we put $G = \bigcap_{i \in I} A_i$, we have

$$(1)\ G \supset A',$$
$$(2)\ \tau\sigma G = \tau\sigma\,(\bigcap A_i) \subset \bigcap \tau\sigma A_i \subset \bigcap A_i = G.$$

Then $G \in \mathscr{A}$ and also $G' = A' \cup \tau\sigma G \in \mathscr{A}$, for

$$(1)\ G' \supset A',$$
$$(2)\ \tau\sigma G' \subset \tau\sigma G \subset A' \cup \tau\sigma G = G'.$$

Then we have $G' \supset G$ since $G' \in \mathscr{A}$, and $G' \subset G$ since

$$G' = A' \cup \tau\sigma G \subset A' \cup G = G.$$

Thus

$$A' \cup \tau\sigma G = G.$$

Consider the mapping α of A into B defined by

$$\alpha x = \sigma x \quad \text{if} \quad x \in G,$$
$$\alpha x = \tau^{-1} x \ \text{if} \quad x \in A - G.$$

In order to show that this is a one-one correspondence between A and B, it is sufficient to show that if we write $\sigma G = H$, then we have $\tau(B-H) = A-G$; and this is immediate, since

$$\tau\,(B-H) = \tau B - (A' \cup \tau\sigma G) = \tau B - G = A - G.$$

EXAMPLE 1. The unit square $[0, 1] \times [0, 1] = [0, 1]^2$ has the power of the continuum. For if $\aleph$ is the cardinal number of $[0, 1]^2$, we have $\aleph_1 \leqq \aleph$, since

$$[0, 1] \sim \{(x, y) \,/\, 0 \leqq x \leqq 1, y = 0\} \subset [0, 1]^2.$$

On the other hand, to each $(x, y) \in [0, 1]^2$ there corresponds an element $\lambda \in [0, 1]$ defined as follows:

$$\left.\begin{aligned} x &= 0{\cdot}x_1x_2x_3x_4x_5x_6\ldots \\ y &= 0{\cdot}y_1y_2y_3y_4y_5y_6\ldots \end{aligned}\right\} \rightarrow \lambda = 0{\cdot}x_1y_1x_2y_2x_3y_3x_4y_4x_5y_5\ldots$$

for example,

$$\left.\begin{array}{l} x = 0{\cdot}4396055\ldots. \\ y = 0{\cdot}8886431\ldots. \end{array}\right\} \rightarrow \lambda = 0{\cdot}48389866045351\ldots$$

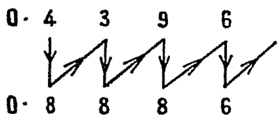


FIG. 13

In writing x and y we avoid, as before, decimal developments which consist only of nines from a certain point onwards; otherwise we get repetitions. Let E be the set of numbers λ such that the decimal development of λ consists of nines in every second position (but not every position) from a certain point onwards. Then

$$[0, 1]^2 \sim [0, 1]-E \subset [0, 1].$$

Therefore $\aleph \leqq \aleph_1$ and so, by Bernstein's theorem, $\aleph = \aleph_1$.

This result can be generalised as follows: *The product $A \times B$ of two sets having the power of the continuum also has the power of the continuum.*

COROLLARY 1. *If the union $A \cup B$ is denumerable or continuum infinite, then $o(A \cup B)$ is the greater of the cardinal numbers $o(A)$ and $o(B)$.*

Proof. Without loss of generality, we can suppose that $o(A) \geqq o(B)$. Since $A \subset A \cup B$, we have $o(A) \leqq o(A \cup B)$. Hence, by Bernstein's Theorem, it is sufficient to prove that the opposite inequality holds.

If $o(A) = \aleph_0$, we have

$$\begin{array}{l} A \sim \{0, 2, 4, 6, \ldots\} \\ B \sim E \subset \{1, 3, 5, \ldots\} \end{array}$$

Therefore $o(A \cup B) \leqq o(\mathrm{N}) = \aleph_0 = o(A)$.

If $o(A) = \aleph_1$, we have

$$\begin{array}{l} A \sim [0, 1/2], \\ B \sim E \subset [2/3, 1]. \end{array}$$

Therefore $o(A \cup B) \leqq o([0, 1]) = \aleph_1 = o(A)$.

COROLLARY 2. *If A is a denumerable or continuum infinite set and if B is a set such that $o(B) < o(A)$, then $o(A-B) = o(A)$.*

Proof. By Corollary 1, $(A-B) \cup B = A \cup B$ has cardinal number $o(A)$ and so $o(A)$ is the greater of the cardinal numbers $o(A-B)$ and $o(B)$. Since $o(A) \neq o(B)$, we have $o(A-B) = o(A)$.

EXAMPLE 2. We shall show that the set $\mathscr{P}(\mathrm{N})$ of subsets of N has cardinal number $\aleph_1$.

Let Q be a subset of N and let ϕ_Q be the characteristic function (cf. Chapter I, § 2) of Q. Writing

$$\lambda = \frac{\phi_Q(0)}{2} + \frac{\phi_Q(1)}{2^2} + \frac{\phi_Q(2)}{2^3} + \ldots$$

we set up a correspondence between $\mathscr{P}(\mathrm{N})$ and $[0, 1]$. This correspondence is not one-one because, for example,

$$\frac{1}{2} = \frac{1}{2}+\frac{0}{2^2}+\frac{0}{2^3}+\ldots = \frac{0}{2}+\frac{1}{2^2}+\frac{1}{2^3}+\ldots$$

but the numbers in $[0, 1]$ which, like $\frac{1}{2}$, have two dyadic developments, form a denumerable set. Let $\mathscr{A}$ be the collection of sets Q such that the $\phi_Q(n)$ are all equal to 1 from a certain point onwards. Then we have

$$\mathscr{P}(\mathrm{N})-\mathscr{A} \sim [0, 1],$$
$$\mathscr{A} \sim \mathrm{N}$$

and so, by Corollary 2,

$$o(\mathscr{P}(\mathrm{N})) = o(\mathscr{P}(\mathrm{N})-\mathscr{A}) = \aleph_1.$$

Cantor's theorem. *For any set A, the family $\mathscr{P}(A)$ of subsets of A satisfies*

$$o(\mathscr{P}(A)) > o(A).$$

Proof. Since $(\{a\} \,/\, a \in A) \subset \mathscr{P}(A)$, we have $o(\mathscr{P}(A)) \geqq o(A)$. Equality is impossible, for if $o(\mathscr{P}(A))$ were equal to $o(A)$ then there would be a one-one correspondence in which a subset $A_a \subset A$ would correspond to $a \in A$; but the subset B defined by

$$B = \{a \,/\, a \in A,\ a \notin A_a\}$$

would then differ from each set A_c by at least the element c and so could not be included in the correspondence.

If A is any infinite set, then it has a denumerable subset; for if a_1 is any element of A, then A has an element a_2 different from a_1, and an element different from both a_1 and a_2, and so on. Therefore $\aleph_0$ is the smallest transfinite cardinal number.

We have seen that $\mathscr{P}(\mathrm{N})$ has cardinal number $\aleph_1$; we denote the cardinal number of $\mathscr{P}([0, 1])$ by $\aleph_2$. More generally, if A is a set of cardinal number $\aleph_k$, we denote the cardinal number of $\mathscr{P}(A)$ by $\aleph_{k+1}$. From Cantor's Theorem we get

$$\aleph_0 < \aleph_1 < \aleph_2 < \ldots$$

The following is called the *generalised hypothesis of the continuum: there is no cardinal number $\aleph$ such that*

$$\aleph_k < \aleph < \aleph_{k+1}.$$

This proposition has never been proved, although it is generally accepted. All the sets which we shall consider have cardinal number $\aleph_0$ or $\aleph_1$ or $\aleph_2$.

REMARK. It is easy to extend the usual operations of elementary arithmetic to transfinite cardinal numbers. If A and A' are sets such that $A \cap A' = \emptyset$ and if $\aleph = o(A)$ and $\aleph' = o(A')$, then the cardinal number $o(A \cup A')$ is called the *sum* $\aleph + \aleph'$ of $\aleph$ and $\aleph'$. The cardinal number $o(A \times A')$ is called the *product* $\aleph \times \aleph'$. We have

$$\aleph_0 + \aleph_0 = 2\,\aleph_0 = \aleph_0$$
$$\aleph_0 \times \aleph_0 = \aleph_0^2 = \aleph_0, \text{ etc.} \ldots$$

§ 4. Ordered sets

A pair $(X, \geqq)$ consisting of a set X and an ordering relation $\geqq$ defined in X is called an **ordered set**. If $a \in X$, the set

$$S(a) = \{x \,/\, x \in X,\ x \geqq a\}$$

is called the **section** from a. The correspondence $a \to S(a)$ is a mapping of X into X, which we shall denote by S.

Theorem. *If B is a subset of X, the correspondence*

$$B \to S(B) = \bigcup_{b \in B} S(b)$$

is an additive closure operation: that is,

(1) $S(B) \supset B$,
(2) $B \supset B' \quad \Rightarrow \quad S(B) \supset S(B')$,
(3) $S(SB) = SB$,
(4) $S(B \cup B') = S(B) \cup S(B')$.

Proof. Condition (1) is immediate since $b \in S(b)$; (2) and (4) follow from the fact that S is a mapping. To prove (3), it is sufficient to observe that

$$x \in S(SB) \;\Rightarrow\; x \geqq a,\ a \geqq b,\ b \in B \;\Rightarrow\; x \geqq b,\ b \in B \;\Rightarrow\; x \in S(B).$$

Then $S(SB) \subset SB$ whence, by (1), $S(SB) = SB$.

Let $(X, \geqq)$ be an ordered set and let B be a subset of X; we say that $a \in X$ is a **majorant** of B if $a \geqq x$ for all $x \in B$. Similarly we say that $a \in X$ is a **minorant** of B if $a \leqq x$ for all $x \in B$. The set $S[B] = \bigcap_{b \in B} S(b)$ is thus the set of majorants of B. If S^- denotes the lower inverse defined by

$$S^-(a) = \{x \,/\, x \in X,\ a \geqq x\} = \{x \,/\, x \in X,\ x \leqq a\},$$

then the set $S^-[B] = \bigcap_{b \in B} S^-(b)$ is the set of minorants of B.

An element a of $S[B] \cap B$ (if there is one) is called a **maximum** of B and we write $a = \max B$. If a exists, it is unique, for if a and a' are two maxima in B, then $a \geqq a'$ and $a' \geqq a$, whence $a = a'$. In a similar manner, an element a of $S^-[B] \cap B$ (which is unique if it exists) is called a **minimum** of B and we write $a = \min B$.

If the set $S[B]$ of majorants has a minimum $a = \min S[B]$, then this element is called the **supremum** of B and we write $a = \sup B$. Intuitively, $\sup B$ is the 'least upper bound' of B. In the case in which B has a maximum, then B has a supremum $\sup B = \max B$. The idea of supremum is an extension of the idea of maximum.

In a similar way, the **infimum** $\inf B$ is defined to be $\max S^-[B]$ (the 'greatest lower bound' of B).

EXAMPLE 1. In the ordered set $(\mathbf{R}, \geqq)$ formed by the straight line $\mathbf{R}$ and the relation 'greater than or equal to', the set

$$B = \{1, 1/2, 1/3, 1/4, \ldots\}$$

has a maximum, which is 1, but has no minimum. On the other hand, it has an infimum, namely 0.

It is known that *in the ordered set* $(\hat{\mathbf{R}}, \geqq)$, *every subset B admits a supremum and an infimum* (Cantor-Dedekind).

If f is a numerical function defined on a set A, the supremum of the set $f(A) = \{f(x) \,/\, x \in A\}$ is, in general, denoted by $\sup\limits_{x \in A} f(x)$; in particular, if the supremum is attained in A, it is denoted by $\max\limits_{x \in A} f(x)$.

EXAMPLE 2. In the ordered set $(\mathscr{P}(X), \supset)$ we can verify that a collection $\mathscr{A} = (A_i \,/\, i \in I)$ of subsets of X admits a supremum and an infimum, namely

$$\sup \mathscr{A} = \bigcup_{i \in I} A_i,$$

$$\inf \mathscr{A} = \bigcap_{i \in I} A_i.$$

If $\mathscr{A}$ is a complete lattice with respect to $\cup$ and $\cap$, we can write

$$\max \mathscr{A} = \bigcup_{i \in I} A_i,$$

$$\min \mathscr{A} = \bigcap_{i \in I} A_i.$$

If every non-empty subset $B \subset X$ has a minimum, we say that the ordered set $(X, \geqq)$ is **well-ordered**; every subset of a well-ordered set is also well-ordered. $(\mathrm{N}, \geqq)$ is well-ordered but $(\mathbf{R}, \geqq)$ is not.

An ordering $\geqq$ defined on a set X, is called **latticial** if

$$(\forall\, x)\,(\forall\, y)\,(\exists\, a) : a = \sup \{x, y\}.$$

An ordering $\geqq$ is called **total** if

$$(\forall\, x)\,(\forall\, y)\,(\exists\, a) : a = \max \{x, y\}.$$

In the case of a total ordering, any pair of elements x, y is such that either $x \geqq y$ or $y \geqq x$; the pair $(X, \geqq)$ is then called a **totally ordered set.** A well-ordered set is totally ordered, for if $B = \{x, y\}$ has a minimum, then either $x \leqq y$ or $y \leqq x$.

EXAMPLE 1. (Economics.) Let X be a set of possible states; if a given individual prefers state x to state y, we write $x \geqq y$ (x is 'preferable' to y). The relation $\geqq$ thus defined is a total pre-ordering, for

$$\begin{array}{lll} (1)\ x \geqq x, & & \\ (2)\ x \geqq y,\ y \geqq z & \Rightarrow & x \geqq z, \\ (3)\ x, y \in X & \Rightarrow & \text{either } x \geqq y \text{ or } y \geqq x \end{array}$$

(such a relation $\geqq$ is sometimes called a 'quasi-ordering' or a 'preference relation').

EXAMPLE 2. Let X be any set and let f be a numerical function defined on X; we write $x \geqq y$ if (and only if) we have $f(x) \geqq f(y)$; this defines a total pre-ordering.

EXAMPLE 3. Let X be the set $\mathbf{R}^2$ and write $(x, y) \geqq (x', y')$ if either (i) $x > x'$ or (ii) $x = x'$ and $y \geqq y'$. We then obtain a total ordering, for

$$\begin{array}{lll} (1)\ (x, y) \geqq (x, y), & & \\ (2)\ (x, y) \geqq (x', y'),\ (x', y') \geqq (x'', y'') & \Rightarrow & (x, y) \geqq (x'', y''), \\ (3)\ (x, y) \geqq (x', y'),\ (x', y') \geqq (x, y) & \Rightarrow & (x, y) = (x', y'), \\ (4)\ \text{either } (x, y) \geqq (x', y') \text{ or } (x', y') \geqq (x, y). & & \end{array}$$

Economists have raised the question of whether a preference relation can always be expressed as a numerical function as in Example 2 above; we can prove that this is not the case. Suppose that a numerical function f does represent the total ordering that we have just defined, so that

$$(x, y) > (x', y') \quad \Leftrightarrow \quad f(x, y) > f(x', y')$$

Let y_1 and y_2 be two numbers such that $y_1 < y_2$. With each number x we associate the proper closed interval

$$I_x = [f(x, y_1), f(x, y_2)]$$

in $\mathbf{R}$. If $x \neq x'$, we have $I_x \cap I_{x'} = \varnothing$, for

$$x > x' \quad \Rightarrow \quad f(x, y_1) > f(x', y_2).$$

The family of disjoint intervals I_x is denumerable (cf. example 4, § 2) and so we have a contradiction, because there cannot be a one-one correspondence between the non-denumerable numbers x and the denumerable intervals I_x.

§ 5.* Transfinite ordinal numbers

Let $(A, \geqq)$ and $(B, \geqq)$ be two well-ordered sets. A one-one mapping σ of A into B is said to be **order-preserving** if

$$a \geqq a' \quad \Leftrightarrow \quad \sigma a \geqq \sigma a'.$$

For such a mapping we also have

$$\sigma a = \sigma a' \Leftrightarrow \begin{cases} a \geqq a' \\ a' \geqq a \end{cases} \Leftrightarrow a = a'$$

and

$$\sigma a > \sigma a' \Leftrightarrow \begin{cases} \sigma a \geqq \sigma a' \\ \sigma a \neq \sigma a' \end{cases} \Leftrightarrow \begin{cases} a \geqq a' \\ a \neq a' \end{cases} \Leftrightarrow a > a'.$$

We say that two well-ordered sets $(A, \geqq)$ and $(B, \geqq)$ are **similar** if there is a one-one order-preserving mapping of A onto B; we then write $(A, \geqq) \simeq (B, \geqq)$. We can see at once that $\simeq$ is an equivalence relation, for

(1) $(A, \geqq) \simeq (A, \geqq)$,
(2) $(A, \geqq) \simeq (B, \geqq)$, $(B, \geqq) \simeq (C, \geqq)$ imply that $(A, \geqq) \simeq (C, \geqq)$,
(3) $(A, \geqq) \simeq (B, \geqq)$ implies that $(B, \geqq) \simeq (A, \geqq)$.

The **ordinal number** of a well-ordered set $(A, \geqq)$ is defined to be the equivalence class which contains it.

EXAMPLE. Consider the set $(4, 5, 6, \ldots; 1, 2, 3)$ ordered by the position of the elements:

$$4 < 5 < 6 < \ldots < 1 < 2 < 3.$$

This is a well-ordered set; it is not similar to the well-ordered set $(5, 6, 7, \ldots; 1, 2, 3, 4)$ for example. The ordinal numbers of various sets of this form are denoted as follows: we say that

$(1, 2, 3, \ldots n)$	has ordinal number n,
$(1, 2, 3, \ldots n, \ldots)$	has ordinal number ω,
$(2, 3, 4, \ldots; 1)$	has ordinal number $\omega+1$,
$(3, 4, 5, \ldots; 1, 2)$	has ordinal number $\omega+2$,
$(1, 3, 5, \ldots; 2, 4, 6, \ldots)$	has ordinal number 2ω,
$(3, 5, 7, \ldots; 2, 4, 6, \ldots; 1)$	has ordinal number $2\omega+1$, etc.

Ordinal numbers constitute above all a convenient system of notation, as we can see in the following example.

EXAMPLE. For each $n \in \mathrm{N}$, let A_n be a set of cardinal number $\aleph_n$. Then we denote by $\aleph_\omega$ the cardinal number of $A = \bigcup_{n=0}^{\infty} A_n$, and by $\aleph_{\omega+1}$ the cardinal number of $\mathscr{P}(A)$, etc.

§ 6.* The different forms of the axiom of choice

Let $\mathscr{A}$ be a family $(A_i \mid i \in I)$. A set B is said to be **comparable** to $\mathscr{A}$ if for any index i we have either $B \supset A_i$ or $B \subset A_i$. If every set in $\mathscr{A}$ is comparable to $\mathscr{A}$, then we say that $\mathscr{A}$ is **comparable to itself**; in this case, the ordered set $(\mathscr{A}, \supset)$ is totally ordered.

Fixed point theorem. *Let A be a non-empty subset of a set X and let σ be a single-valued mapping of $\mathscr{P}(X)$ into X; we say that a family $\mathscr{A} = (A_i \mid i \in I)$ satisfies property $L(A, \sigma)$ whenever*

(1) $\bigcap_{i \in I} A_i = A \in \mathscr{A}$,

(2) $A_i \cup \{\sigma A_i\} \in \mathscr{A}$,

(3) *every sub-family $(A_j \mid j \in J)$ of $\mathscr{A}$ comparable to itself satisfies* $\bigcup_{j \in I} A_j \in \mathscr{A}$.

Then if $\mathscr{A}$ satisfies the property $L(A, \sigma)$, there exists an index i_0 such that $\sigma A_{i_0} \in A_{i_0}$.

Proof. Let $\mathscr{B}$ be the intersection of all sub-families of $\mathscr{A}$ which satisfy $L(A, \sigma)$; then $\mathscr{B}$ satisfies $L(A, \sigma)$. The theorem can be proved by showing that $\mathscr{B}$ is comparable to itself; for if this is so, then, putting $A_i \cup \{\sigma A_i\} = f(A_i)$, we have

$$\bigcup_{B \in \mathscr{B}} B = B_0 \in \mathscr{B} \quad\Rightarrow\quad f(B_0) \in \mathscr{B} \quad\Rightarrow\quad f(B_0) \subset B_0 \quad\Rightarrow\quad \sigma B_0 \in B_0.$$

That $\mathscr{B}$ is comparable to itself is a consequence of the following two lemmas.

LEMMA 1. *For each set $B \in \mathscr{B}$ comparable to $\mathscr{B}$, we have*

$$\mathscr{C}_B = \{C \mid C \in \mathscr{B};\ C \subset B \text{ or } C \supset f(B)\} = \mathscr{B}.$$

Proof. We show that $\mathscr{C}_B$ satisfies the property $L(A, \sigma)$.

(1) Since $A \in \mathscr{C}_B$, we have $\bigcap_{C \in \mathscr{C}_B} C = A$.

(2) If $C \in \mathscr{C}_B$, we have either $C \supset f(B)$ or $C \subset B$; in the first case we have

$$C \supset f(B) \quad\Rightarrow\quad f(C) \supset C \supset f(B) \quad\Rightarrow\quad f(C) \in \mathscr{C}_B$$

and in the second case

$$C \subset B \Rightarrow \begin{cases} f(C) \subset B, \\ \text{or } C \subset B \subset f(C) \end{cases} \Rightarrow \begin{cases} f(C) \subset B, \\ \text{or } f(C) = B, \\ \text{or } C = B \end{cases}$$
$$\Rightarrow \begin{cases} f(C) \subset B, \\ \text{or } f(C) = f(B) \end{cases} \Rightarrow f(C) \in \mathscr{C}_B.$$

Thus in either case, we have $f(C) \in \mathscr{C}_B$.

(3) Let $\mathscr{D} = (D_k \mid k \in K)$ be a sub-family of $\mathscr{C}_B$ comparable to itself. Then either $D_k \subset B$ for all k or $D_k \supset f(B)$ for one index k. Since

$$(\forall k) : D_k \subset B \quad\Rightarrow\quad \bigcup_k D_k \subset B \quad\Rightarrow\quad \bigcup_k D_k \in \mathscr{C}_B,$$

$$(\exists k) : D_k \supset f(B) \quad\Rightarrow\quad f(B) \subset \bigcup_k D_k \quad\Rightarrow\quad \bigcup_k D_k \in \mathscr{C}_B,$$

it follows that $\mathscr{C}_B$ satisfies (3).

We therefore conclude that $\mathscr{B} \subset \mathscr{C}_B$, and since $\mathscr{C}_B \subset \mathscr{B}$ by definition, we have $\mathscr{B} = \mathscr{C}_B$.

LEMMA 2. *If $\mathscr{B}_1$ is the collection of sets in $\mathscr{B}$ comparable to $\mathscr{B}$, then $\mathscr{B}_1 = \mathscr{B}$.*

Proof. $\mathscr{B}_1$ satisfies $L(A, \sigma)$, for

(1) since $A \in \mathscr{B}_1$, we have $\bigcap_{B \in \mathscr{B}_1} B = A$,
(2) if $B \in \mathscr{B}_1$, we have $f(B) \in \mathscr{B}_1$ since, by Lemma 1,
$$C \in \mathscr{B} \quad \Rightarrow \quad \begin{cases} C \subset B \\ \text{or } C \supset f(B) \end{cases} \quad \Rightarrow \quad \begin{cases} C \subset f(B) \\ \text{or } C \supset f(B), \end{cases}$$
(3) the union of every sub-family of $\mathscr{B}_1$ comparable to itself is in $\mathscr{B}_1$.

Therefore $\mathscr{B}_1 \supset \mathscr{B}$ and so, since $\mathscr{B}_1 \subset \mathscr{B}$, we have $\mathscr{B}_1 = \mathscr{B}$.

COROLLARY. *Let $(X, \geqq)$ be an ordered set in which every totally ordered subset B has a supremum $b = \sup B$. If f is a single-valued mapping of X into X such that $f(a) \geqq a$ for all $a \in X$, then there exists an element a_0 in X such that $f(a_0) = a_0$.*

Proof. Let x_0 be an element in X and let $\mathscr{A} = (A_i \,/\, i \in I)$ be the collection of subsets of X such that

(1) $x_0 \in A_i$,
(2) A_i is totally ordered with respect to the relation $\geqq$.

Let σ be the single-valued mapping of $\mathscr{A}$ into X defined as follows:
$$\sigma(A_i) \begin{cases} = f(\max A_i) \text{ if } \max A_i \text{ exists,} \\ = \sup A_i \text{ otherwise.} \end{cases}$$
Then $\mathscr{A}$ satisfies property $L(\{x_0\}, \sigma)$, for if $(A_i \,/\, i \in J)$ is a sub-family of $\mathscr{A}$ comparable to itself, the union of this sub-family is in $\mathscr{A}$. By the fixed point theorem there exists an index i such that
$$\sigma(A_i) \in A_i;$$
hence there exists an element $\max A_i = a_0$ such that $f(a_0) = \sigma(A_i) \in A_i$. Then
$$a_0 \geqq f(a_0) \geqq a_0$$
and so we have $f(a_0) = a_0$.

In an ordered set $(X, \geqq)$, an element a of X is called **maximal** if
$$x \in X,\ x \geqq a \quad \Rightarrow \quad x = a.$$
If the element $\max X$ exists, then it is clearly a maximal element; moreover it is then the only maximal element, for if $x = \max X$ and a is another maximal element, then $x \in X$ and $x \geqq a$, whence $x = a$. In general, however, there may be more than one maximal element.

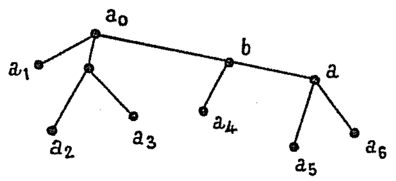


FIG. 14

EXAMPLE. Consider the tree shown in figure 14. Put $a \geqq b$ to signify that a is further from the origin than b; then $\geqq$ is an ordering relation. $(X \geqq)$ is not totally ordered (the elements a and a_2, for example, are not comparable). The maximal elements are a_1, a_2, a_3, a_4, a_5, a_6.

Fundamental theorem (Zorn). *The axiom of choice can be stated in any one of the following equivalent forms:*

(1) *Given a set X, there exists a single-valued mapping γ in which to every non-empty subset B there corresponds an element $\gamma(B)$ in X such that $\gamma(B) \in B$.*

(2) *If X is an ordered set such that every totally ordered subset of X has a supremum, then X has a maximal element.*

(3) *If $(X, \geqq)$ is an ordered set, the collection $(\mathscr{A}, \supset)$ of all totally ordered subsets of X contains a maximal element.*

(4) *If $(X, \geqq)$ is an ordered set such that every totally ordered subset of X has a majorant, then X has a maximal element.*

(5) *If X is any set, there exists an ordering relation $\geqq$ such that $(X, \geqq)$ is a well-ordered set.*

Proof. The axiom of choice, which was stated above (§ 3, Chapter I) is clearly equivalent to condition (1); we shall prove that

$$(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (5) \Rightarrow (1).$$

(1) implies (2). Let $(X, \geqq)$ be an ordered set such that every totally ordered subset has a supremum and let f be the function defined as follows:

$$f(x)\begin{cases} = \gamma(A_x) & \text{if} \quad A_x = \{y \,/\, y \in X,\ y > x\} \neq \varnothing, \\ = x & \text{if} \quad A_x = \varnothing. \end{cases}$$

Then $f(x) \geqq x$ and so, by the corollary to the fixed point theorem, there exists an element a such that $a = f(a)$. Therefore $A_a = \varnothing$ and hence the element a is maximal.

(2) implies (3). Let $(X, \geqq)$ be an ordered set and let $\mathscr{A}$ be the collection of subsets of X totally ordered with respect to $\geqq$. Since $(\mathscr{A}, \supset)$ satisfies the hypotheses of (2), $\mathscr{A}$ contains a maximal element A_0:

$$A \supset A_0,\ A \in \mathscr{A} \quad \Rightarrow \quad A = A_0.$$

(3) implies (4). Let $(X, \geqq)$ be an ordered set satisfying the hypothesis of (4); then, if (3) holds, X contains a maximal totally ordered subset A_0. By hypothesis, A_0 has a majorant a and

$$x \geqq a \Rightarrow A_0 \cup \{x\} \text{ totally ordered} \Rightarrow x \in A_0 \Rightarrow \begin{Bmatrix} a \geqq x \\ \text{and} \\ x \geqq a \end{Bmatrix} \Rightarrow x = a.$$

Then the element a is maximal in X.

(4) implies (5). Let X be any set and let $\mathscr{A} = (A_i \,/\, i \in I)$ be the family of subsets A_i which can be well-ordered, the ordering relation for A_i being denoted by $\geqq^i$. Put $A_i \succ A_j$ if

$$\begin{aligned}
&(1)\ A_i \supset A_j \\
&(2)\ y \in A_i - A_j,\ x \in A_j \quad \Rightarrow \quad y \geqq^i x \\
&(3)\ x, y \in A_j,\ y \geqq^j x \quad \Rightarrow \quad y \geqq^i x
\end{aligned}$$

We can verify at once that $\succ$ is an ordering relation, for

$$\begin{aligned}
&A_i \succ A_i, \\
&A_i \succ A_k,\ A_k \succ A_j \quad \Rightarrow \quad A_i \succ A_j, \\
&A_i \succ A_j,\ A_j \succ A_i \quad \Rightarrow \quad A_i = A_j.
\end{aligned}$$

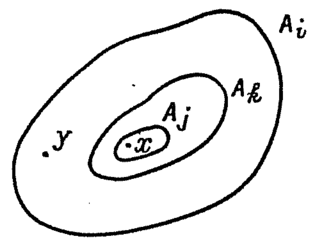


FIG. 15

If $(A_i \,/\, i \in J)$ is a sub-family totally ordered by $\succ$ and if $A = \bigcup_{i \in J} A_i$, then

$$x, y \in A \quad \Rightarrow \quad (\exists_J i, j) : x \in A_i,\ y \in A_j \quad \Rightarrow \quad (\exists_J k) : x, y \in A_k$$

(for either $A_i \supset A_j$ or $A_j \supset A_i$). Writing $x \geqq y$ if $x \geqq^k y$, we define a well-ordered set $(A, \geqq)$. Since A is a majorant of $(A_i \,/\, i \in J)$, $\mathscr{A}$ satisfies the hypotheses of (4). Therefore amongst the well-ordered sets $(A_i, \geqq^i)$ there exists a maximal set $(A_0, \geqq^0)$.

Suppose that A_0 is not the full set X. Then there is an element a in $X - A_0$ and $A_0 \cup \{a\}$ is well-ordered by the relation $\geqq$ defined by

$$\begin{aligned}
&(1)\ x \geqq y \quad \text{if} \quad x, y \in A_0;\ x \geqq^0 y, \\
&(2)\ a > x \quad \text{if} \quad x \in A_0.
\end{aligned}$$

This implies that A_0 is not a maximal totally ordered set and so leads to a contradiction; hence $A_0 = X$. Thus $(X, \geqq^0)$ is a well-ordered set as required.

(5) implies (1). Let X be any set; if the set $(X, \geqq)$ is well-ordered with respect to the ordering relation $\geqq$ and if $B \subset X$, $B \neq \varnothing$, then there is a single-valued mapping γ of $\mathscr{P}(X)$ into X defined by

$$\gamma(B) = \min B \in B.$$

This completes the proof of the theorem.

A very convenient form of the axiom of choice is given in the following theorem.

Zorn's Theorem. *If every totally pre-ordered subset of a pre-ordered set X has a majorant, then there is an element a in X such that*

$$x \geqq a,\ x \in X \quad \Rightarrow \quad x \equiv a.$$

Proof. If x^* denotes the equivalence class, with respect to $\equiv$, which contains x, then we write $x^* \geqq^* y^*$ whenever $x \geqq y$. The set X^* of equivalence classes is then ordered by $\geqq^*$ and a subset B^* of X^* admits a majorant if it is totally ordered. Therefore X^* contains a maximal element a^* and so we have

$$x \geqq a \quad \Rightarrow \quad x^* \geqq^* a^* \quad \Rightarrow \quad x^* =^* a^* \quad \Rightarrow \quad x \equiv a.$$

COROLLARY (Bourbaki). *Given a filter base $\mathscr{B}$, there exists an ultra-filter base $\mathscr{B}_0$ which is finer than $\mathscr{B}$ (that is, $\mathscr{B}_0 \vdash \mathscr{B}$).*

Proof. Let $\mathscr{B}$ be a filter base in a set A and let $\Phi = \{\mathscr{B}_i \,/\, i \in I\}$ be the collection of filter bases in A which are finer than $\mathscr{B}$. As can easily be seen, $\vdash$ is a pre-ordering in Φ. Moreover, if $\{\mathscr{B}_j \,/\, j \in J\}$ is a sub-collection totally pre-ordered by $\vdash$, it admits as a majorant the filter base consisting of sets of the form

$$B_K = \bigcap_{i \in K} B_i,$$

where $B_i \in \mathscr{B}_i$ for all $i \in K$ and K is a finite subset of J.

By Zorn's theorem, there is a base $\mathscr{B}_0$ in Φ such that, for any filter base $\mathscr{B}_1$,

$$\mathscr{B}_1 \vdash \mathscr{B}_0 \quad \Rightarrow \quad \mathscr{B}_1 \vdash \mathscr{B}_0 \vdash \mathscr{B} \quad \Rightarrow \quad \begin{Bmatrix} \mathscr{B}_1 \in \Phi \\ \mathscr{B}_1 \vdash \mathscr{B}_0 \end{Bmatrix} \quad \Rightarrow \quad \mathscr{B}_1 \simeq \mathscr{B}_0.$$

Hence, by Theorem 5, § 6, Chapter I, $\mathscr{B}_0$ is an ultra-filter base.

CHAPTER IV

TOPOLOGICAL SPACES

§ 1. Metric spaces

Let X be any set. A numerical function d defined on $X \times X$ is called a **distance function** or **metric** if

$$\begin{aligned}
&(1) \quad d(x, y) \geqq 0,\\
&(2) \quad d(x, y) = 0 \quad \Leftrightarrow \quad x = y,\\
&(3) \quad d(x, y) = d(y, x),\\
&(4) \quad d(x, y) + d(y, z) \geqq d(x, z).
\end{aligned}$$

The inequality (4) is the only condition likely to cause difficulty; it is called the **triangular inequality**.

The pair (X, d) consisting of the set X and the metric d, is called a **metric space** and $d(x, y)$ is called the **distance** between x and y. If in any given situation only one metric on X is involved and no confusion is possible, we follow the convention of referring to X as the metric space.

EXAMPLE 1. Let X be the set of towns on a geographical map and let $d(x, y)$ be the length of the shortest route by road from town x to town y. Clearly d satisfies the axioms (1) to (4) for a metric.

EXAMPLE 2. Let $X = \mathbf{C}$, the set of complex numbers. We can define a metric d by writing $d(z_1, z_2) = |z_1 - z_2|$, where $|z_1 - z_2|$ is the modulus of $z_1 - z_2$.

EXAMPLE 3. Let X be the Euclidean space $\mathbf{R}^n$ of n dimensions; if

$$x = (x^1, x^2, \ldots, x^n)$$

and

$$y = (y^1, y^2, \ldots, y^n)$$

are two points of $\mathbf{R}^n$, we write

$$d(x, y) = \sqrt{(x^1 - y^1)^2 + (x^2 - y^2)^2 + \ldots + (x^n - y^n)^2}$$

and this determines a metric d on $\mathbf{R}^n$. This is called the **Euclidean metric** on $\mathbf{R}^n$ and, in what follows, $\mathbf{R}^n$ will always be considered as a metric space with this function d as metric. In $\mathbf{R}$, the Euclidean metric is simply given by $d(x, y) = |x - y|$.

The triangular inequality in $\mathbf{R}^n$ can be proved by a similar method to that used in proving it for $\mathbf{R}^3$. It follows from a general inequality which we shall prove later (Chapter VIII, § 9).

EXAMPLE 4. Let X be any set. Then a metric d can be defined by

$$d(x, y) \begin{cases} = 0 & \text{if } x = y, \\ = 1 & \text{if } x \neq y. \end{cases}$$

EXAMPLE 5. Let X be the set Φ of numerical functions defined on the segment $[0, 1]$. A metric d on X is defined as follows:

$$d(f, g) = \sup \{ |f(x)-g(x)| \,/\, x \in [0, 1]\}.$$

The triangular inequality is satisfied, since

$$\begin{aligned} d(f, g)+d(g, h) &= \sup |f(x)-g(x)| + \sup |g(x)-h(x)| \\ &\geqq \sup (|f(x)-g(x)|+|g(x)-h(x)|) \geqq \sup |f(x)-h(x)|. \end{aligned}$$

In a metric space X, the set

$$B_\lambda(x_0) = \{x \,/\, d(x, x_0) \leqq \lambda\}$$

is called the **ball** of centre x_0 and radius λ. The set

$$S_\lambda(x_0) = \{x \,/\, d(x, x_0) = \lambda\}$$

is called the **sphere** of centre x_0 and radius λ.

A point a in X is called an **interior point** of the set A if there exists a strictly positive number ε such that $B_\varepsilon(a) \subset A$; the set of interior points of A is denoted by $\mathring{A}$ and is called the **interior** of A. Clearly $\mathring{A} \subset A$.

A point a in X is called a **point of closure** of a set A if for any strictly positive number ε, we have $B_\varepsilon(a) \cap A \neq \emptyset$. The set of points of closure of the set A is denoted by $\bar{A}$ and is called the **closure** of A. Clearly $\bar{A} \supset A$.

It is possible for a point a to be a point of closure of A without being in A; for example if X is the real line $\mathbf{R}$ and A is the set $\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$, the point 0 is a point of closure of A but does not belong to A.

A set G such that every point of G is an interior point of G is said to be **open**. A set F such that every point of closure of F is in F is said to be **closed**.

EXAMPLE 1. Let X be the real line $\mathbf{R}$ and let d be the Euclidean metric defined by $d(x, y) = |x-y|$. The segment $[\lambda, \mu] = \{x \,/\, x \geqq \lambda,\ x \leqq \mu\}$ is a closed set; the interval $]\lambda, \mu[$ is an open set; the union $]\lambda, \mu[\cup]\lambda', \mu'[$ is an open set; the intervals $]\lambda, \mu]$ and $[\lambda, \mu[$ are neither open nor closed.

Using Example 4, § 2, Chapter III, we can prove the following result: *a subset of* $\mathbf{R}$ *is open if and only if it is the union of a countable number of disjoint open intervals.*

EXAMPLE 2. Let X be any metric space. The set

$$A = B_\lambda(a) - S_\lambda(a)$$

is open. For suppose that $x_0 \in A$; then $d(x_0, a) < \lambda$. Hence there is a number ε such that $0 < \varepsilon < \lambda - d(x_0, a)$; since

$$x \in B_\varepsilon(x_0) \quad \Rightarrow \quad d(x, x_0) \leqq \varepsilon \quad \Rightarrow$$
$$d(x, a) \leqq d(x, x_0)+d(a, x_0) < \varepsilon+(\lambda-\varepsilon) = \lambda \quad \Rightarrow \quad x \in A,$$

we have $B_\varepsilon(x_0) \subset A$ and therefore x_0 is an interior point of A, whence A is open.

EXAMPLE 3. Let X be any metric space. The set $\{a\}$ which has only one element a is closed, for if x is a point of closure of this set, then

$$(\forall\varepsilon) : a \in B_\varepsilon(x) \quad \Rightarrow \quad (\forall\varepsilon) : d(x, a) \leqq \varepsilon \quad \Rightarrow \quad d(x, a) = 0 \quad \Rightarrow \quad x = a.$$

Theorem 1. *If $\mathscr{G} = (G_i \mid i \in I)$ is the collective family of open sets, then*

$$(1) \quad J \subset I \quad \Rightarrow \quad \bigcup_{i \in J} G_i \in \mathscr{G},$$
$$(2) \quad i_1, i_2, \ldots, i_n \in I \quad \Rightarrow \quad \bigcap_{i=i_1}^{i_n} G_i \in \mathscr{G},$$
$$(3) \quad \emptyset \in \mathscr{G}, \ X \in \mathscr{G}.$$

Proof. Let a be a point of $\bigcup_{i \in J} G_i$. Then a is an interior point of this set, because

$$(\exists i) : a \in G_i \quad \Rightarrow \quad (\exists i)(\exists\varepsilon) : B_\varepsilon(a) \subset G_i \quad \Rightarrow \quad (\exists\varepsilon) : B_\varepsilon(a) \subset \bigcup_{i \in J} G_i.$$

Therefore the union of the G_i is an open set.

Suppose now that b is a point of the finite intersection $\bigcap_{i=i_1}^{i_n} G_i$. For each $i = i_1, i_2, \ldots, i_n$ there exists a strictly positive number ε_i such that $B_{\varepsilon_i}(b) \subset G_i$. If we denote the smallest of the numbers $\varepsilon_{i_1}, \varepsilon_{i_2}, \ldots, \varepsilon_{i_n}$ by ε, we have

$$B_\varepsilon(b) \subset \bigcap_{i=i_1}^{i_n} G_i.$$

Therefore b is an interior point of the set $\bigcap_{i=i_1}^{i_n} G_i$ and so this set is open.

This proves parts (1) and (2) of the theorem; the proof of part (3) is immediate.

Theorem 2. *The complement of an open set is closed and the complement of a closed set is open.*

Proof. Let F be a closed set and write $A = -F$. If $a \notin F$, there exists a ball $B_\varepsilon(a)$ which does not meet F. Therefore $B_\varepsilon(a) \subset -F = A$ and so a is an interior point of A, whence A is an open set.

Let G be an open set and let a be a point of closure of $-G$. Then $a \in -G$, for otherwise there would be a ball $B_\varepsilon(a)$ such that $B_\varepsilon(a) \subset G$; this implies that $B_\varepsilon(a) \cap -G = \emptyset$ and so contradicts the fact that a is a point of closure of $-G$.

COROLLARY 1. *If $\mathscr{F} = (F_i \mid i \in I)$ is the collective family of closed sets, then*

$$(1) \quad J \subset I \quad \Rightarrow \quad \bigcap_{i \in J} F_i \in \mathscr{F},$$

$$(2) \quad i_1, i_2, \ldots, i_n \in I \quad \Rightarrow \quad \bigcup_{i=i_1}^{i_n} F_i \in \mathscr{F},$$

$$(3) \quad \varnothing \in \mathscr{F}, \ X \in \mathscr{F}.$$

Proof. The set $F_i \in \mathscr{F}$ if and only if $-F_i \in \mathscr{G}$. Then the intersection $\bigcap_{i \in J} F_i$ belongs to $\mathscr{F}$, for

$$-\bigcap_{i \in J} F_i = \bigcup_{i \in J} (-F_i) \in \mathscr{G}.$$

A similar reasoning can be applied to the finite union $\bigcup_{i=i_1}^{i_n} F_i$.

COROLLARY 2. *The ball $B_\lambda(a)$ is a closed set.*

Proof. If $x_0 \in -B_\lambda(a)$, then $d(a, x_0) > \lambda$ and so there exists a number ε such that $0 < \varepsilon < d(a, x_0) - \lambda$. Then

$$x \in B_\varepsilon(x_0) \Rightarrow d(x, a) \geqq d(a, x_0) - d(x, x_0) \geqq d(a, x_0) - \varepsilon > \lambda \Rightarrow x \in -B_\lambda(a)$$

and therefore $-B_\lambda(a)$ is an open set.

Let X and Y be two metric spaces, having metrics d_X and d_Y respectively. A single-valued mapping σ of X into Y is said to be **continuous at the point** x_0 if to each number $\varepsilon > 0$ there corresponds a number $\eta > 0$ such that

$$d_X(x, x_0) \leqq \eta \quad \Rightarrow \quad d_Y(\sigma x, \sigma x_0) \leqq \varepsilon.$$

A single-valued mapping σ is called **continuous (on X)** if it is continuous at each point of X. If $X = Y = \mathbf{R}$, this definition reduces to the well-known one of continuity of a numerical function.

Theorem 3. *A necessary and sufficient condition for the single-valued mapping σ of X into Y to be continuous at x_0 is that, for any open set G of Y containing $y_0 = \sigma x_0$, the point x_0 is an interior point of $\sigma^{-1}G$.*

Proof. Suppose that x_0 is an interior point of $\sigma^{-1}G$ for each open set G of Y containing $y_0 = \sigma x_0$. Choose G to be the set $B_\varepsilon(y_0) - S_\varepsilon(y_0)$. Then there exists a ball $B_\eta(x_0)$ contained in $\sigma^{-1}G$ and we have

$$d_X(x, x_0) \leqq \eta \Rightarrow x \in B_\eta(x_0) \Rightarrow \sigma x \in G \Rightarrow d_Y(\sigma x, y_0) \leqq \varepsilon.$$

Suppose conversely that σ is continuous at x_0 and let G be an open set of Y containing $y_0 = \sigma x_0$. Then there exists a number ε such that $B_\varepsilon(y_0) \subset G$. To this number ε there corresponds a number η such that

$$d_X(x, x_0) \leqq \eta \quad \Rightarrow \quad d_Y(\sigma x, y_0) \leqq \varepsilon,$$

whence

$$\sigma B_\eta(x_0) \subset B_\varepsilon(y_0)$$

and so

$$B_\eta(x_0) \subset \sigma^{-1} B_\varepsilon(y_0) \subset \sigma^{-1} G.$$

Therefore x_0 is an interior point of $\sigma^{-1}G$ and so the set $\sigma^{-1}G$ is open.

By means of this theorem, we can study the idea of continuity of a single-valued function in a very general setting.

§ 2.* L*- and L⁰- spaces

A sequence $(x_n) = (x_1, x_2, x_3, \ldots)$ of elements of a set X is an element of the product set $X^\omega = X \times X \times X \times \ldots$. We say that a sequence $(x'_n) = (x'_1, x'_2, x'_3, \ldots)$ is a **sub-sequence** of the sequence (x_n) if

$$\begin{cases} x'_1 = x_{k_1}; \quad x'_2 = x_{k_2}; \quad x'_3 = x_{k_3}; \text{ etc.} \ldots \\ 1 \leqq k_1 < k_2 < k_3 < \ldots \end{cases}$$

and we write $(x'_n) \vdash (x_n)$.

A mapping Λ of X^ω into X is called a **convergence** if the following conditions are satisfied:

(1) $\left.\begin{matrix} x_0 \in \Lambda(x_n) \\ (x'_n) \vdash (x_n) \end{matrix}\right\} \quad \Rightarrow \quad x_0 \in \Lambda(x'_n),$

(2) $(\forall i) : x_i = x_0 \quad \Rightarrow \quad x_0 \in \Lambda(x_n),$

(3) $x_0 \in \Lambda(x_n)$ whenever $(\forall (x'_n);\ (x'_n) \vdash (x_n))\ (\exists (x''_n);\ (x''_n) \vdash (x'_n)) : x_0 \in \Lambda(x''_n)$.

If $x_0 \in \Lambda(x_n)$, then we say that (x_n) **converges** to x_0 and this is sometimes written $(x_n) \to x_0$.

A pair (X, Λ) consisting of a set X and a convergence Λ is called an **L*-space**; where the convergence Λ is fixed, so that there is no possibility of confusing it with another convergence, we shall refer to the set X itself as an L^*-space.

EXAMPLE 1. *Numerical convergence.*

Let X be the real line **R**. A sequence $(x_n) = (x_1, x_2, x_3, \ldots)$ is said to **converge numerically** to a if, given any strictly positive number ε, there is an integer $n(\varepsilon)$ such that

$$k \geqq n(\varepsilon) \quad \Rightarrow \quad |x_k - a| \leqq \varepsilon.$$

In other words, we have

$$(\forall \varepsilon;\ \varepsilon > 0)\ (\exists n)\ (\forall k;\ k \geqq n) : |x_k - a| \leqq \varepsilon.$$

We remark that **R** is a metric space with metric

$$d(x, y) = |x - y|,$$

and that a convergence can be defined in any metric space in a similar manner: the sequence (x_n) converges to x_0 if, given any strictly positive number ε, there is an integer $n(\varepsilon)$ such that

$$k \geqq n(\varepsilon) \quad \Rightarrow \quad d(x_k, a) \leqq \varepsilon.$$

Axioms (1) and (2) are clearly satisfied. To prove axiom (3), suppose that $x_0 \notin \Lambda(x_n)$; then there exists a number ε such that $d(x_k, x_0) > \varepsilon$ for an infinity of values of k, say $k_1, k_2, k_3, \ldots$, such that $k_1 < k_2 < k_3 < \ldots$ We have

$$(x_{k_n}) \vdash (x_n)$$

and (x_{k_n}) contains no subsequences converging to x_0.

EXAMPLE 2. *Simple convergence on* [0, 1].

Let $X = \Phi$, the set of numerical functions defined on [0, 1]; we say that a sequence (f_n) of elements of Φ **converges simply** to f_0 on [0, 1] if, for any x in [0, 1], the numerical sequence $(f_n(x))$ converges numerically to $f_0(x)$; in other words we have

$$(\forall x)\,(\forall \varepsilon)\,(\exists n)\,(\forall k;\ k \geqq n) : |f_k(x) - f_0(x)| \leqq \varepsilon.$$

Clearly axioms (1) and (2) for a convergence are satisfied. To prove axiom (3), suppose that (f_n) does not converge simply to f_0; then

$$(\exists x_0)\,(\exists \varepsilon_0)\,(\forall n)\,(\exists k;\ k > n) : |f_k(x_0) - f_0(x_0)| > \varepsilon_0.$$

Thus, taking $n = 1$, there is an integer k_1 such that

$$|f_{k_1}(x_0) - f_0(x_0)| > \varepsilon_0;$$

then, taking $n = k_1$, there is an integer k_2 such that

$$|f_{k_2}(x_0) - f_0(x_0)| > \varepsilon_0.$$

Continuing this process in the obvious way, there is a sequence $(f_{k_1}, f_{k_2}, f_{k_3}, \ldots)$ which does not contain a sub-sequence converging to f_0. Therefore axiom (3) is satisfied.

EXAMPLE 3. *Uniform convergence on* [0, 1].

Let $X = \Phi$, as in example 2. A sequence (f_n) is said to **converge uniformly** on [0, 1] to $f_0 \in \Phi$ if

$$(\forall \varepsilon)\,(\exists n_0)\,(\forall k;\ k \geqq n_0) : \ \sup \{\, |f_k(x) - f_0(x)| \,/\, x \in [0, 1] \,\} \leqq \varepsilon.$$

This convergence is that determined by the metric

$$d(f, g) = \sup \{\, |f(x) - g(x)| \,/\, x \in [0, 1] \,\}.$$

We note that a sequence of functions can converge on $[0, 1]$ without converging uniformly. For example the sequence of functions f_n defined by

$$f_n(x) \begin{cases} = x^n & \text{if } x < 1 \\ = 0 & \text{if } x = 1 \end{cases}$$

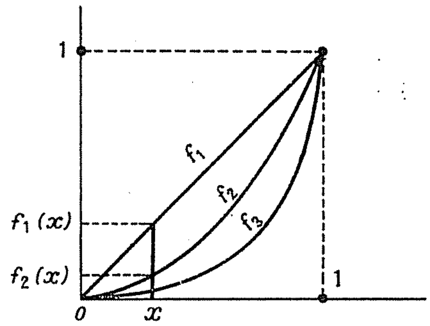


FIG. 16

converges to the function f_0 which is identically zero on $[0, 1]$; however, it does not converge uniformly, since, for each n,

$$\sup \{ |f_n(x)| \ / \ x \in [0, 1] \} = 1.$$

EXAMPLE 4. *Uniform convergence on a family of sets.*

A sequence $(f_1, f_2, f_3, \ldots)$ of functions in Φ is said to *converge uniformly on a family* $\mathscr{A} = (A_i \ / \ i \in I)$ contained in $\mathscr{P}([0, 1])$ if, for each i, the sequence (f_n) converges uniformly on A_i; in other words

$$(\forall i)\,(\forall \varepsilon)\,(\exists n)\,(\forall k;\ k \geqq n) : \sup_{x \in A_i} |f_k(x) - f_0(x)| \leqq \varepsilon.$$

As in the previous examples, we can easily verify that the axioms for a convergence are satisfied.

If $\mathscr{A}$ is the partition determined by the single set $[0, 1]$ we have uniform convergence; if $\mathscr{A}$ is the partition determined by the sets A_i each consisting of a single element, we have simple convergence.

In an L^*-space, a set F is said to be **closed** if

$$x_n \in F,\ (x_n) \to a \text{ imply that } a \in F.$$

Theorem 1. *The family* $\mathscr{F} = (F_i \ / \ i \in I)$ *of closed sets satisfies*

$$(1)\quad J \subset I \quad \Rightarrow \quad \bigcap_{i \in J} F_i \in \mathscr{F},$$

$$(2)\quad i_1, i_2, \ldots, i_k \in I \quad \Rightarrow \quad \bigcup_{i=i_1}^{i_k} F_i \in \mathscr{F},$$

$$(3)\quad \varnothing \in \mathscr{F};\ X \in \mathscr{F}.$$

Proof. The set $\bigcap_{i \in J} F_i$ is closed, for

$$x_n \in \bigcap F_i,\ (x_n) \to a \quad \Rightarrow \quad a \in \bigcap F_i.$$

The union $\bigcup_{i=i_1}^{i_k} F_i$ is closed, for if (x_n) is a sequence of elements in this union converging to a point a, then there is an infinity of elements (x_n) belonging to the same set F_i; these elements form a sequence $(y_n) \vdash (x_n)$. Therefore, by axiom (1) for a convergence, $(y_n) \to a$ and so we have $a \in F_i \subset \bigcup_{i=i_1}^{i_k} F_i$. Finally, $\varnothing$ and X are closed; this is trivial.

Suppose now that we call the complement of a closed set 'open'. Then, by taking complements in Theorem 1, we can prove the following result.

COROLLARY. *The family $\mathscr{G}$ of open sets satisfies*:

$$\begin{aligned}
&(1)\quad J \subset I \quad\Rightarrow\quad \bigcup_{i\in J} G_i \in \mathscr{G},\\
&(2)\quad i_1, i_2, \ldots, i_n \in I \quad\Rightarrow\quad \bigcap_{i=i_1}^{i_n} G_i \in \mathscr{G},\\
&(3)\quad \varnothing \in \mathscr{G};\ X \in \mathscr{G}.
\end{aligned}$$

If X and Y are two L^*-spaces, a single-valued mapping σ of X into Y is called **L*-continuous at the point** x_0 if

$$(x_n) \to x_0 \text{ implies that } (\sigma x_n) \to \sigma x_0.$$

A mapping σ is called **L*-continuous (on** X**)** if it is L^*-continuous at each point of X. This concept has certain similarities to the concept of continuity studied in § 1, the most interesting cases being those in which the convergence Λ satisfies the additional axiom:

(4) *For any subset A of X, the intersection $\bar{A}$ of the closed sets which contain A consists of the points a for which there exists a sequence of elements of A converging to a.*

For a convergence Λ which satisfies (4), we say that the pair (X, Λ) is an L^0-space; it is easily verified that a metric space is an L^0-space (with respect to the convergence introduced in example 1).

We now prove the following theorem.

Theorem 2. *If X and Y are two L^*-spaces and σ is a single-valued L^*-continuous mapping of X into Y, then $\sigma^{-1}F$ is closed for each closed set F of Y and $\sigma^{-1}G$ is open for each open set G of Y. If Y is an L^0-space, then the converse is true: that is, if σ is a single-valued mapping of X into Y such that $\sigma^{-1}F$ is closed for each closed set F of Y, then σ is L^*-continuous.*

Proof. Let σ be a single-valued L^*-continuous mapping of X into Y and let F be a closed set in Y; then $\sigma^{-1}F$ is closed, since

$$\left\{\begin{matrix}(x_n)\to x_0\\ x_n\in\sigma^{-1}F\end{matrix}\right\} \Rightarrow \left\{\begin{matrix}(\sigma x_n)\to\sigma x_0\\ \sigma x_n \in F\end{matrix}\right\} \Rightarrow \sigma x_0 \in F \Rightarrow x_0 \in \sigma^{-1}F$$

Suppose now that Y is an L^0-space and that σ is a single-valued mapping of X into Y such that $\sigma^{-1}F$ is closed for each closed set F in Y. If σ is not L^*-continuous, there exists a sequence (x_n) converging to x_0 such that (σx_n) does not converge to $\sigma x_0 = y_0$. By axiom (3), there exists a sub-sequence (σx_{k_n}) which does not contain any sub-sequence converging to y_0. If $A = \{\sigma x_{k_1}, \sigma x_{k_2}, \ldots\}$, then, by axiom (4), $y_0 \notin \bar{A}$, whence $x_0 \notin \sigma^{-1}\bar{A}$. But $\sigma^{-1}\bar{A}$ is a closed set, and so

$$(x_{k_n}) \to x_0, \quad x_{k_n} \in \sigma^{-1}\bar{A} \quad\Rightarrow\quad x_0 \in \sigma^{-1}\bar{A},$$

which gives a contradiction. Therefore σ is L^*-continuous.

§ 3. Topological spaces

DEFINITION 1. A **topological structure** or a **topology** in a set X is a collective family $\mathscr{G} = (G_i \,/\, i \in I)$ of subsets of X satisfying

(1) $J \subset I \quad \Rightarrow \quad \bigcup_{i \in J} G_i \in \mathscr{G}$,

(2) J finite; $J \subset I \quad \Rightarrow \quad \bigcap_{i \in J} G_i \in \mathscr{G}$,

(3) $\emptyset \in \mathscr{G}, \; X \in \mathscr{G}$.

The pair $(X, \mathscr{G})$ is called a **topological space**(1) and the sets in $\mathscr{G}$ are called **open sets.** Where the topology $\mathscr{G}$ is fixed and no confusion is possible, we shall refer to the set X itself as the topological space. A metric space will always be considered to have the topology $\mathscr{G}$ defined in § 1. Thus the concept of topological space can be regarded as a generalisation of that of metric space.

EXAMPLE 1. Let X be any set and let $\mathscr{G} = \mathscr{P}(X)$, the family of all subsets of X. The axioms (1), (2) and (3) are easily seen to be satisfied and so we have a topology; this is called the **discrete topology** on X.

EXAMPLE 2. Let X be any set and let $\mathscr{G} = (\emptyset, X)$ consist only of the empty set and the whole set X. This topology is called the **coarsest topology.**

More generally, if two topologies $\mathscr{G}_1$ and $\mathscr{G}_2$ are such that $\mathscr{G}_1 \supset \mathscr{G}_2$, we say that $\mathscr{G}_1$ is **finer** than $\mathscr{G}_2$ or that $\mathscr{G}_2$ is **coarser** than $\mathscr{G}_1$.

EXAMPLE 3. Let X be any set; if $\mathscr{A} = (A_i \,/\, i \in I)$ is a family of sets in X, then the 'smallest' topology $\mathscr{G}$ which contains the A_i is formed as follows: let $\mathscr{B}$ be the collective family $(B_j \,/\, j \in J)$, where B_j is the intersection of a finite number of the sets A_i; then let $\mathscr{C}$ be the collective family $(C_k \,/\, k \in K)$, where C_k is a union of sets B_j or the empty set or the whole set. Since

(1) $\bigcup_k C_k \in \mathscr{C}$,

(2) $\bigcap_{k=1}^{n} C_k = \bigcap_{k=1}^{n} \bigcup_{j \in J_k} B_j^k = \bigcup_{\substack{j_1 \in J_1 \\ j_2 \in J_2 \\ \cdots\cdots \\ j_n \in J_n}} (B_{j_1}^1 \cap B_{j_2}^2 \cap \cdots \cap B_{j_n}^n) \in \mathscr{C}$,

(3) $\emptyset \in \mathscr{C}, \; X \in \mathscr{C}$,

the family $\mathscr{C}$ determines a topology, which is called the **topology generated by the family** $\mathscr{A}$.

(1) The idea of a topological space is introduced in different places in the literature with some slight variations in the axioms. The axioms used here are those used by P. Alexandroff and H. Hopf (*Topologie*, 1). We have followed N. Bourbaki (*Topologie*) for terminology.

DEFINITION 2. A set F is said to be **closed** if its complement $-F$ is open. From the axioms for open sets it follows immediately that the family $\mathscr{F} = (F_i \mid i \in I)$ of closed sets satisfies

$$(1') \quad J \subset I \quad \Rightarrow \quad \bigcap_{i \in J} F_i \in \mathscr{F},$$

$$(2') \quad J = \{i_1, i_2, \ldots, i_n\} \subset I \quad \Rightarrow \quad \bigcup_{i=i_1}^{i_n} F_i \in \mathscr{F},$$

$$(3') \quad \emptyset \in \mathscr{F}, \; X \in \mathscr{F}.$$

REMARK. (1′), (2′) and (3′) are sufficient to characterise a topology. In fact, every family $\mathscr{F}$ satisfying these properties determines a topology $\mathscr{G}$, for which $\mathscr{F}$ is the family of closed sets.

DEFINITION 3. Given any subset A of a topological space X, the intersection $\bar{A}$ of all the closed sets containing A is called the **topological closure** of A. From (1′) above, it follows that $\bar{A}$ is a closed set.

The following properties of topological closure are easily verified:

$$\begin{aligned}
&(1) \quad \bar{A} \supset A,\\
&(2) \quad \overline{A \cup B} = \bar{A} \cup \bar{B},\\
&(3) \quad \overline{(\bar{A})} = \bar{A},\\
&(4) \quad \bar{\emptyset} = \emptyset.
\end{aligned}$$

Properties (1), (3) and (4) are immediate. To prove (2), we observe that $\overline{A \cup B}$ is a closed set containing A and B and so it contains $\bar{A}$ and $\bar{B}$, whence $\overline{A \cup B} \supset \bar{A} \cup \bar{B}$. Similarly $\bar{A} \cup \bar{B}$ is a closed set which contains A and B, whence $\bar{A} \cup \bar{B} \supset \overline{A \cup B}$.

REMARK 1. The correspondence $A \to \bar{A}$ is a closure operation of the kind defined above (§ 7, Chapter I), for, in addition to the above properties, we have

$$A \supset B \quad \Rightarrow \quad A = A \cup B \quad \Rightarrow \quad \bar{A} = \overline{A \cup B} = \bar{A} \cup \bar{B} \quad \Rightarrow \quad \bar{A} \supset \bar{B}.$$

We observe that the lattice of closed sets (§ 8, Chapter I), in which

$$\begin{aligned}
F_1 \wedge F_2 &= F_1 \cap F_2,\\
F_1 \vee F_2 &= \overline{F_1 \cup F_2} = F_1 \cup F_2,
\end{aligned}$$

is distributive in this case, since

$$F_1 \cap (F_2 \cup F_3) = (F_1 \cap F_2) \cup (F_1 \cap F_3).$$

REMARK 2. *A set A is closed if and only if $A = \bar{A}$*; this follows from the definition of closure.

REMARK 3. Properties (1), (2), (3) and (4) of a topological closure operation are sufficient to characterise a topology: given a closure operation $A \to \bar{A}$ which is additive (that is, one which satisfies (2)), and defining a set A to be closed if and only if $\bar{A} = A$, we can define a topology $\mathscr{G}$ for which the topological closure operation is $A \to \bar{A}$.

DEFINITION 4. Given a set A, the union $\mathring{A}$ of all the open sets contained in A is called the **interior** of A; it is easily seen that

$$\mathring{A} = -(\overline{-A}).$$

A point x in $\mathring{A}$ is called an **interior point** of the set A.

DEFINITION 5. An open set containing a point x of X is called an **open neighbourhood** of x; such a set will be denoted by $U(x)$ or $V(x)$ and the collective family of open neighbourhoods of x will be denoted by $\mathscr{V}(x)$. Clearly $\mathscr{V}(x)$ is a filter base. Any set containing a set $V(x)$ is called a **neighbourhood** of x; such a set will be denoted by $N(x)$ and the collective family of neighbourhoods of x will be denoted by $\mathscr{N}(x)$.

A family of sets $\mathscr{B}(x) = (N_i \,/\, i \in I)$ is called a **fundamental base** of neighbourhoods of x if it is equivalent in the exterior sense (cf. § 3, Chapter I) to $\mathscr{V}(x)$. If $\mathscr{B}(x) \simeq \mathscr{V}(x)$, then every set N_i contains a $U(x)$ and every $U(x)$ contains an N_i; therefore every N_i is a neighbourhood of x. A family $(\mathscr{B}(x) \,/\, x \in X)$ of fundamental bases satisfies the following properties:

(1) $x \in X,\ N \in \mathscr{B}(x) \quad \Rightarrow \quad x \in N,$
(2) $\mathscr{B}(x)$ is a filter base,
(3) for each $N_0 \in \mathscr{B}(x)$, there exists an $N \in \mathscr{B}(x)$ such that

$$(\forall_N y)\,(\exists M(y);\ M(y) \in \mathscr{B}(y)) : M(y) \subset N_0.$$

Property (1) is immediate, for if $\mathscr{V}(x) \vdash \mathscr{B}(x)$, then in each N_i there exists an open neighbourhood of x, which contains x. Property (2) can be deduced from the fact that a family equivalent to a filter base is itself a filter base (Theorem 1, § 6, Chapter I). To prove property (3), we observe that if $N_0 \in \mathscr{B}(x)$, then there exists a $V(x)$ contained in N_0; then, choosing N to be a set of $\mathscr{B}(x)$ contained in $V(x)$, we have

$$y \in N \quad \Rightarrow \quad V(x) \in \mathscr{V}(y) \quad \Rightarrow \quad (\exists M(y);\ M(y) \in \mathscr{B}(y)) : M(y) \subset V(x) \subset N_0.$$

REMARK 1. *A point $x \in \bar{A}$ if and only if every neighbourhood of x meets A.*

Proof. Suppose that $x \in \bar{A}$ and that there is a set $V(x)$ such that $V(x) \cap A = \varnothing$; the closed set $-V(x)$ then contains A and so contains $\bar{A}$. Since $x \in \bar{A}$ and $x \notin -V(x)$, we have a contradiction.

Conversely, suppose that each set $V(x)$ meets A. Since the open set $-(\bar{A})$ does not meet A it cannot be an open neighbourhood of x and therefore $x \in \bar{A}$.

REMARK 2. *A set is open if and only if it is a neighbourhood of each of its points.*

Proof. If G is open, then by definition it is an open neighbourhood of each of its points. Conversely, if a set A is a neighbourhood of each of its points, then given $a \in A$ there exists a set $V(a) \subset A$ and so

$$A = \bigcup_{a \in A} \{a\} \subset \bigcup_{a \in A} V(a) \subset A.$$

Therefore $A = \bigcup_{a \in A} V(a)$ and so A is open.

REMARK 3. It can be shown that the properties (1), (2) and (3) of a family of bases $(\mathscr{B}(x) \,/\, x \in X)$ are sufficient to characterise a topology: given a family $(\mathscr{B}(x) \,/\, x \in X)$ satisfying (1), (2) and (3) and using the property mentioned in Remark 2, we can define a topological structure $\mathscr{G}$ which admits $(\mathscr{B}(x) \,/\, x \in X)$ as the family of fundamental bases.

EXAMPLE. Let X be a metric space; for fundamental base at x we can take the family

$$\mathscr{B}(x) = (B_1(x),\ B_{1/2}(x),\ B_{1/3}(x), \dots)$$

(where $B_\lambda(x)$ is the ball of centre x and radius λ).

DEFINITION. Let σ be a single-valued mapping of a topological space X into a topological space Y; then σ is said to be **continuous at a point** x_0 of X if, for each neighbourhood $V(\sigma x_0) = V(y_0)$ in Y there exists a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad \sigma x \in V(y_0):$$

that is,

$$(\forall V(y_0))\, (\exists U(x_0)) : \sigma U(x_0) \subset V(y_0).$$

A mapping σ which is continuous at each point of X is said to be a **continuous mapping** (on X). (In the special cases when X is a metric space or an L^0-space, we have the type of continuity already discussed.)

A numerical function f defined on a topological space X is said to be **continuous** if it is a continuous mapping of X into $\mathbf{R}$. In this case the following property is satisfied: given any point $x_0 \in X$ and any strictly positive number ε, there exists a neighbourhood $V(x_0)$ such that

$$x \in V(x_0) \quad \Rightarrow \quad |f(x) - f(x_0)| \leqq \varepsilon.$$

A one-one correspondence σ between two spaces X and Y is called a **homeomorphism** if σ and σ^{-1} are continuous mappings. If two spaces X and Y are such that there exists a homeomorphism between them, then we say that X and Y are **homeomorphic**, or **topologically equivalent**; we can verify that the relation of being homeomorphic is an equivalence relation.

EXAMPLE. In $\mathbf{R}^3$, it is sometimes easy to see intuitively if two surfaces A and B are homeomorphic; they will be so related if we can obtain B by deforming A, without tears or joins, as if it were a perfectly elastic rubber sheet. The sphere is homeomorphic to the cube but not to the torus; from figure 17 we can see that the cylinder and the Möbius band (which can both be obtained by sticking together the ends of a strip of paper $abcd$) belong to two different topological equivalence classes.

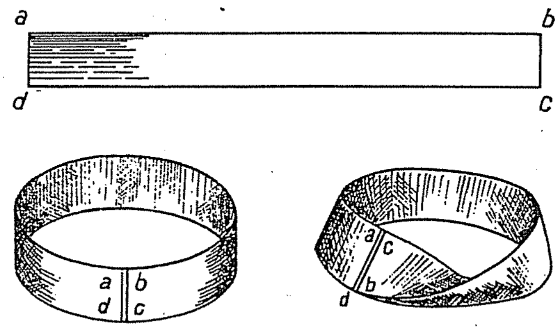


FIG. 17

Theorem. *A single-valued mapping σ of a topological space X into a topological space Y is continuous if and only if, for each open set $G \subset Y$, the set $\sigma^{-1}G$ is open; likewise σ is continuous if and only if for each closed set $F \subset Y$ the set $\sigma^{-1}F$ is closed.*

Proof. Suppose that σ is a mapping of X into Y such that $\sigma^{-1}G$ is open for each open set $G \subset Y$. Given any neighbourhood $V(\sigma x_0) = V(y_0)$, then $\sigma^{-1}V(y_0)$ is open; therefore it is a set of the form $U(x_0)$ and we have

$$\sigma U(x_0) \subset V(y_0).$$

Hence σ is continuous at x_0.

Conversely, suppose that σ is continuous at each point x of X. Let G be an open set of Y and let x_0 be an element of $\sigma^{-1}G$. Since $y_0 = \sigma x_0 \in G$, there exists a set $U(x_0)$ such that

$$\sigma U(x_0) \subset G, \quad \text{i.e.} \quad U(x_0) \subset \sigma^{-1}G.$$

Therefore x_0 is an interior point of the set $\sigma^{-1}G$ and so $\sigma^{-1}G$ is open.

This proves the first part of the theorem; the second part can easily be proved by taking complements.

COROLLARY 1. *If a continuous mapping σ maps a subset A of X into a subset B of Y, then $\sigma\bar{A} \subset \bar{B}$.*

Proof. The set $F = \sigma^{-1}\bar{B}$ is closed and contains A; therefore $F \supset \bar{A}$ and

$$\sigma\bar{A} \subset \sigma F = \bar{B}.$$

COROLLARY 2. *A one-one correspondence σ between two topological spaces X and Y is a homeomorphism if and only if it induces (in the obvious way) a one-one correspondence between the open sets of X and the open sets of Y; likewise σ is a homeomorphism if and only if it induces a one-one correspondence between the closed sets of X and the closed sets of Y.*

This is an immediate consequence of the theorem.

DEFINITION. Let A be a subset of a topological space X; we say that an element x of X is a **point of closure** of A if every $V(x)$ meets A: that is, if $x \in \bar{A}$. We say that an element $x \in X$ is a **point of accumulation** of the set A if every $V(x)$ meets the set $A-\{x\}$: that is, if $x \in \overline{A-\{x\}}$. The set A' of points of accumulation of A is called the **derived set** of A.

We observe that $\bar{A} = A \cup A'$, for if $x \in A$ or $x \in A'$, then $x \in \bar{A}$ and conversely if $x \in \bar{A}$ and $x \notin A$, then

$$x \in \bar{A} = \overline{A-\{x\}},$$

whence $x \in A'$.

Given a set A, the **frontier** of A is the set Fr A of points of closure of A which are not interior points of A; we have

$$\bar{A} = \mathring{A} \cup \mathrm{Fr}\, A; \quad \mathrm{Fr}\, A \cap \mathring{A} = \emptyset.$$

We observe that Fr A is a closed set, for

$$\mathrm{Fr}\, A = \bar{A} \cap (-\mathring{A}) = \bar{A} \cap \overline{(-A)}.$$

EXAMPLE. In a metric space, the frontier of the ball $B_\lambda(a)$ is the sphere $S_\lambda(a) = \{x \,/\, d(x, a) = \lambda\}$.

§ 4. Sequences and filtered families

Let $(x_n) = (x_1, x_2, x_3, \dots)$ be a sequence of elements of X. Then we say that an element $a \in X$ is a **limit point** of (x_n) if

$$(\forall U(a))\,(\exists n_0)\,(\forall k;\ k \geqq n_0) : x_k \in U(a).$$

In other words, any neighbourhood of a contains all the x_n from a certain value of n onwards. If a is a limit point of (x_n), we say that (x_n) **converges** to a and we write $(x_n) \to a$.

Given a sequence (x_n), we say that a point a is a **cluster point** of (x_n) if

$$(\forall U(a))\,(\forall n)\,(\exists k;\ k \geqq n) : x_k \in U(a).$$

In other words, any neighbourhood of a contains members x_k of the sequence with indices as large as we please. It is an immediate consequence of the definitions that a limit point is always a cluster point.

EXAMPLE. Let $X = \mathbf{R}$; then if $x_n = 1/n$, the sequence (x_n) has 0 as a limit point; if $x_n = (-1)^n(1-1/n)$, the sequence (x_n) has two cluster points $+1$ and -1, but has no limit point.

These ideas can be extended to any family $(x_i \mid i \in I)$. To do this[1] we associate with the family a filter base $\mathscr{B}$ on I; we then obtain a **filtered family** $(x_i) = (x(i) \mid i \in I, \mathscr{B})$. Unless otherwise stated, a sequence (x_n) will always be considered as a filtered family with the Fréchet base $\mathscr{S} = (S_k \mid k \in \mathbf{N})$ as the filter base on the index set $\mathbf{N}$; here S_k is the section $\{n \mid n \in \mathbf{N},\ n \geqq k\}$. The above definition of limit point of a sequence (x_n) can be stated as follows: a is a limit point of (x_n) if

$$(\forall U(a))\,(\exists S;\ S \in \mathscr{S}) : x(S) \doteq \bigcup_{i \in S} \{x(i)\} \subset U(a).$$

Likewise a cluster point of a sequence (x_n) can be defined to be a point a such that

$$(\forall U(a))\,(\forall S;\ S \in \mathscr{S}) : x(S) \cap U(a) \neq \varnothing.$$

The purpose of restating the definitions in these forms is to make them suitable for generalisation to any filtered family. Thus we say that an element a is a **limit point** of the filtered family $(x(i) \mid i \in I, \mathscr{B})$ if

$$(\forall U(a))\,(\exists B;\ B \in \mathscr{B}) : x(B) \subset U(a).$$

In other words, the filter base $x(\mathscr{B})$ is partially contained in every neighbourhood of a: thus $x(\mathscr{B}) \vdash \mathscr{V}(a)$.

As in the particular case of sequences, we say that (x_i) **converges** to a if a is a limit point of the filtered family (x_i) and this is written $(x_i) \to a$.

An element a is called a **cluster point** of the filtered family $(x(i) \mid i \in I, \mathscr{B})$ if

$$(\forall U(a))\,(\forall B,\ B \in \mathscr{B}) : x(B) \cap U(a) \neq \varnothing.$$

In other words, given any neighbourhood $U(a)$, the filter base $x(\mathscr{B})$ is not partially contained in $-U(a)$.

Let $(x_i) = (x(i) \mid i \in I, \mathscr{B})$ and $(y_i) = (y(i) \mid i \in J, \mathscr{B}')$ be two filtered families. If $x(\mathscr{B}) \vdash y(\mathscr{B}')$, then we say that (x_i) is a sub-family of (y_i) and we write $(x_i) \vdash (y_i)$. It is easily seen that a sub-sequence (x_{k_n}) of a sequence (x_n) is also a filtered sub-family of (x_n).

If $x(\mathscr{B})$ is an ultra-filter base, then we say that (x_i) is an **ultra-filtered family.**

We now introduce a particular type of filtered family which is of some importance. Let $(x_i \mid i \in I)$ be a family of elements of X and let $\geqq$ be a latticial ordering relation (cf. § 4, Chapter III) on the index set I. Then the sections $S_k = \{i \mid i \in I,\ i \geqq k\}$ form a filter base, since

(1) $S_k \neq \varnothing$ (because $k \in S_k$),
(2) $(\forall i)\,(\forall j)\,(\exists k;\ k = \sup\{i, j\}) : S_k \subset S_i \cap S_j$,
(because $l \geqq \sup\{i, j\} \quad \Rightarrow \quad l \geqq i, l \geqq j$).

[1] (Translator's note.) An alternative development is through the concept of directed set. For an account of this, see J. L. Kelley (*General Topology*).

The pair formed by the family $(x_i \,/\, i \in I)$ and the filter base determined by the sections S_k is called a **Moore-Smith family.** By definition it is a filtered family; we denote it by $(x_i \,/\, i \in I, \geqq)$. For a Moore-Smith family, the definitions of limit point and cluster point can be stated more simply as follows. A point $a \in X$ is a limit point if

$$(\forall V(a))\,(\exists_I j) \,:\, x_i \in V(a) \text{ for all } i \geqq j;$$

a point $b \in X$ is a cluster point if

$$(\forall V(b))\,(\forall_I j) \,:\, x_i \in V(b) \text{ for at least one } i \geqq j.$$

EXAMPLE. Let $u_1, u_2, u_3, \ldots$ be elements of **R** and write

$$x_n = u_1 + u_2 + u_3 + \ldots + u_n.$$

Then we say that the series Σu_n is **convergent** if the sequence (x_n) converges to a point x_0; in other words if, given $\varepsilon > 0$, there exists an integer m such that

$$n \geqq m \quad \Rightarrow \quad |x_n - x_0| \leqq \varepsilon.$$

Let $K = \{n_1, n_2, \ldots n_j\}$ be a finite subset of **N**. Then we write

$$x_K = u_{n_1} + u_{n_2} + \ldots + u_{n_j}.$$

Let $\mathscr{K}$ be the collection of finite sets $K \subset \mathbf{N}$. The inclusion relation $\supset$ is a latticial ordering relation on $\mathscr{K}$, since

$$\sup\{K_1, K_2\} = K_1 \cup K_2 \in \mathscr{K}.$$

We say that the series Σu_n is **summable** if the Moore-Smith family $(x_K \,/\, K \in \mathscr{K}, \supset)$ converges to a point x_0: that is, to each $\varepsilon > 0$ there corresponds a finite set $K_0 \subset \mathbf{N}$ such that

$$K\,(\text{finite}) \supset K_0 \quad \Rightarrow \quad |x_K - x_0| \leqq \varepsilon.$$

It can be shown that *the series Σu_n is convergent independently of the order in which the terms u_n are taken if and only if it is summable.* This result is very general and is still valid if **R** is replaced by a 'topological vector space' (cf. Chapter IX).

In the theorems which follow, we can use indifferently the idea of 'filtered family' or that of 'Moore-Smith family'.

Theorem 1. *If $(y_j) \vdash (x_i)$, every limit point of (x_i) is a limit point of (y_j) and every cluster point of (y_j) is a cluster point of (x_i).*

Proof. If $(y_j) \vdash (x_i)$, then $y(\mathscr{B}') \vdash x(\mathscr{B})$; if a is a limit point of (x_i), then there is a set $B \in \mathscr{B}$ such that $x(B) \subset U(a)$ and so there is a set $B' \in \mathscr{B}'$ such that $y(B') \subset x(B) \subset U(a)$. Hence a is a limit point of (y_j).

Suppose now that b is a cluster point of (y_j). Then, for each neighbourhood $U(x)$ and for each $B' \in \mathscr{B}'$, we have

$$y(B') \cap U(b) \neq \varnothing.$$

Let B be any set in $\mathscr{B}$. Then there exists $B' \in \mathscr{B}'$ such that $y(B') \subset x(B)$ and therefore $x(B) \cap U(a) \neq \varnothing$. Hence b is a cluster point of (x_i).

Theorem 2. *A point a belongs to $\bar{A}$ if and only if there exists a filtered family $(x_i) = \big(x(i) \,/\, i \in I, \mathscr{B}\big)$ contained in A and having a as a limit point.*

Proof. Suppose that $a \in \bar{A}$. Let V be any set in $\mathscr{V}(a)$; then, by the axiom of choice, there exists an element $x(V)$ in $V \cap A$. Since $\subset$ is a latticial ordering in $\mathscr{V}(a)$, $(x_V) = \big(x(V) \,/\, V \in \mathscr{V}(a);\, \subset\big)$ is a Moore-Smith family. For any neighbourhood $U(a)$, we have

$$V \subset U(a) \quad\Rightarrow\quad x(V) \in V \subset U(a)$$

and so a is a limit point of (x_V).

Conversely suppose that a is a limit point of a filtered family

$$(x_i) = \big(x(i) \,/\, i \in I;\, \mathscr{B}\big).$$

Then, for every neighbourhood $U(a)$, we have

$$(\forall B)\,(\exists_B i) : x_i \in U(a)$$

and so $A \cap U(a) \neq \varnothing$, whence a is a point of closure of A.

COROLLARY 1. *A point a is a point of accumulation of A if and only if there exists a filtered family contained in $A - \{a\}$ and having a as a limit point.*

This result is an immediate consequence of Theorem 2 and the definition of point of accumulation.

COROLLARY 2. *A set F is closed if and only if for each limit point a of a filtered family of F we have $a \in F$.*

This result follows from Theorem 2 and the fact that a set F is closed if and only if $F = \bar{F}$.

Theorem 3. *A point a is a cluster point of (x_i) if and only if there exists a filtered family (y_j) admitting a as a limit point and satisfying $(y_j) \vdash (x_i)$.*

Proof. Suppose that $(y_j) = \big(y(j) \,/\, j \in J, \mathscr{B}'\big)$ admits a as a limit point and is such that $y(\mathscr{B}') \vdash x(\mathscr{B})$. Then, if a is not a cluster point of (x_i), the filter base $x(\mathscr{B})$ is partially contained in a set $-U(a)$ and so there exist sets $B \in \mathscr{B}$ and $B' \in \mathscr{B}'$ such that

$$y(B') \subset x(B) \subset -U(a),$$

which implies that (y_j) does not admit a as a limit point, contrary to hypothesis. Therefore a is a cluster point of (x_i).

Suppose conversely that a is a cluster point of (x_i). Then

$$(\forall B)\,\big(\forall V \in \mathscr{V}(a)\big)\,\big(\exists x(B, V)\big) : x(B, V) \in x(B) \cap V$$

and $(x_{B,V}) = \big(x(B, V) \,/\, (B, V) \in \mathscr{B} \times \mathscr{V}(a);\, \subset\big)$ is a Moore-Smith family.

For any set $x(B_0) \in x(\mathscr{B})$, we have

$$B \times V \subset B_0 \times V_0 \quad \Rightarrow \quad x(B, V) \in x(B_0)$$

and so $(x_{B,V}) \vdash (x_i)$. For any set $U(a) \in \mathscr{V}(a)$, we have

$$B \times V \subset B_0 \times U(a) \quad \Rightarrow \quad x(B, V) \in V \subset U(a)$$

and so a is a limit point of $(x_{B,V})$.

COROLLARY. *Every limit point is a cluster point.*

This follows at once from the fact that $(x_i) \vdash (x_i)$.

REMARK. *Every cluster point of an ultra-filtered family (x_i) is a limit point.*

Proof. If $x(\mathscr{B})$ is an ultra-filter base, then either $x(\mathscr{B}) \vdash V(a)$ or $x(\mathscr{B}) \vdash [-V(a)]$. If a is a cluster point the second case is impossible and so a is a limit point of (x_i).

Theorem 4. *A necessary and sufficient condition for a single-valued mapping σ of a topological space X into a topological space Y to be continuous at x_0 is that*

$$(x_i) \to x_0 \Rightarrow (\sigma x_i) \to \sigma x_0.$$

Proof. If σ is continuous at x_0, then

$$(\forall V(y_0))\,(\exists U(x_0)) : \sigma U(x_0) \subset V(y_0),$$

where $y_0 = \sigma x_0$. If $(x_i) \to x_0$ there exists a set $B \in \mathscr{B}$ such that $x(B) \subset U(x_0)$; we have

$$\sigma x(B) \subset \sigma U(x_0) \subset V(y_0)$$

and therefore $(\sigma x_i) \to y_0$.

Conversely, suppose that σ is not continuous at x_0. Then

$$(\exists V(y_0))\,(\forall U(x_0))\,(\exists x_U \in U(x_0)) : \sigma x_U \notin V(y_0).$$

The filtered family $(x_U \,/\, U \in \mathscr{V}(x_0);\ \subset)$ has x_0 as a limit point but (σx_U) does not have $y_0 = \sigma x_0$ as a limit point.

Theorem 5. *Convergence of filtered families in a topological space satisfies the following conditions:*

(1) $(x_i) \to x_0,\ (y_j) \vdash (x_i) \quad \Rightarrow \quad (y_j) \to x_0,$
(2) $(\forall i) : x_i = x_0 \quad \Rightarrow \quad (x_i) \to x_0,$
(3) $(x_i) \to x_0$ *whenever* $(\forall (y_j);\ (y_j) \vdash (x_i))\,(\exists (z_k);\ (z_k) \vdash (y_i)) : (z_k) \to x_0,$
(4) *if* $(x_i) \to x_0$ *and, for each* $i \in I$, *we have* $(x_k^i \,/\, k \in K_i, \mathscr{B}_i) \to x_i$, *then the set* $D = \{x_k^i \,/\, k \in K_i,\ i \in I\}$ *contains a filtered family* (y_j) *having* x_0 *as a limit point.*

Proof. The proofs of properties (1) and (2) are immediate. To prove (3), suppose that x_0 is not a limit point of (x_i). Then

$$(\exists V(x_0))\,(\forall B) : x(B) \not\subset V(x_0).$$

If $I_0 = \{i \mid x(i) \notin V(x_0)\}$ then the sets $B \cap I_0$ form a filter base $\mathscr{B}_0$ (by the Corollary to Theorem 3, § 6, Chapter I). The family $(x'_i) = (x(i) \mid i \in I_0, \mathscr{B}_0)$ is a sub-family of (x_i) and, since $x(B \cap I_0) \subset -V(x_0)$, it does not admit x_0 as a cluster point. Thus we have a contradiction and so x_0 is a limit point of (x_i).

To prove property (4) it is sufficient to prove that x_0 is a point of closure of the set D; the result then follows from Theorem 2. Any neighbourhood $U(x_0)$ contains an element x_i; hence, since $U(x_0)$ is a neighbourhood of x_i, it contains an element x_k^i and so $U(x_0) \cap D \neq \emptyset$. Therefore x_0 is a point of closure of D.

REMARK. Properties (1), (2), (3) and (4) of a filtered family are sufficient to characterise a topology. Starting with any convergence $(x_i) \to x_0$ satisfying these properties and using Corollary 2 of Theorem 2, we can define a topology for which x_0 is a limit point of a filtered family (x_i) if and only if $(x_i) \to x_0$.

§ 5. Separated, quasi-separated, regular and normal spaces

We now study certain important properties which characterise various types of topological spaces. Let X be a topological space with $\mathscr{G} = (G_i \mid i \in I)$ as the topology. Let A be a subset of X. For the family $\mathscr{G}_A$ defined by

$$\mathscr{G}_A = (G_i \cap A \mid i \in I)$$

we have

$$\bigcup_{i \in J} (G_i \cap A) = A \cap \bigcup_{i \in J} G_i \in \mathscr{G}_A,$$

$$\bigcap_{i=i_1}^{i_n} (G_i \cap A) = A \bigcap_{i=i_1}^{i_n} G_i \in \mathscr{G}_A,$$

$$A = A \cap X \in \mathscr{G}_A, \qquad \emptyset = A \cap \emptyset \in \mathscr{G}_A.$$

Therefore $\mathscr{G}_A$ determines a topology on A; we call $(A, \mathscr{G}_A)$ a **topological subspace** of $(X, \mathscr{G})$.

If the topological space $(A, \mathscr{G}_A)$ satisfies a property (L), we say that, in X, A **satisfies the property** (L). Given a point $x \in X$, we say that an **arbitrarily small neighbourhood** (or a neighbourhood as small as we please) of x satisfies (L) if each $U(x)$ contains an $N(x)$ satisfying (L). We also say that X **satisfies** (L) **locally** if, for each $x \in X$, an arbitrarily small neighbourhood of x satisfies (L).

A topological space X is called a **separated** space, or a **Hausdorff** space, if, given any two distinct points x_1 and x_2, there exist disjoint neighbourhoods

$U(x_1)$ and $U(x_2)$. A topological space X is said to be **quasi-separated** if, given any two distinct points x_1 and x_2, there exists a neighbourhood $U(x_1)$ not containing x_2. Clearly a separated space is also quasi-separated.

EXAMPLE 1. A metric space is separated, for if $x_1 \neq x_2$, we have $d(x_1, x_2) = \varepsilon > 0$, and so the neighbourhoods

$$U(x_1) = B_{\varepsilon/2}(x_1) - S_{\varepsilon/2}(x_1), \quad U(x_2) = B_{\varepsilon/2}(x_2) - S_{\varepsilon/2}(x_2)$$

are disjoint.

EXAMPLE 2. A space X having the coarsest topology $\mathscr{G} = (X, \varnothing)$ is not separated unless it has only one point or is empty. On the other hand, a space with the discrete topology $\mathscr{G} = \mathscr{P}(X)$ is always separated.

EXAMPLE 3. (Alexandroff-Urysohn). Let X be the real line $\mathbf{R}$. Given any $\varepsilon > 0$, we write

$$N_\varepsilon(x) = [x-\varepsilon, \; x+\varepsilon]$$

if $x \leqq 0$ and

$$N_\varepsilon(x) = [x-\varepsilon, \; x+\varepsilon] \cup [-x-\varepsilon, \; -x+\varepsilon] - \{-x\}$$

if $x > 0$. Then the sets $N_\varepsilon(x)$ satisfy the axioms for fundamental bases and so determine a topology. Since

$$N_\varepsilon(x) \cap N_{\varepsilon'}(-x) \neq \varnothing$$

the space is not separated; however it is easily shown that it is quasi-separated.

Theorem 1. *A topological space X is quasi-separated if and only if every set $\{x\}$ consisting of a single element is closed.*

Proof. Suppose that the sets consisting of single elements are closed. Then, if $x \neq y$, we have $x \in -\{y\} = U(x)$ and $y \notin U(x)$; therefore X is quasi-separated.

Conversely, suppose that X is quasi-separated. Let x_0 be a fixed point of X and let x be any point of X such that $x_0 \neq x$. Then there exists a neighbourhood $U(x)$ which does not contain x_0. The union of the $U(x)$ for all $x \in X$ such that $x_0 \neq x$ is an open set and it is equal to the complement of the set $\{x_0\}$. Therefore the set $\{x_0\}$ is closed.

Theorem 2. *A space X is separated if and only if every filtered family (x_i) has at most one limit point.*

Proof. Suppose that X is a space which is not separated. Then there exist two points x_1 and x_2 such that, for each neighbourhood U of x_1 and each neighbourhood V of x_2, we have $U \cap V \neq \varnothing$. The filtered family defined by

$$(x_{U,V}) = (x(U, V) \,/\, (U, V) \in \mathscr{V}(x_1) \times \mathscr{V}(x_2); \; \subset),$$

where $x(U, V)$ denotes an element of $U \cap V$, has x_1 as a limit point, since for any neighbourhood $U(x_1)$, we have

$$U \times V \subset U(x_1) \times V_0 \quad \Rightarrow \quad x(U, V) \in U \subset U(x_1).$$

Similarly x_2 is a limit point of the family and so the family has at least two limit points. Therefore if every filtered family of a space X has at most one limit point, the space X is separated.

Conversely, suppose that X is a separated space. Let $(x_i) = (x(i) \mid i \in I, \mathscr{B})$ be a filtered family in X. Suppose that x_1 and x_2 are two limit points of the family such that $x_1 \neq x_2$. Then there exist disjoint neighbourhoods $U(x_1)$ and $U(x_2)$ and therefore there exist sets B and B' in $\mathscr{B}$ such that

$$x(B) \subset U(x_1),$$
$$x(B') \subset U(x_2).$$

Then $x(B) \cap x(B') \subset U(x_1) \cap U(x_2) = \varnothing$, which leads to a contradiction, for $x(\mathscr{B})$ is a filter base and so $x(B) \cap x(B') \neq \varnothing$.

In what follows all the topological spaces which we consider will be assumed to be separated.

An **open neighbourhood** of a set A is any open set containing A and will be denoted by $U(A)$ or $V(A)$, just as an open neighbourhood of a point x is denoted by $U(x)$ or $V(x)$. A topological space X is said to be **regular** if it is separated and if, for any point x and any closed set F not containing x, there exist disjoint neighbourhoods $U(x)$ and $U(F)$. A topological space X is said to be **normal** if it is separated and if, for any two disjoint closed sets F and F' there exist disjoint neighbourhoods $U(F)$ and $U(F')$. Clearly a normal space is regular.

EXAMPLE. A space with the discrete topology is normal.

Theorem 3. *A separated space X is regular if and only if it is locally closed: that is, every neighbourhood of x contains a closed neighbourhood of x.*

Proof. Suppose that X is regular and let $U(x)$ be an open neighbourhood of x. Since $F = -U(x)$ is closed, there exist disjoint neighbourhoods $V(x)$ and $V(F)$. Then $V(x) \subset -V(F)$ and so

$$\overline{V(x)} \subset -V(F) \subset U(x).$$

Therefore $U(x)$ contains a closed neighbourhood $\overline{V}(x)$.

Conversely, suppose that for each $x \in X$ and each neighbourhood $U(x)$ there exists a closed neighbourhood of x contained in $U(x)$. Let F be a closed set not containing x. Then the open set $U(x) = -F$ contains x and so contains a closed neighbourhood $N(x)$. Writing $V(F) = -N(x)$, we have $V(F) \cap \mathring{N}(x) = \varnothing$. Therefore X is regular.

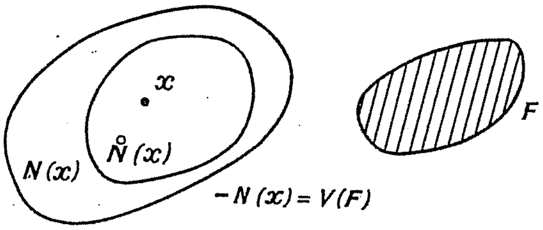


FIG. 18

Theorem 4. *A subspace X of a separated (resp. quasi-separated, regular, normal) space is a separated (resp. quasi-separated, regular, normal) space.*

Proof. We prove the result in the case in which X is separated; the other cases can be treated in a similar manner.

If $A \subset X$ and $a, b \in A$, then there exist disjoint neighbourhoods $U(a)$ and $U(b)$; therefore

$$[U(a) \cap A] \cap [U(b) \cap A] = \varnothing$$

and so the space $(A, \mathscr{G}_A)$ is separated.

§ 6. Compact sets

Again all the topological spaces which we are considering are assumed to be separated. A topological space X is said to be **compact** if it is separated and if the following axiom is satisfied:

(1) (Borel-Lebesgue axiom). *Every family of open sets $(G_i \mid i \in I)$ forming a covering of X, contains a finite covering:*

$$(G_{i_1}, G_{i_2}, \ldots, G_{i_k}).$$

A **compact set** K is a subset of a topological space X such that $(K, \mathscr{G}_K)$ is compact; in other words, every family of open sets whose union contains K has a finite sub-family whose union contains K.

EXAMPLE 1. Let X be the set $\mathbf{R}$, with the usual topology. Then the set $K = [0, 1]$ is compact. To prove this we suppose that $\mathscr{A} = (G_i \mid i \in I)$ is an open covering of $[0, 1]$ and that it contains no finite sub-covering. Divide K into two segments $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$; at least one of these segments cannot be covered by a finite sub-family of $\mathscr{A}$; let $[a_1, b_1]$ denote this segment. Sub-divide $[a_1, b_1]$ into two equal segments $[a_1, \frac{1}{2}(a_1+b_1)]$ and $[\frac{1}{2}(a_1+b_1), b_1]$; at least one of these segments cannot be covered by a finite sub-family of $\mathscr{A}$; let $[a_2, b_2]$ denote this segment. Continuing this process, we obtain a sequence of segments $[a_n, b_n]$. The sequence $(a_1, a_2, a_3, \ldots)$ is increasing and is bounded above by 1; the sequence $(b_1, b_2, b_3, \ldots)$ is decreasing and is bounded below by 0. Thus (a_n) tends to a limit a_0 and

(b_n) tends to a limit b_0; but $|b_n - a_n|$ tends to zero and so $a_0 = b_0$. Let G_k be an open set of $\mathscr{A}$ which contains a_0 (since $\mathscr{A}$ is an open covering, such a set exists). For n sufficiently large, the interval $[a_n, b_n]$ is contained in G_k, which contradicts the definition of $[a_n, b_n]$. Therefore the hypothesis that there is no finite sub-covering is false.

Example 2. The straight line $\mathbf{R}$, with the usual topology, is a locally compact space; in fact, from the above, each point x of $\mathbf{R}$ possesses a fundamental base of compact neighbourhoods, consisting of the sets $N_\varepsilon(x) = [x-\varepsilon, x+\varepsilon]$.

Example 3. Let $\hat{\mathbf{R}}$ be the augmented real line (that is, the line $\mathbf{R}$ together with the points $+\infty$ and $-\infty$). A topology in $\hat{\mathbf{R}}$ is generated (see Example 3 of § 3) by the following sets:

(1) the open sets in $\mathbf{R}$,

(2) the union of $\{+\infty\}$ with an open set of $\mathbf{R}$ containing an interval $]\lambda, +\infty[$,

(3) the union of $\{-\infty\}$ with an open set of $\mathbf{R}$ containing an interval $]-\infty, \lambda[$.

By means of an argument similar to that given in example 1, we can show that $\hat{\mathbf{R}}$, together with the topology just defined, is a compact space.

Example 4. In $\mathbf{R}$ a set A which is not bounded above is not compact, for a covering of A consisting of open intervals of length unity does not contain a finite sub-covering.

Example 5. Any finite set $\{x_1, x_2, \ldots, x_k\}$ in a space X is compact, for an open covering of this set contains a finite open covering consisting of at most k sets.

Bolzano-Weierstrass theorem. *If X is a compact space, every infinite subset A of X possesses a point of accumulation.*

Proof. Suppose that A is a set which possesses no points of accumulation. Then, if x is any point of X, we have $x \notin A'$ and so there exists a neighbourhood $U(x)$ disjoint from $A-\{x\}$. The family $(U(x) \,/\, x \in X)$ is an open covering of X, which contains a finite open covering $(U(x_1), U(x_2), \ldots, U(x_k))$, since X is compact. Then $X = \bigcup_{i=1}^{k} U(x_i)$ and, since $U(x_i)$ contains at most one point of A (the point x_i being the only possibility) it follows that A is finite.

Theorem 1. *A compact set is closed.*

Proof. Let K be a compact set and let x be a point not belonging to K. If $y \in K$, there exist disjoint neighbourhoods $U^y(x)$ and $V(y)$, since X is a separated space. Then $(V(y) \,/\, y \in K)$ is an open covering of K and so contains a finite open covering $(V(y_1), V(y_2), \ldots, V(y_n))$. Since the neighbourhood $U(x) = \bigcap_{k=1}^{n} U^{y_k}(x)$ does not meet $V(K) = \bigcup_{k=1}^{n} V(y_k)$ it follows that K is closed.

Theorem 2. *If F is a closed set contained in a compact set K, then F is compact.*

Proof. If $(G_i \mid i \in I)$ is an open covering of F, then we can obtain an open covering of K by adjoining $-F = G_0$. Since K is compact, this covering contains a finite covering $(G_0, G_{i_1}, \ldots, G_{i_n})$ and then $(G_{i_1}, \ldots, G_{i_n})$ is a finite open covering of F. Therefore F is compact.

COROLLARY. *A subset of the straight line* **R** *(with the usual topology) is compact if and only if it is closed and bounded.*

Proof. By Theorem 1, a compact set is closed and, by Example 4, a compact set in **R** is bounded. Conversely, if A is a closed and bounded subset of **R**, then it is contained in a segment, which is compact, and so, by Theorem 2, A is compact.

Theorem 3. *A compact space is normal.*

Proof. Let X be a compact space. We first prove that X is regular. If F is a closed subset of X, then, by Theorem 2, F is compact. Let x be a point not in F. For each point y of F there exist disjoint neighbourhoods $V(y)$ and $U^y(x)$. Because F is compact the covering $(V(y) \mid y \in F)$ contains a finite covering $(V(y_1), \ldots, V(y_n))$. Then, writing $U(x) = \bigcap_{k=1}^{n} U^{y_k}(x)$ and $V(F) = \bigcup_{k=1}^{n} V(y_k)$, we have disjoint neighbourhoods containing x and F respectively, and so X is regular.

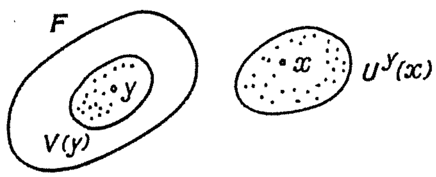


FIG. 19

We now complete the proof by using the regularity of X to prove that it is normal. Let F and F' be disjoint closed sets. If $x \in F$, there exist disjoint open sets $U(x)$ and $V^x(F')$, since X is regular. Since F is compact, the open covering $(U(x) \mid x \in F)$ contains a finite open covering $U(x_1), U(x_2), \ldots, U(x_n)$ and then the sets $U(F) = \bigcup_{k=1}^{n} U(x_k)$ and $V(F') = \bigcap_{k=1}^{n} V^{x_k}(F')$ are disjoint open neighbourhoods containing F and F' respectively.

Theorem 4. *The union of a finite number of compact sets is compact. The intersection of any number of compact sets is compact.*

Proof. Let $K_1, K_2, \ldots, K_n$ be compact sets and let A be their union. If $\mathscr{A} = (G_i \mid i \in I)$ is an open covering of A, then the sets of $\mathscr{A}$ which cover K_1 contain a finite covering, those which cover K_2 contain a finite covering and so on. Thus $\mathscr{A}$ contains a finite open covering of A and so A is compact.

Suppose now that K_i $(i \in I)$ is any collection of compact sets. Then $B = \bigcap_{i \in I} K_i$ is closed since it is the intersection of closed sets; but B is contained in each set K_i and so, since K_i is compact, B is compact by Theorem 2.

Theorem 5. *Let σ be a single-valued continuous mapping of a space X into a space Y. If K is a compact subset of X, then the image σK is a compact subset of Y.*

Proof. Let $(G_i \mid i \in I)$ be an open covering of σK; then the family $(\sigma^{-1}G_i \mid i \in I)$ is an open covering of K, and since K is compact, this contains a finite open covering $(\sigma^{-1}G_{i_1}, \sigma^{-1}G_{i_2}, \ldots, \sigma^{-1}G_{i_n})$. Then $(G_{i_1}, G_{i_2}, \ldots, G_{i_n})$ is an open covering of σK and therefore σK is compact.

COROLLARY. *If f is a numerical function defined and continuous on a space X, and if K is a compact subset of X, then f attains the value $\sup_{x \in K} f(x)$ in K (in other words, the supremum is a maximum). Similarly, f attains the value $\inf_{x \in K} f(x)$ in K.*

Proof. If f is a continuous mapping of X into $\mathbf{R}$, then, by Theorem 5, $f(K)$ is compact. Therefore, by the Corollary to Theorem 2, $f(K)$ is closed and bounded. Hence f has a maximum $f(a)$ and a minimum $f(b)$, where both a and b are in K.

Broadly speaking, it is this result which accounts for the interest of topologists in compact sets.

Theorem 6. *Let X be a topological space. The axiom* (1) *for compactness is equivalent to each of the following:*

(2) (Finite intersection axiom). *If $(F_i \mid i \in I)$ is a family of closed sets in X for which every finite intersection is non-empty, then $\bigcap_{i \in I} F_i \neq \emptyset$.*

(3) (Cluster axiom.) *Every filtered family (x_i) of elements of X admits at least one cluster point.*

(4) (Limit axiom.) *Every ultra-filtered family of elements of X admits a limit point.*

Proof. One method of proving the theorem is to show that

$$(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (1).$$

However, it is simpler here to prove the following:

$$(1) \Leftrightarrow (2) \Leftrightarrow (3) \Leftrightarrow (4).$$

(1) implies (2). Let $(F_i \,/\, i \in I)$ be a family of closed sets for which every finite intersection is non-empty. If $\bigcap_{i \in I} F_i = \emptyset$ then the family

$$(-F_i \,/\, i \in I)$$

is an open covering of X, which contains a finite covering $(-F_{i_1}, -F_{i_2}, \ldots, -F_{i_n})$ by (1). Then $\bigcup_{k=1}^{n} (-F_{i_k}) = X$ and so $\bigcap_{k=1}^{n} F_{i_k} = \emptyset$, contrary to hypothesis. Therefore $\bigcap_{i \in J} F_i \neq \emptyset$.

(2) implies (1). Let $\mathscr{G} = (G_i \,/\, i \in I)$ be an open covering of X and write $F_i = -G_i$. If $(G_i \,/\, i \in I)$ does not contain a finite open covering, then every finite intersection of the closed sets F_i is non-empty. Therefore, by (2), $\bigcap_{i \in I} F_i \neq \emptyset$, which contradicts the fact that $\mathscr{G}$ is a covering. Hence (1) is satisfied.

(2) implies (3). Let $(x(i) \,/\, i \in I, \mathscr{B})$ be a filtered family in X. Since $x(\mathscr{B})$ is a filter base, every finite intersection of sets of type $x(B)$ is non-empty; hence every finite intersection of sets $\overline{x(B)}$ is non-empty and so, by (2), we have

$$\bigcap_{B \in \mathscr{B}} \overline{x(B)} \neq \emptyset.$$

Therefore every point a of this intersection is a cluster point, since

$$(\forall V(a))\,(\forall B) : V(a) \cap x(B) \neq \emptyset.$$

(3) implies (2). Let $(F_i \,/\, i \in I)$ be a family of closed sets for which every finite intersection is non-empty; then the finite intersections of the sets F_i form a filter base $\mathscr{B}$. Choose a point $x(B)$ in each set $B \in \mathscr{B}$; then, by hypothesis, the filtered family $(x(B) \,/\, B \in \mathscr{B}, \subset)$ admits a cluster point a, and so

$$(\forall V(a))\,(\forall B_0)\,(\exists B;\ B \subset B_0) : x(B) \in V(a),$$

whence

$$(\forall V(a))\,(\forall B_0) : B_0 \cap V(a) \neq \emptyset.$$

Since B_0 is closed, it follows that a belongs to B_0; this is true for all sets B_0 and so a belongs to all sets F_i. Hence condition (2) is satisfied.

(3) implies (4). If $x(\mathscr{B})$ is an ultra-filter base, then every cluster point is a limit point, by Corollary 2 to Theorem 3 of § 4.

(4) implies (3). By Theorem 6 of § 6, Chapter I, there is an ultra-filter base $x(\mathscr{B}_0)$ such that $x(\mathscr{B}_0) \vdash x(\mathscr{B})$; then $(x(i) \,/\, i \in I, \mathscr{B})$ admits the limit point of $(x(i) \,/\, i \in I, \mathscr{B}_0)$ as a cluster point.

COROLLARY. *A necessary and sufficient condition for a filtered family (x_i) in a compact space to admit a limit point is that (x_i) admits one and only one cluster point.*

Proof. The condition is clearly necessary. To prove that it is sufficient, suppose that (x_i) admits only one cluster point a; then the theorem just proved shows that for any family $(y_j) \vdash (x_i)$ there exists a family $(z_k) \vdash (y_j)$ possessing a limit point, which can only be a; then a is a limit point of (x_i) by Theorem 5 of § 4.

§ 7. Connected sets

A topological space is said to be **disconnected** if it satisfies any one of the following equivalent conditions:

(1) There is a partition (G_1, G_2) consisting of two open sets.
(2) There exists a set G, neither empty nor the whole space, which is both open and closed.
(3) There exists a set G, neither empty nor the whole space, whose frontier is empty.

A topological space which is not disconnected is said to be **connected.** A connected subset of a space X is a subset A for which $(A, \mathscr{G}_A)$ is connected; in other words, A is connected if the following conditions cannot be satisfied simultaneously:

$$\begin{cases} G_1, G_2 \in \mathscr{G}, \\ A \subset G_1 \cup G_2, \\ G_1 \cap G_2 \cap A = \varnothing, \\ A \cap G_1 \neq \varnothing, \\ A \cap G_2 \neq \varnothing. \end{cases}$$

EXAMPLE 1. The space $\mathbf{R}$ is connected. For suppose that G_1 and G_2 are two disjoint open sets whose union is $\mathbf{R}$. The set G_1 (resp. G_2) is a union of disjoint open intervals D_1^i (resp. D_2^j). If G_1 and G_2 are both non-empty, there exist $a \in G_1$ and $b \in G_2$; without loss of generality we can assume that $a < b$. Since the segment $[a, b]$ is compact,[1] it is covered by a finite number of open intervals of the form D_1^i or D_2^j. If $a \in D_1^i =]\lambda, \mu[$ then $\mu \in [a, b]$; but μ does not belong to any of the selected intervals, which leads to a contradiction. Thus one of G_1 and G_2 is empty and so $\mathbf{R}$ is connected.

EXAMPLE 2. In any topological space a set $\{x\}$ containing a single element is connected and the null set $\varnothing$ is also connected.

EXAMPLE 3. In a quasi-separated space X, the set $A = \{x_1, x_2\}$ (where $x_1 \neq x_2$) is not connected, for there exist neighbourhoods $U(x_1)$ not containing x_2 and $U(x_2)$ not containing x_1. Denoting these by G_1 and G_2, we see that the above conditions for disconnectedness are satisfied.

[1] (Translator's note.) This result can be proved without appealing to the compactness of $[a, b]$.

EXAMPLE 4. Consider the sets in $\mathbf{R}^2$ represented in figure 20.

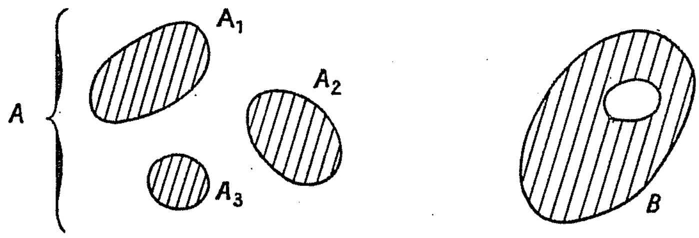


FIG. 20

Without arguing rigorously, we can see that there exist disjoint open sets G_1, G_2, G_3 covering A_1, A_2, A_3 respectively, so that A is not connected. But $-A$ is connected, and B is connected, although $-B$ is not connected.

Theorem 1. *If A_1 and A_2 are two intersecting connected sets, then the set $A = A_1 \cup A_2$ is connected.*

Proof. Suppose that A is disconnected. Then there exist open sets G_1 and G_2 such that

$$\begin{cases} A \subset G_1 \cup G_2, \\ A \cap G_1 \cap G_2 = \emptyset, \\ A \cap G_1 \neq \emptyset, \\ A \cap G_2 \neq \emptyset. \end{cases}$$

Let a be an element of $A_1 \cap A_2$; without loss of generality, we can suppose that

$$a \in G_1 \; ; \quad G_2 \cap A_1 \neq \emptyset.$$

Then we have

$$\begin{cases} A_1 \subset G_1 \cup G_2, \\ A_1 \cap G_1 \cap G_2 = \emptyset, \\ A_1 \cap G_1 \neq \emptyset & \text{(because it contains } a), \\ A_1 \cap G_2 \neq \emptyset. \end{cases}$$

Therefore A_1 is not connected, which is contrary to hypothesis.

COROLLARY 1. *If $A_1, A_2, \ldots, A_n$ are connected sets such that $A_i \cap A_{i+1} \neq \emptyset$ for $i = 1, 2, \ldots, n-1$, the set $A = \bigcup_{i=1}^{n} A_i$ is connected.*

Proof. By the theorem, $A_1 \cup A_2$ is connected; hence $A_1 \cup A_2 \cup A_3 = (A_1 \cup A_2) \cup A_3$ is connected and so on.

COROLLARY 2. *If $\mathscr{A} = (A_i \mid i \in I)$ is a family of connected sets whose intersection is non-empty, their union is a connected set.*

A similar argument can be used in this case.

COROLLARY 3. *If a family of connected sets $\mathscr{A} = (A_i \mid i \in I)$ contains a set A_0 which meets all the A_i, the union of the A_i is a connected set.*

Proof. The family $\mathscr{A}_0 = (A_0 \cup A_i \,/\, i \in I)$ satisfies the conditions of Corollary 2.

Theorem 2. *If A is a connected set, so is its closure $\bar{A}$.*

Proof. If $\bar{A}$ is not connected, there exist open sets G_1 and G_2 such that

$$\begin{cases} \bar{A} \subset G_1 \cup G_2, \\ \bar{A} \cap G_1 \cap G_2 = \varnothing, \\ \bar{A} \cap G_1 \neq \varnothing, \\ \bar{A} \cap G_2 \neq \varnothing. \end{cases}$$

Then

$$\begin{cases} A \subset G_1 \cup G_2, \\ A \cap G_1 \cap G_2 = \varnothing, \\ A \cap G_1 \neq \varnothing \text{ (for otherwise, } A \subset -G_1\text{, so that } \bar{A} \subset -G_1), \\ A \cap G_2 \neq \varnothing \text{ (for otherwise, } A \subset -G_2\text{, so that } \bar{A} \subset -G_2). \end{cases}$$

This implies that A is not connected, contrary to hypothesis.

Theorem 3. *If σ is a single-valued continuous mapping of a topological space X into a topological space Y, and A is a connected subset of X, then the set σA is connected.*

Proof. If σA is not connected, there exist open sets G_1 and G_2 such that

$$\begin{cases} \sigma A \subset G_1 \cup G_2, \\ \sigma A \cap G_1 \cap G_2 = \varnothing, \\ \sigma A \cap G_1 \neq \varnothing, \\ \sigma A \cap G_2 \neq \varnothing. \end{cases}$$

Then

$$\begin{cases} A \subset \sigma^{-1}G_1 \cup \sigma^{-1}G_2, \\ A \cap \sigma^{-1}G_1 \cap \sigma^{-1}G_2 = \varnothing, \\ A \cap \sigma^{-1}G_1 \neq \varnothing, \\ A \cap \sigma^{-1}G_2 \neq \varnothing \end{cases}$$

and so A is not connected, contrary to hypothesis.

Theorem 4. *The only connected sets in* **R** *are the intervals (these include, in particular, the sets*

$$\varnothing =]\lambda, \lambda[\quad \text{and} \quad \{\lambda\} \doteq [\lambda, \lambda]).$$

Proof. As in example 1 above, we can prove that an interval is connected. Suppose now that A is any connected subset of **R**. To show that A is an interval, it is sufficient to show that if $a \in A$ and $b \in B$, then each point λ between a and b belongs to A. If this were not true, then we should have a point $\lambda \in]a, b[$, such that $\lambda \notin A$; then $G_1 =]-\infty, \lambda[$ and $G_2 =]\lambda, +\infty[$ would be disjoint open sets covering A, which would contradict the hypothesis that A is connected.

COROLLARY. *If A is a connected subset of X, then any numerical function f defined and continuous on X attains in A every value strictly between* $\sup_{x \in A} f(x)$ *and* $\inf_{x \in A} f(x)$.

Proof. If f is a continuous mapping of X into $\mathbf{R}$, then $f(A)$ is connected by Theorem 3 and is therefore an interval by Theorem 4.

Broadly speaking, it is this result which accounts for the interest of topologists in connected sets.

§ 8. Numerical functions defined on a topological space

The ideas of continuity introduced above apply in particular to numerical functions and we say that a numerical function f, defined on a topological space X, is **continuous at a point** x_0 if it is a mapping of X into $\mathbf{R}$ and, as such, is continuous at x_0; in other words if, to each $\varepsilon > 0$, there corresponds a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad |f(x)-f(x_0)| < \varepsilon.$$

In certain situations this idea turns out to be too restrictive and we are forced to consider the concepts of lower and upper semi-continuity,[1] as given in the following definitions.

A numerical function defined on X is said to be **lower semi-continuous at a point** x_0 if, to each $\varepsilon > 0$, there corresponds a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad f(x) > f(x_0)-\varepsilon.$$

Similarly a numerical function f is said to be **upper semi-continuous at** x_0 if, to each $\varepsilon > 0$, there corresponds a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad f(x) < f(x_0)+\varepsilon.$$

We note that f is continuous at x_0 if and only if it is both lower and upper semi-continuous at that point.

Properties of upper semi-continuous functions can easily be deduced from those of lower semi-continuous functions by replacing $f(x)$ by $-f(x)$; we shall therefore consider only the case of lower semi-continuous functions.

We say that a function f is **lower semi-continuous (in X)** if it is lower semi-continuous at each point of X.

EXAMPLE 1. As an intuitive example, consider the photographic image X of an opaque object Y. The object, a set of points in three-dimensional space, can be regarded as a topological space; the image is a set of points in

[1] (Translator's note.) The terms 'lower semi-continuous' and 'upper semi-continuous' for single-valued numerical functions must not be confused with similar phrases introduced in Chapter VI and referring to different concepts concerning multi-valued functions. For the latter, the abbreviations l.s.c. and u.s.c. are used subsequently; these abbreviations are not used for the other types of semi-continuity.

a plane and can also be regarded as a topological space. Let σ be the mapping in which there corresponds to each point x of X the point y of Y which is seen on the photograph at x. Given any point $x \in X$, let $f(x)$ be the actual distance of the observer from the point σx of the object. The numerical function f so defined is not necessarily continuous in X, but it is upper semi-continuous.

EXAMPLE 2. Let f be a numerical function defined in $[0, 1]$ and having a continuous derivative. The set $C_f = \{(x, f(x)) \,/\, x \in [0, 1]\}$ in $\mathbf{R}^2$ is called the **representative curve** of f. The number

$$l(C_f) = \int_0^1 \sqrt{1+\{f'(x)\}^2}\,\mathrm{d}x$$

is called the **length** of the curve C_f. If C_f and C_g are two curves, then

$$d(C_f, C_g) = \sup\{\,|f(x)-g(x)|\,/\,x \in [0, 1]\}$$

is called the **distance** between the curves. With this definition of distance, curves form a metric space (and so a topological space) which we denote by $\mathscr{C}$.

Let $C_1, C_2, \ldots$ be a sequence of curves defined as follows: C_1 is the semi-circle on the unit segment $[0, 1]$ as diameter, as indicated in figure 21; C_2 is the union of the semi-circles on $[0, \frac{1}{2}]$, $[\frac{1}{2}, 1]$ as diameters; C_3 is the union of the semi-circles on $[0, \frac{1}{4}]$, $[\frac{1}{4}, \frac{1}{2}]$, $[\frac{1}{2}, \frac{3}{4}]$, $[\frac{3}{4}, 1]$ as diameters and so on. Clearly this sequence of curves in $\mathscr{C}$ tends to a limit, which is the segment $[0, 1]$. However, the lengths of the curves are given by

$$l(C_1) = \frac{\pi}{2},$$

$$l(C_2) = 2\times\pi\times\frac{1}{4} = \frac{\pi}{2},$$

$$l(C_3) = 4\times\pi\times\frac{1}{8} = \frac{\pi}{2}, \text{ etc.}$$

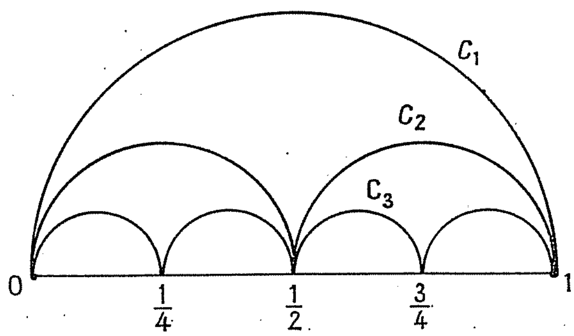


FIG. 21

Therefore the sequence $(l(C_n))$ does not tend to $l([0, 1])$; this shows that $l(C)$ is not a continuous function in $\mathscr{C}$, but we can prove that it is lower semi-continuous.

Theorem 1. *A necessary and sufficient condition for a numerical function f to be lower semi-continuous is that, for each number λ, the set $S_\lambda = \{x \,/\, x \in X, f(x) > \lambda\}$ is open.*

Proof. The family $\mathscr{D}^-$ of intervals

$$(]\lambda, +\infty[\,/\, \lambda \in \mathbf{R})$$

determines a topology in $\mathbf{R}$. Then it follows quickly from the definition of lower semi-continuity that f is lower semi-continuous if and only if it is a continuous mapping of X into the topological space $(\mathbf{R}, \mathscr{D}^-)$. Since $S_\lambda = f^{-1}(]\lambda, +\infty[)$, the theorem stated is equivalent to this result.

COROLLARY. *A numerical function f is lower semi-continuous if and only if, for each number λ, the set $T_\lambda = \{x \,/\, f(x) \leqq \lambda\}$ is closed.*

Theorem 2. *If K is a compact subset of X, a lower semi-continuous function f attains in K the value $m = \inf_{x \in K} f(x)$ (in other words, $\inf_{x \in K} f(x)$ is a minimum).*

Proof. Let μ be a number such that $\mu > m$. Then $S_\mu = \{x \,/\, x \in K, f(x) \leqq \mu\}$ is not empty and is a closed subset of the compact set K; hence, by the finite intersection axiom, we have

$$\bigcap_{\mu > m} S_\mu \neq \varnothing.$$

For any point x_0 of this intersection, we have $f(x_0) = m$.

Theorem 3. *If $(f_i \,/\, i \in I)$ is a family of lower semi-continuous functions, the function g such that $g(x) = \sup_{i \in I} f_i(x)$ is lower semi-continuous.*

Proof. Let x_0 be any point of X. Given $\varepsilon > 0$, there exists an index i such that

$$f_i(x_0) > g(x_0) - \varepsilon.$$

Since f_i is lower semi-continuous, there exists a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad f_i(x) > f_i(x_0) - \varepsilon > g(x_0) - 2\,\varepsilon$$

and therefore

$$x \in U(x_0) \quad \Rightarrow \quad g(x) \geqq g(x_0) - 2\,\varepsilon.$$

Hence g is lower semi-continuous at x_0; since x_0 is arbitrary, the theorem is proved.

Theorem 4. *If $f_1, f_2, \ldots, f_n$ are lower semi-continuous functions then the function $h(x) = \inf_{i \in I} f_i(x)$ (where $I = \{1, 2, \ldots, n\}$) is lower semi-continuous.*

Proof. Let x_0 be any point of X. Given $\varepsilon > 0$, there exists a neighbourhood $U_i(x_0)$ such that

$$x \in U_i(x_0) \quad \Rightarrow \quad f_i(x) > f_i(x_0) - \varepsilon.$$

If $x \in \bigcap_{i \in I} U_i(x_0)$, then, for all i,

$$f_i(x) > f_i(x_0) - \varepsilon,$$

whence

$$h(x) = \inf_i f_i(x) \geqq \inf_i (f_i(x_0) - \varepsilon) = h(x_0) - \varepsilon$$

and so h is lower semi-continuous.

Theorem 5. *If f and g are two lower semi-continuous numerical functions, the function $f+g$ is lower semi-continuous.*

Proof. The set

$$\{x \,/\, f(x) + g(x) > a\} = \bigcup_\lambda \{x \,/\, f(x) > a - \lambda\} \cap \{x \,/\, g(x) > \lambda\}$$

is a union of open sets and so is open.

Theorem 6. *If f and g are two positive lower semi-continuous numerical functions, the function h defined by $h(x) = f(x)\, g(x)$ is lower semi-continuous.*

Proof. If $\alpha > 0$, the set

$$\{x \,/\, f(x)\, g(x) > \alpha\} = \bigcup_{\lambda > 0} \{x \,/\, f(x) > \lambda\} \cap \left\{x \,/\, g(x) > \frac{\alpha}{\lambda}\right\}$$

is a union of open sets and so is open.

§ 9. Products and sums of topological spaces

In order to fix our ideas, we first consider a countable family $(X_1, X_2, \ldots)$ of topological spaces. If $\xi = (x_1, x_2, \ldots)$ is an element of the set $X = X_1 \times X_2 \times \ldots$, the **projection** of X on X_i is the single-valued mapping π_i (or proj_{X_i}) of X into X_i defined by $\pi_i(x_1, x_2, \ldots, x_i, \ldots) = x_i$.

We wish to define a topological structure $\mathscr{G}$ on X for which π_i is a continuous mapping; in other words, if

$$\mathscr{G}_i = (G_i^k \,/\, k \in K_i)$$

is the topology for X_i, then $\mathscr{G}$ must contain the sets

$$(\pi_i)^{-1} G_i^k = X_1 \times X_2 \times \ldots \times X_{i-1} \times G_i^k \times X_{i+1} \times X_{i+2} \times \ldots$$

The coarsest topological structure $\mathscr{G}$ which satisfies this condition is that generated by the sets $(\pi_i)^{-1} G_i^k$, where i and k vary (cf. example 3, § 3). A finite intersection of these sets $(\pi_i)^{-1}G_i^k$ is called an **elementary open set** in X; thus an elementary open set is a set E of the form

$$E = X_1 \times G_2^k \times X_3 \times X_4 \times G_5^l \times X_6 \times \ldots$$

where only a finite number of the G_i differ from X_i. Then the topology $\mathscr{G}$ is obtained by forming unions of elementary open sets. With this definition of $\mathscr{G}$, we call $(X, \mathscr{G})$ the **topological product** of the spaces $(X_i, \mathscr{G}_i)$ and this new topological space is denoted by

$$X = \prod X_i.$$

It is easily seen that this definition can be extended to the cases in which the family of spaces X_i is not countable.

Example. The plane $\mathbf{R}^2$, with the metric topology defined above, is the topological product of the space $\mathbf{R}$ with itself; it is easily seen that every open set of $\mathbf{R}^2$ is a union of elementary open sets $E = G_1^k \times G_2^k$.

Let $(X_i, \mathscr{G}_i)$ be a family of topological spaces. Then the pair $(X, \mathscr{G})$ consisting of the set

$$X = \{(i, x_i) \mid i \in I,\ x_i \in X_i\}$$

and the topology $\mathscr{G}$ on X generated by the sets (i, G_i^k), where $k \in K_i$ and $i \in I$, is called the **topological sum** of the X_i. This new space is denoted by

$$X = \sum X_i.$$

Theorem 1. *If A is a subset of a topological space X and B is a subset of a topological space Y, then*

$$\bar{A} \times \bar{B} = \overline{A \times B}.$$

Proof. π_X is a continuous mapping of $X \times Y$ into X and $\pi_X(A \times B) \subset A$; therefore, by Corollary 1 to the theorem of § 3, we have

$$\pi_X(\overline{A \times B}) \subset \bar{A}.$$

Similarly, the projection π_Y of $X \times Y$ into Y satisfies

$$\pi_Y(\overline{A \times B}) \subset \bar{B},$$

whence

$$\overline{A \times B} \subset \bar{A} \times \bar{B}.$$

If $(x, y) \in \bar{A} \times \bar{B}$ and $(x, y) \notin \overline{A \times B}$, there exists an elementary neighbourhood $U(x) \times V(y)$ in $X \times Y$ such that

$$(U(x) \times V(y)) \cap (A \times B) = \varnothing.$$

Then either $U(x) \cap A = \varnothing$ or $V(y) \cap B = \varnothing$; therefore $(x, y) \notin \bar{A} \times \bar{B}$, contrary to hypothesis. It follows that

$$\bar{A} \times \bar{B} = \overline{A \times B}.$$

Theorem 2. *If G is an open set of $X = \prod_{i \in I} X_i$, the projection $\pi_i G$ is an open set of X_i.*

Proof. If G is an elementary open set, the result is true trivially. If G is any open set, it is a union of elementary open sets E^k and then

$$\pi_i G = \pi_i \Big(\bigcup_k E^k\Big) = \bigcup_k \pi_i E^k$$

is clearly an open set.

LEMMA. *Let x_1^0 be a point of X_1. Then the single-valued mapping σ of X_2 into $X_1 \times X_2$ defined by $\sigma x_2 = (x_1^0, x_2)$ is continuous.*

Proof. If G is an open set in $X_1 \times X_2$, then $\sigma^{-1}G$ is the union of projections (on X_2) of elementary open sets of G meeting $\{x_1^0\} \times X_2$.

Theorem 3. *If F is a closed set in $X_1 \times X_2$ and X_2 is a compact space, then $\pi_1 F$ is a closed set in X_1.*

Proof. Let x_1^0 be a point of X_1. Then, by the lemma, $\{x_1^0\} \times X_2$ is compact, since it is the image of a compact set under a continuous mapping.

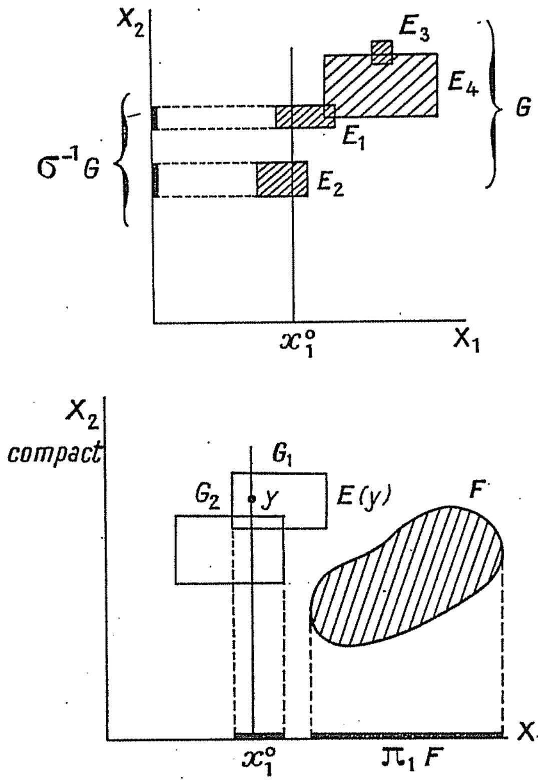


FIG. 22

If $x_1^0 \notin \pi_1 F$ and $y \in \{x_1^0\} \times X_2$, then $y \notin F$. Therefore there exists an elementary open neighbourhood $E(y)$ disjoint from F. The covering $(E(y) \,/\, y \in \{x_1^0\} \times X_2)$ contains a finite covering $E(y_1), E(y_2), \ldots, E(y_n)$; the set $\bigcap_{k=1}^{n} \pi_1 E(y_k)$ is an open neighbourhood of x_1^0 disjoint from $\pi_1 F$ and so the set $\pi_1 F$ is closed.

Tychonoff's Theorem. *A necessary and sufficient condition for the topological product $X = \prod_{i \in I} X_i$ to be compact is that each space X_i is compact.*

Proof. If X is compact, then X_i is separated, for if x_i^0 and y_i^0 are distinct elements of X_i we can find disjoint elementary open neighbourhoods for the points

$$(x_i) = (x_i \mid i \in I) \quad \text{and} \quad (y_i) = (y_i \mid i \in I)$$

(with $x_j = y_j$ if $j \neq i$, $x_i = x_i^0$ and $y_i = y_i^0$).

Thus X_i is a separated space and, since $X_i = \pi_i X$ is the image of a compact space under a continuous mapping, it is itself a compact space.

Suppose conversely that each space X_i is compact. Then it is easily shown that X is separated. Let $x = (x(k) \mid k \in K, \mathscr{B})$ be an ultra-filtered family in X and consider the projection $\pi_i x = (x_i(k) \mid k \in K, \mathscr{B})$. The latter is the image under a single-valued mapping of an ultra-filtered family and so is itself an ultra-filtered family. Since X_i is compact, this family admits a limit point $x_i^0 \in X$. Then

$$(x(k) \mid k \in K, \mathscr{B}) \to x^0 = (x_i^0 \mid i \in I).$$

Therefore every ultra-filtered family in X admits a limit point and so X is compact by Theorem 6 of § 6.

EXAMPLE 1. The square $[0, 1] \times [0, 1]$ is compact; more generally, the n dimensional cube $[0, 1]^n$ in $\mathbf{R}^n$ is compact.

EXAMPLE 2. The plane $\mathbf{R}^2$ is locally compact, for each point $x = (x^1, x^2)$ admits a fundamental base of compact neighbourhoods, consisting of the sets

$$N_{\varepsilon, \varepsilon'}(x) = [x^1 - \varepsilon, x^1 + \varepsilon] \times [x^2 - \varepsilon', x^2 + \varepsilon'].$$

More generally, the space $\mathbf{R}^n$ is locally compact.

Theorem 4. *The topological sum $X = \sum_{i \in I} X_i$ is not compact if I is infinite. The topological sum $X = \sum_{i=1}^{n} X_i$ of a finite number of topological spaces is compact if and only if each space X_i is compact.*

Proof. If I is infinite, the sets (i, X_i) determine an open covering of X and this covering does not contain a finite open covering; therefore ΣX_i is not compact.

Suppose now that $X = \sum_{i=1}^{n} X_i$ is the sum of a finite number of spaces X_i and that X is compact. Then X_i is clearly separated. Let $(G_i^k \mid k \in K_i)$ be an open covering of X_i; then $((i, G_i^k) \mid k \in K_i)$ is an open covering of (i, X_i) and so it contains a finite covering $(i, G_i^1), (i, G_i^2), \ldots, (i, G_i^m)$, for (i, X_i), being a closed subset of a compact space, is compact. Then the sets $G_i^1, G_i^2, \ldots, G_i^m$ form a finite open covering of K_i and therefore X_i is compact.

Suppose conversely that $X = \sum_{i=1}^{n} X_i$ is the sum of a finite number of compact spaces. It is easily shown that X is separated, and if $(G_k \mid k \in K)$ is an open covering of X, then we see that there exists a finite covering of $(1, X_1)$; similarly there exists a finite covering of $(2, X_2)$ and so on, whence there exists a finite covering of X.

The remaining theorems of this section are analogous results concerning connectedness.

Theorem 5. *The topological product $X_1 \times X_2$ is connected if and only if X_1 and X_2 are connected.*

Proof. If $X_1 \times X_2$ is connected, then $X_1 = \pi_1(X_1 \times X_2)$ is the image of a connected space under a single-valued continuous transformation π_1 and therefore X_1 is connected. Similarly X_2 is connected.

Suppose now that X_1 and X_2 are connected. Let a_1 be a point of X_1; by the lemma immediately preceding Theorem 3, the single-valued mapping of X_2 into $X_1 \times X_2$ defined by $\sigma y = (a_1, y)$ is continuous. Therefore the set $A_1 = \{a_1\} \times X_2$, being the image of the connected space X_2 under this mapping, is connected. Similarly, if $y \in X_2$, the set $A(y) = X_1 \times \{y\}$ is connected. Moreover, we have

$$A_1 \cap A(y) = \{(a_1, y)\} \neq \emptyset$$

and so, by Corollary 3 to Theorem 1 of § 7,

$$A_1 \cup \bigcup_{y \in X_2} A(y) = X_1 \times X_2$$

is also connected.

Theorem 6. *The topological sum $X_1 + X_2$ is not connected.*

Proof. In the space $X_1 + X_2$, the set

$$(1, X_1) = \{(1, x_1) \,/\, x_1 \in X_1\}$$

is both open and closed but is neither the null set nor the whole space.

CHAPTER V

TOPOLOGICAL PROPERTIES OF METRIC SPACES

§ 1. Topology of a metric space

We have already seen that a **metric**[1] for a set X is a numerical function d defined on $X \times X$ such that

$$\begin{array}{ll} (1) & d(x, y) \geqq 0, \\ (2) & d(x, y) = 0 \quad \Leftrightarrow \quad x = y, \\ (3) & d(x, y) = d(y, x), \\ (4) & d(x, y) + d(y, z) \geqq d(x, z) \text{ (triangular inequality).} \end{array}$$

If d is a generalised numerical function (with values in $\hat{\mathbf{R}}$) satisfying axioms (1), (2), (3) and (4) we say that d is a **generalised metric.**

With each metric space (X, d) we associate a topology $\mathscr{G}_d$, defined as above (Chapter IV, § 1). Conversely, we say that a topological space $(X, \mathscr{G})$ **admits a metric** d or is **metrisable** if the topology $\mathscr{G}_d$ of the metric space (X, d) coincides with $\mathscr{G}$. Clearly certain topological spaces do not admit a metric (for example, non-separated spaces). On the other hand a topological space can admit several metrics. Two metrics d and d' on the same set X are said to be **equivalent** if for each strictly positive number ε there exist numbers η and η' such that

$$\begin{array}{lll} (1)\ d(x, y) \leqq \eta & \Rightarrow & d'(x, y) \leqq \varepsilon, \\ (2)\ d'(x, y) \leqq \eta' & \Rightarrow & d(x, y) \leqq \varepsilon. \end{array}$$

Theorem 1. *If d and d' are equivalent metrics for X, we have*

$$\mathscr{G}_d = \mathscr{G}_{d'},$$

Proof. If $G \in \mathscr{G}_d$, then to each $a \in G$ there correspond numbers $\varepsilon > 0$ and $\eta' > 0$ such that

$$d'(x, a) \leqq \eta' \quad \Rightarrow \quad d(x, a) \leqq \varepsilon \quad \Rightarrow \quad x \in G.$$

Hence a is an interior point of G relative to the topology $\mathscr{G}_{d'}$, and so $G \in \mathscr{G}_{d'}$. Therefore, by symmetry, we have $\mathscr{G}_d = \mathscr{G}_{d'}$.

[1] The idea of a metric space is due to M. Fréchet (*Rend. Circ. Mat. di Palermo*, 22, 1906); also that of complete space introduced in § 4 (thesis). Important contributions to the theory of metric spaces have been made by Urysohn, Hausdorff, Alexandroff and Hopf.

EXAMPLE. Let

$$x = (x^1, x^2, x^3) \quad \text{and} \quad y = (y^1, y^2, y^3)$$

be two points in $\mathbf{R}^3$ and let d, d', d'' be the three metrics defined by

$$\begin{aligned} d(x, y) &= \sqrt{(x^1-y^1)^2+(x^2-y^2)^2+(x^3-y^3)^2}, \\ d'(x, y) &= |x^1-y^1| + |x^2-y^2| + |x^3-y^3|, \\ d''(x, y) &= \max \{ |x^1-y^1|, |x^2-y^2|, |x^3-y^3| \}. \end{aligned}$$

Then

$$\begin{aligned} d(x, y) &\leqq d'(x, y);\ d'(x, y) \leqq 3\, d(x, y); \\ d'(x, y) &\leqq 3\, d''(x, y);\ d''(x, y) \leqq d'(x, y); \end{aligned}$$

and so these three metrics are equivalent.

Theorem 2. *If (X, d) is a metric space, then d is a continuous numerical function on the topological product $X \times X$.*

Proof. Let (x_0, y_0) be a point of $X \times X$ and consider the neighbourhood

$$V_\varepsilon(x_0, y_0) = \mathring{B}_\varepsilon(x_0) \times \mathring{B}_\varepsilon(y_0).$$

If (x, y) belongs to this neighbourhood, then

$$\begin{aligned} |d(x, y)-d(x_0, y_0)| &\leqq |d(x, y)-d(x, y_0)| + |d(x, y_0)-d(x_0, y_0)| \\ &\leqq d(y, y_0)+d(x, x_0) \leqq 2\,\varepsilon, \end{aligned}$$

and so d is continuous in $X \times X$.

If A is a subset of a metric space (X, d), then clearly the topological subspace $(A, \mathscr{G}_A)$ admits d restricted to $A \times A$ as a metric.

The **distance** between a point x and a set A is defined to be

$$d(x, A) = \inf \{d(x, y) \,/\, y \in A\}.$$

If A is compact, then $\{x\} \times A$ is compact in $X \times X$ and so, by Theorem 2, this infimum is a minimum.

If A and B are two sets, then we define the **distance** between them to be the number

$$d(A, B) = \inf \{d(x, y) \,/\, x \in A,\ y \in B\}.$$

If A and B are compact, then $A \times B$ is a compact subset of $X \times X$ and so, by Theorem 2, this infimum is a minimum.

The number

$$\delta(A) = \sup \{d(x, y) \,/\, x \in A,\ y \in A\}$$

is called the **diameter** of the set A. As in the previous cases, we can easily see that this supremum is a maximum if A is compact.

A set A is said to be **bounded** if $\delta(A) < +\infty$; any metric d (or even generalised metric) is equivalent to a metric d' for which X is a bounded set. For if we write

$$\begin{aligned} d'(x, y) &= d(x, y) \quad \text{if} \quad d(x, y) \leqq 1, \\ d'(x, y) &= 1 \qquad\quad\ \ \text{if} \quad d(x, y) > 1, \end{aligned}$$

then d' is a metric, since

$$\begin{aligned} &d'(x, y) \geqq 0, \\ &d'(x, y) = 0 \quad \Leftrightarrow \quad x = y, \\ &d'(x, y) = d'(y, x), \\ &d'(x, y)+d'(y, z) \geqq d'(x, z). \end{aligned}$$

The triangular inequality is immediate if $d'(x, y) = 1$ or $d'(y, z) = 1$; in the other cases we have

$$d'(x, y)+d'(y, z) = d(x, y)+d(y, z) \geqq d(x, z) \geqq d'(x, z).$$

Moreover, it is clear that d and d' are equivalent.

A property of a metric space (X, d) is called **topological** if it is unchanged whenever d is replaced by an equivalent metric d'. The above remarks show that the property of being bounded is not topological.

In what follows we can always suppose, in studying the topology of a metric space (X, d), that X is bounded.

Theorem 3. *If A is a subset of a metric space X, the distance $d(x, A)$ from a point x to this subset determines a continuous function f in X, where $f(x) = d(x, A)$.*

Proof. Let ε be a strictly positive number and let $x, y \in X$ be such that $d(x, y) \leqq \varepsilon$. There exists an element $a \in A$ such that

$$d(x, a) \leqq d(x, A)+\varepsilon.$$

Therefore

$$d(y, A) \leqq d(y, a) \leqq d(y, x)+d(x, a) \leqq d(x, A)+2\,\varepsilon$$

and so, by symmetry, we have

$$\left| d(y, A)-d(x, A) \right| \leqq 2\,\varepsilon.$$

Since this inequality holds whenever $d(x, y) \leqq \varepsilon$, it follows that the function f defined by $f(x) = d(x, A)$ is continuous.

A metric space possesses certain interesting topological properties; one of these is that there exists a countable fundamental base of neighbourhoods, namely

$$\mathcal{N}(x) = (B_1(x), B_{\frac{1}{2}}(x), B_{\frac{1}{3}}(x), \ldots)$$

and another is given in the following theorem.

Theorem 4. *A metric space is normal.*

Proof. Let (X, d) be a metric space. Then it is separated, for if $x \in X$, $y \in X$ and $x \neq y$, then $d(x, y) = \varepsilon > 0$; since

$$d(z, x) \leqq \frac{\varepsilon}{4} \quad \Rightarrow \quad d(z, y) \geqq d(x, y) - d(z, x) \geqq \varepsilon - \frac{\varepsilon}{4} = \frac{3\varepsilon}{4},$$

we have

$$B_{\frac{\varepsilon}{4}}(x) \cap B_{\frac{\varepsilon}{4}}(y) = \emptyset.$$

Suppose now that F_1 and F_2 are two disjoint closed sets. If $x \in F_1$, there exists a positive number ε such that $B_\varepsilon(x) \subset -F_2$ and we have

$$d(x, F_2) \geqq \varepsilon > 0 = d(x, F_1).$$

It follows that

$$V(F_1) = \{x \,/\, d(x, F_1) < d(x, F_2)\}$$

and

$$V(F_2) = \{x \,/\, d(x, F_2) < d(x, F_1)\}$$

are disjoint sets containing F_1 and F_2 respectively. Moreover, since $d(x, F_1) - d(x, F_2)$ determines a continuous function of x, $V(F_1)$ is open; similarly $V(F_2)$ is open. Therefore X is normal.

§ 2. Sums and products of metric spaces

We consider a family of metric spaces (X_i, d_i) and ask whether the topological sum $\sum_{i \in I} X_i$ and the topological product $\prod_{i \in I} X_i$ are metrisable spaces.

Theorem 1. *The topological sum $\sum_{i \in I} X_i$ of the metric spaces (X_i, d_i) admits a metric d, which is defined as follows: for each $i \in I$, let a_i be an element of X_i and write*

$$d(x, y) \begin{cases} = d_i(x_i, y_i) & \text{if } x = (i, x_i),\ y = (i, y_i), \\ = d_i(x_i, a_i) + 1 + d_j(y_j, a_j) & \text{if } x = (i, x_i),\ y = (j, y_j),\ i \neq j. \end{cases}$$

Proof. The function d is a metric, for

$$\begin{aligned} &d(x, y) \geqq 0, \\ &d(x, y) = 0 \quad \Leftrightarrow \quad x = y, \\ &d(x, y) = d(y, x), \\ &d(x, y) + d(y, z) \geqq d(x, z). \end{aligned}$$

To prove the triangular inequality, it is sufficient to consider the following cases separately:

Case 1. If $x = (i, x_i)$, $y = (i, y_i)$, $z = (i, z_i)$ we have

$$d(x, y) + d(y, z) = d_i(x_i, y_i) + d_i(y_i, z_i) \geqq d_i(x_i, z_i) = d(x, z).$$

Case 2. If $x = (i, x_i)$, $y = (i, y_i)$, $z = (j, z_j)$ where $i \neq j$, we have

$$\begin{aligned} d(x, y)+d(y, z) &= d_i(x_i, y_i)+d_i(y_i, a_i)+1+d_j(a_j, z_j) \\ &\geqq d_i(x_i, a_i)+1+d_j(a_j, z_j) = d(x, z). \end{aligned}$$

Case 3. If $x = (i, x_i)$, $y = (j, y_j)$, $z = (k, z_k)$ where $i \neq j$ and $j \neq k$, we have

$$d(x, y)+d(y, z) = d_i(x_i, a_i)+1+2\,d_j(a_j, y_j)+1+d_k(a_k, z_k) \geqq d(x, z).$$

To complete the proof, we verify that the topology $\mathscr{G}_d$ is that of the topological sum.

Theorem 2. *The topological product* $\prod_{i=1}^{n} X_i$ *of the metric spaces*

$$(X_1, d_1), (X_2, d_2), (X_3, d_3), \ldots, (X_n, d_n)$$

admits a metric d, defined by writing

$$d(x, y) = \max_i d_i(x^i, y^i)$$

for the distance between $x = (x^1, x^2, \ldots, x^n)$ *and* $y = (y^1, y^2, \ldots, y^n)$.

Proof. The function d is a metric, for

$$\begin{aligned} &d(x, y) \geqq 0, \\ &d(x, y) = 0 \quad \Leftrightarrow \quad x = y, \\ &d(x, y) = d(y, x), \\ &d(x, y)+d(y, z) \geqq d(x, z). \end{aligned}$$

To prove the triangular inequality, we observe that, for all $j \leqq n$,

$$\max_i d_i(x^i, y^i)+\max_i d_i(y^i, z^i) \geqq d_j(x^j, y^j)+d_j(y^j, z^j) \geqq d_j(x^j, z^j),$$

whence

$$d(x, y)+d(y, z) \geqq \max_j d_j(x^j, z^j) = d(x, z).$$

To complete the proof, we verify that the topology is that of the topological product.

Remark. We could equally well have proved that $\prod_{i=1}^{n} X_i$ admits the metric d_p defined by

$$d_p(x, y) = \sqrt[p]{\sum_{i=1}^{n} [d_i(x^i, y^i)]^p}.$$

The triangular inequality is immediate if $p = 1$; in the other cases, it follows from a general result (cf. § 9, Chapter VIII). The metric d_p is equivalent to the metric d defined in Theorem 2, for

$$\begin{aligned} d_p(x, y) &\leqq d(x, y) \times \sqrt[p]{n}, \\ d(x, y) &\leqq d_p(x, y). \end{aligned}$$

Applications. The ideas contained in the above remarks enable us to define a new class of metric spaces. Let p be an integer and let L_p be the set of numerical sequences (x^n) such that $\sum_{n=1}^{\infty} | x^n |^p < +\infty$. If $x = (x^n) \in L_p$, then $\lambda x = (\lambda x^n) \in L_p$ for each number λ, since

$$\sum_{n=1}^{\infty} | \lambda x^n |^p = | \lambda |^p \sum_{n=1}^{\infty} | x^n |^p < +\infty.$$

If $x = (x^n) \in L_p$ and $y = (y^n) \in L_p$, we have $x+y = (x^n+y^n) \in L_p$, since

$$\sum_{n=1}^{\infty} | x^n+y^n |^p \leqq 2^p \sum_{n=1}^{\infty} (| x^n |^p + | y^n |^p) < +\infty.$$

We also note that $0 = (0, 0, 0, \ldots) \in L_p$.

We observe that $L_1 \subset L_2 \subset L_3 \subset \ldots$ For, if $x \in L_p$, then $\sum_{n=1}^{\infty} | x^n |^p < +\infty$, so that $| x^n | \leqq 1$ for all n from a certain number n_0 onwards, whence

$$\sum_{n=n_0}^{\infty} | x^n |^{p+1} \leqq \sum_{n=n_0}^{\infty} | x^n |^p < +\infty.$$

For two points x and y of L_p, the function d_p defined by

$$d_p(x, y) = \sqrt[p]{\sum_{n=1}^{\infty} | x^n - y^n |^p}$$

is a metric, for

$$\sum_{n=1}^{\infty} | x^n - y^n |^p \leqq 2^p \sum_{n=1}^{\infty} (| x^n |^p + | y^n |^p) < +\infty.$$

The space L_2 is called the **real Hilbert space.**

§ 3. Sequences of elements

The ideas of limit point and cluster point for a sequence (x_n) were introduced in the preceding chapter. Since a metric space is a separated space, a sequence in such a space cannot have more than one limit point.

All the results established for filtered families can be applied to sequences when X is a metric space. The proofs are the same if we replace $\mathscr{V}(x)$ by the denumerable base

$$\mathscr{N}(x) = (B_1(x), B_{\frac{1}{2}}(x), B_{\frac{1}{3}}(x), \ldots)$$

We now state, without proofs, the results of § 4, Chapter IV, modified to the case of sequences in metric spaces.

Theorem 1. *Let (x_{k_n}) be a sub-sequence of (x_n). Then every cluster point of (x_{k_n}) is a cluster point of (x_n).*

Theorem 2. *A point $a \in \bar{A}$ if and only if there exists a sequence (x_n) in A converging to a.*

COROLLARY 1. *A point a is a point of accumulation of A if and only if there exists a sequence in $A-\{a\}$ converging to a.*

COROLLARY 2. *A set F is closed if and only if*

$$x_n \in F,\ (x_n) \to a \quad \text{imply that} \quad a \in F.$$

Theorem 3. *A point a is a cluster point of a sequence (x_n) if and only if there exists a sub-sequence (x_{k_n}) converging to a.*

Theorem 4. *If σ is a single-valued mapping of a metric space X into a metric space Y, then σ is continuous if and only if*

$$(x_n) \to x_0 \quad \text{implies that} \quad (\sigma x_n) \to \sigma x_0.$$

Theorem 5. *In a metric space X, the following properties hold:*

(1) $(x_n) \to x_0$ *implies that* $(x_{k_n}) \to x_0$,
(2) $(\forall n) : x_n = x_0$ *implies that* $(x_n) \to x_0$,
(3) *if every sub-sequence of (x_n) has x_0 as a cluster point then*
$$(x_n) \to x_0,$$
(4) *if*
$$(x_n^i) \to x_0^i,$$
$$(x_0^n) \to x_0,$$
then the set $\{x_k^i \,/\, i \geqq 1,\ k \geqq 1\}$ contains a sequence which converges to x_0.

In a metric space X, we say that a sequence (x_n) is **Cauchy-convergent** or that it is a **Cauchy sequence**, if to each $\varepsilon > 0$ there corresponds an integer m such that

$$n \geqq m,\ p \geqq m \quad \Rightarrow \quad d(x_n, x_p) \leqq \varepsilon.$$

Theorem 6. *A sub-sequence (x_{k_n}) of a Cauchy-convergent sequence is Cauchy-convergent.*

Proof. This result is immediate, since $k_n \geqq m$ and $k_p \geqq m$ imply that $d(x_{k_n}, x_{k_p}) \leqq \varepsilon$.

Theorem 7. *Every convergent sequence is Cauchy-convergent.*

Proof. Suppose that $(x_n) \to x_0$, and let m be such that

$$n \geqq m \quad \Rightarrow \quad d(x_n, x_0) \leqq \frac{\varepsilon}{2}.$$

Then if $n \geqq m$ and $p \geqq m$, we have

$$d(x_n, x_p) \leqq d(x_n, x_0) + d(x_p, x_0) \leqq \varepsilon.$$

Theorem 8. *If a Cauchy-convergent sequence (x_n) contains a sub-sequence (x_{k_n}) converging to x_0, then $(x_n) \to x_0$.*

Proof. Let ε be a strictly positive number; then there exists an integer m such that

$$n \geqq m, p \geqq m \quad \Rightarrow \quad d(x_n, x_p) \leqq \frac{1}{2}\varepsilon$$

and an integer m' such that

$$n \geqq m' \quad \Rightarrow \quad d(x_{k_n}, x_0) \leqq \frac{1}{2}\varepsilon.$$

Let l be an index for the sub-sequence such that $k_l \geqq m$ and $l \geqq m'$; then if $n \geqq l$, we have

$$d(x_n, x_0) \leqq d(x_n, x_{k_n}) + d(x_{k_n}, x_0) \leqq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

and therefore $(x_n) \to x_0$.

Theorem 9. *A metric space X is compact if and only if every sequence has a cluster point.*

Proof. If X is compact, then every sequence has a cluster point by Theorem 6 of § 6, Chapter IV.

Suppose now that X is a metric space such that every sequence has a cluster point. If X is not compact, there exists an open covering $(G_i \,/\, i \in I)$ of X which does not contain a finite sub-covering. Write

$$\lambda_i(x) = d(x, -G_i)$$

and

$$\lambda(x) = \sup_{i \in I} \lambda_i(x).$$

We first show that the mapping λ is continuous. For each index i,

$$\lambda_i(x) \leqq d(x, x') + \lambda_i(x') \leqq d(x, x') + \lambda(x')$$

and therefore

$$\lambda(x) - \lambda(x') \leqq d(x, x').$$

Hence, by symmetry, we have

$$| \lambda(x) - \lambda(x') | \leqq d(x, x')$$

and so λ is a continuous numerical function in X.

Write $\mu = \inf_{x \in X} \lambda(x)$. If $\mu = 0$, there exists a sequence (x_n) such that $(\lambda(x_n)) \to 0$. Let x be a cluster point of this sequence. Then $\lambda(x) = 0$, which implies that x is in none of the sets G_i, contrary to the hypothesis that these sets form a covering. Therefore $\mu > 0$.

Let x_1 be a point of X. Then there exists an index i_1 such that $\lambda_{i_1}(x_1) > \mu / 2$. Since G_{i_1} does not cover X, there exists a point $x_2 \notin G_{i_1}$;

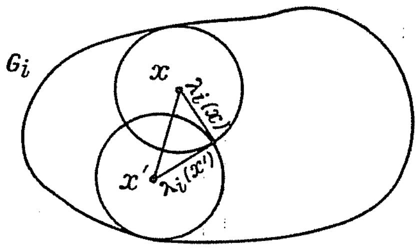


FIG. 23

corresponding to x_2 there exists an index i_2 such that $\lambda_{i_2}(x_2) > \mu / 2$. Since $G_{i_1} \cup G_{i_2}$ does not cover X, there exists a point $x_3 \notin G_{i_1} \cup G_{i_2}$; corresponding to x_3 there is an integer i_3 such that $\lambda_{i_3}(x_3) > \mu / 2$ and so on. The sequence (x_n) so formed has no cluster points, since if (x_{k_n}) is a sub-sequence, then

$$m \neq n \quad \Rightarrow \quad d(x_{k_m}, x_{k_n}) > \frac{\mu}{2}.$$

Thus we have a contradiction and so the hypothesis that X is not compact is false.

COROLLARY 1. *A necessary and sufficient condition for a metric space to be compact is that every infinite subset admits at least one point of accumulation.*

Proof. The condition is necessary by the Bolzano-Weierstrass theorem of § 6, Chapter IV.

To prove the sufficiency, suppose that X is a metric space such that every infinite subset admits a point of accumulation. In order to prove that X is compact, it is sufficient to show that any sequence (x_n) admits a cluster point. If infinitely many of the x_n coincide with a point a, then the sequence admits a as a cluster point and our aim is achieved. If not, the set $A = \{x_1, x_2, \ldots\}$ is infinite and so, by hypothesis, it admits a point of accumulation x_0. Let k_1 be an index such that $d(x_{k_1}, x_0) < 1$; let k_2 be an index such that $k_2 > k_1$ and $d(x_{k_2}, x_0) < \frac{1}{2}$; let k_3 be an index such that $k_3 > k_2$ and $d(x_{k_3}, x_0) < \frac{1}{3}$ and so on. Then $(x_{k_n}) \to x_0$ and so the sequence (x_n) has x_0 as a cluster point.

COROLLARY 2. *A necessary and sufficient condition for a sequence (x_n) in a compact metric space to be convergent is that the set of cluster points of the sequence reduces to a single element.*

This follows from the Corollary to Theorem 6 of § 6, Chapter IV.

§ 4. Totally bounded spaces and complete spaces

If X is a metric space such that every sequence (x_n) contains a Cauchy-convergent sub-sequence, then we say that X is **totally bounded.** A metric space X for which every Cauchy-convergent sequence is convergent is said to be **complete.**

EXAMPLE 1. **R** is complete.

If (x_n) is Cauchy-convergent, then to each $\varepsilon > 0$ there corresponds an integer m such that

$$n \geqq m \quad \Rightarrow \quad x_n \in B_\varepsilon(x_m).$$

Since $B_\varepsilon(x_m)$ is compact, (x_n) contains a convergent sub-sequence and hence, by Theorem 8 of § 3, (x_n) is convergent.

On the other hand, **R** is not totally bounded, because the sequence (x_n) in which $x_n = n$ does not contain a Cauchy-convergent sub-sequence.

A fundamental property of totally bounded spaces is contained in the following theorem.

Theorem 1. *Let X be a totally bounded space. Then, to each $\varepsilon > 0$, there corresponds a finite set $A \subset X$ such that*

$$(\forall x)\,(\exists_A a) : d(x, a) \leqq \varepsilon.$$

Proof. Suppose that there is an $\varepsilon > 0$ such that, for each finite set A,

$$(\exists x)\,(\forall_A a) : d(x, a) > \varepsilon.$$

Let x_0 be any point of X. Then, taking $A = \{x_0\}$, there exists a point x_1 such that

$$d(x_0, x_1) > \varepsilon.$$

Taking $A = \{x_0, x_1\}$, we see that there exists a point x_2 such that

$$d(x_0, x_2) > \varepsilon, \quad d(x_1, x_2) > \varepsilon.$$

Continuing in this way, we can define a sequence (x_n) which has no Cauchy-convergent sub-sequence. Therefore X is not totally bounded.

COROLLARY. *A totally bounded space is bounded.*

Proof. Taking $\varepsilon = 1$ in the theorem, let $A = \{a_1, a_2, a_3, \ldots, a_n\}$ be a finite set such that

$$(\forall x)\,(\exists_A a_k) : d(x, a_k) \leqq 1.$$

$\Rightarrow X \subseteq B(a_0, \delta)$

Then, if a_0 is any fixed point of X, we have

$$d(x, a_0) \leqq d(x, a_k) + d(a_k, a_0) \leqq 1 + \underbrace{\max \{d(a_i, a_0) \,/\, i = 1, 2, \ldots n\}}_{= \delta}$$

for all X, and so X is bounded.

Theorem 2. *A metric space X is compact if and only if it is complete and totally bounded.*

Proof. If X is compact, any sequence (x_n) contains a sub-sequence which is convergent and so Cauchy-convergent. Therefore X is totally bounded. Moreover, by the same property, namely that any sequence contains a convergent sub-sequence, every Cauchy-sequence is convergent (see Theorem 8, § 3); therefore X is complete.

Conversely, let X be complete and totally bounded. Then any sequence (x_n) contains a sub-sequence which is Cauchy-convergent and therefore convergent; hence X is compact.

Theorem 3. *If $(X_1, d_1), (X_2, d_2), \ldots, (X_k, d_k)$ are complete metric spaces, then the product space $\prod_{i=1}^{k} X_i$ is complete.*

Proof. The distance between

$$x = (x^1, x^2, \ldots, x^k) \text{ and } y = (y^1, y^2, \ldots, y^k)$$

in $X = \prod X_i$ is defined to be

$$d(x, y) = \max_i d_i(x^i, y^i).$$

If (x_n) is a Cauchy-convergent sequence in X, then there exists m such that for all $n \geqq m$,

$$d_i(x_n^i, x_{n+p}^i) \leqq d(x_n, x_{n+p}) \leqq \varepsilon.$$

Hence (x_n^i) is Cauchy-convergent, and therefore convergent, since X_i is complete. If x_0^i is the limit point of this sequence, we have

$$n \geqq m \quad \Rightarrow \quad d_i(x_n^i, x_0^i) \leqq \varepsilon.$$

Writing $x_0 = (x_0^1, x_0^2, \ldots, x_0^k)$, we have

$$n \geqq m \quad \Rightarrow \quad d(x_n, x_0) = \max_i d_i(x_n^i, x_0^i) \leqq \varepsilon$$

and therefore $(x_n) \to x_0$.

Complete subsets possess certain properties analogous to those of compact subsets; in particular, we have the following theorems.

Theorem 4. *If F is a closed set contained in a complete space X, then F is complete.*

Proof. If (x_n) is a Cauchy-convergent sequence of elements of F, it converges to a point $a \in X$, since the space X is complete. Since F is closed, we have $a \in F$ and so F is complete.

Theorem 5. *A complete subset A of X is closed in X.*

Proof. If $a \in \bar{A}$, there exists a sequence (x_n) in A converging to a. Since A is complete, we have $a \in A$ and so A is closed.

Theorem 6. *If $A_1, A_2, \ldots, A_k$ are complete sets, so is their union $A = \bigcup_{i=1}^{k} A_i$.*

Proof. If (x_n) is a Cauchy sequence in A, there exists an infinity of indices n such that

$$x_n \in A_{i_0}$$

for a suitably chosen index i_0. Therefore the Cauchy-convergent sequence (x_n) contains a convergent sub-sequence and so, by Theorem 8 of § 3, it converges to an element $a \in A_{i_0} \subset A$. Hence the set A is complete.

Theorem 7. *If $(A_i \mid i \in I)$ is a family of complete sets, the intersection $A = \bigcap_{i \in I} A_i$ is complete.*

Proof. By Theorem 5, the set A is closed; since A is closed and contained in a complete set A_{i_0}, the set A is complete by Theorem 4.

§ 5. Separable sets

A subset B of a space X is said to be **dense** in X if $\bar{B} = X$. A metric space X is said to be **separable** if it contains a countable dense subset B; we then have

$$(\forall x)\,(\forall \varepsilon)\,(\exists_B b) : d(x, b) \leqq \varepsilon.$$

EXAMPLE 1. $\mathbf{R}$ is separable, for the set $\mathbf{R}_r$ of positive or negative rational numbers is dense and countable.

EXAMPLE 2. $\mathbf{R}^n$ is separable, for the points whose coordinates are rational numbers form a countable dense subset.

EXAMPLE 3. Hilbert space L_2 (cf. § 2) is separable, for the points which have a finite number of their coordinates rational numbers and the remainder zero, form a countable dense subset.

EXAMPLE 4. The metric space consisting of $\mathbf{R}$ together with the metric d defined by $d(x, y) = 0$ if $x = y$ and $d(x, y) = 1$ if $x \neq y$ is not separable.

A **fundamental family** for a topological space X is a family of open sets $\mathscr{A} = (A_i \mid i \in I)$ such that, given any non-empty open set G there exists a subset J of I for which

$$G = \bigcup_{i \in J} A_i.$$

Theorem 1. *A metric space is separable if and only if it possesses a countable fundamental family.*

Proof. Let X be a separable metric space and let B be a countable dense subset. Then, if $\mathring{B}_\lambda(b)$ is the open ball of radius λ and centre b, the family

$$\mathscr{A} = (\mathring{B}_\lambda(b) \mid b \in B,\ \lambda \in \mathbf{R}_r^+)$$

is countable. If G is an open set and $x \in G$, there exists a rational number λ such that

$$\mathring{B}_\lambda(x) \subset G,$$

and an element $b \in B$ such that $d(b, x) < \lambda / 2$. Then

$$\mathring{B}_{\frac{\lambda}{2}}(b) \subset \mathring{B}_\lambda(x) \subset G$$

and so G is a union of open sets in the family $\mathscr{A}$, whence $\mathscr{A}$ is a fundamental family.

Conversely, suppose that X possesses a countable fundamental family $\mathscr{A} = (A_1, A_2, \ldots)$. Let a_i be an element of A_i. The set $B = \{a_1, a_2, \ldots\}$ is dense, for if $\bar{B} \neq X$, then the set $X - \bar{B}$ is a non-empty open set, which therefore contains a set A_i and so contains an element a_i, contrary to hypothesis.

Theorem 2 (Lindelöf). *If X is a separable metric space, then any open covering $\mathscr{A} = (G_k \mid k \in K)$ of X contains a countable covering.*

Proof. Let $(A_i \mid i \in \mathrm{N}_0)$, where N_0 is a subset of the set N of positive integers, be a countable fundamental family. Then the set

$$I = \{i \mid i \in \mathrm{N}_0;\ (\exists_K k) : A_i \subset G_k\}$$

is countable. The family $(A_i \mid i \in I)$ is a covering of X, since

$$x \in X \Rightarrow (\exists k) : x \in G_k \Rightarrow (\exists_I i) : x \in A_i.$$

For each $i \in I$, let k_i be an index such that $G_{k_i} \supset A_i$. The family $(G_{k_i} \mid i \in I)$ is then a countable covering of X.

Theorem 3. *The cardinal number of a separable metric space is at most that of the continuum.*

Proof. Let B be a countable subset such that $\bar{B} = X$. If $x \in X$ and $x \notin B$, there exists an element b_1 of B such that $d(x, b_1) \leqq 1$, an element b_2 of B such that $d(x, b_2) \leqq \frac{1}{2}$, an element b_3 of B such that $d(x, b_3) \leqq \frac{1}{3}$ and so on. The set $A = \{b_1, b_2, b_3, \ldots\}$ has the point x as its only point of accumulation. Thus we can set up a one-one correspondence between X and a subset of $\mathscr{P}(B)$. By Cantor's Theorem (§ 3, Chapter III) $\mathscr{P}(B)$ has the power of the continuum (unless B is finite, in which case $\mathscr{P}(B)$ is also finite). Therefore the cardinal number of X is at most that of the continuum.

Theorem 4. *A totally bounded space X is separable.*

Proof. Putting $\varepsilon = 1/n$ in Theorem 1 of § 4, we see that there is a finite set $A_{1/n}$ such that

$$(\forall x)\,(\exists a;\ a \in A_{1/n}) : d(x, a) \leqq \frac{1}{n}.$$

Then $B = \bigcup_{n=1}^{\infty} A_{1/n}$ is countable; moreover $\bar{B} = X$, for

$$(\forall x)\,(\forall 1/n)\,(\exists_{A_{1/n}} b) : d(x, b) \leqq \frac{1}{n}.$$

§ 6. Compact sets

Compact subsets and connected subsets of a metric space possess certain interesting topological and metrical properties. These are the subject of this section and the next.

Theorem 1. *A compact subset K of a metric space is closed and bounded.*

Proof. If K is compact, then, as we have already seen, it is closed. Suppose that K is not bounded. Let a be any point; then

$$(\forall n)\,(\exists_K x_n) : d(a, x_n) > n.$$

Since K is compact, the sequence (x_n) admits a cluster point x_0 in K; if (y_n) is a sub-sequence converging to x_0, there exists an integer n_0 such that $n > n_0$ implies that

$$d(a, y_n) \leqq d(a, x_0) + d(x_0, y_n) \leqq 2\, d(a, x_0).$$

Since $d(a, y_n)$ increases indefinitely with n, we have a contradiction.

COROLLARY. *In $\mathbf{R}^n$ a necessary and sufficient condition for a set A to be compact is that it is closed and bounded.*

Proof. By the preceding theorem, it is sufficient to show that if A is closed and bounded it is compact. Now in $\mathbf{R}$ there exist closed intervals $D_1, D_2, \ldots, D_n$ such that

$$A \subset D_1 \times D_2 \times \ldots \times D_n.$$

Since the D_i are compact sets in $\mathbf{R}$, their topological product is a compact subset of $\mathbf{R}^n$ (by Tychonoff's theorem). Thus the set A is closed and is contained in a compact set; therefore it is compact.

Lebesgue's theorem. *Let $(F_1, F_2, \ldots, F_m)$ be a finite closed covering of a compact metric space X. Then there exists a number ε such that for each set A of diameter $\delta(A) \leqq \varepsilon$ the intersection of the sets F_i meeting A is not empty.*

Proof. Suppose that no such number ε exists. Then, for any $k \in \mathbf{N}$, there is a set A_k such that

$$(1) \qquad \delta(A_k) \leqq \frac{1}{k},$$

$$(2) \qquad I_k = \{i \,/\, F_i \cap A_k \neq \varnothing\} \Rightarrow \bigcap_{i \in I_k} F_i = \varnothing.$$

Since there is only a finite number of the sets F_i, there exists a set I_0 such that $I_k = I_0$ for infinitely many values of k. Let $k_1, k_2, \ldots$ be these values of k, with $k_1 < k_2 < \ldots$ Let a_{k_n} be an element of A_{k_n}. Since X is compact, the sequence (a_{k_n}) admits a cluster point a_0 in X; there is therefore a sub-sequence (a_{h_n}) converging to a_0. Any ball $B_\lambda(a_0)$ contains an A_{h_n}; it is sufficient to choose n to be sufficiently large for the conditions

$$a_{h_n} \in B_{\frac{\lambda}{2}}(a_0); \quad \delta(A_{h_n}) \leqq \frac{1}{h_n} \leqq \frac{\lambda}{2}$$

to be satisfied. Then, if $i \in I_0$, the set F_i meets all the $B_\lambda(a_0)$ and so

$$a_0 \in \overline{F}_i = F_i.$$

This is true for all i in I_0 and therefore

$$\bigcap_{i \in I_0} F_i \neq \emptyset,$$

which is contrary to hypothesis.

Another result, which we shall use later, is the following.

Theorem 2. *Let $K_1, K_2, \ldots$ be a decreasing sequence of compact sets and let K be their intersection. Then, for each $\varepsilon > 0$, there exists an integer n_0 such that*

$$n \geqq n_0 \Rightarrow K_n \subset B_\varepsilon(K) = \bigcup_{x \in K} B_\varepsilon(x).$$

Proof. Suppose that the theorem is false. Then there exist indices n_k such that

$$n_1 < n_2 < \ldots < n_k < \ldots$$

and points x_{n_k} such that

$$x_{n_k} \in K_{n_k}; \; d(x_{n_k}, K) > \varepsilon.$$

The sequence formed by the x_{n_k} contains a sequence (y_n) converging to a point x_0, which belongs to all the K_i and so

$$x_0 \in K = \bigcap_{i=1}^{\infty} K_i.$$

On the other hand, because $d(x, K)$ is continuous, we have

$$d(x_0, K) \geqq \varepsilon$$

and so we have a contradiction.

§ 7. Connected sets[1]

A finite family of elements $(a_1, a_2, \ldots, a_m)$ is called an **ε-chain** if $d(a_1, a_2) \leqq \varepsilon$, $d(a_2, a_3) \leqq \varepsilon, \ldots, d(a_{m-1}, a_m) \leqq \varepsilon$. Two points a and b in a set A are said to be **ε-connected** in A if there exists an ε-chain $(a_1, a_2, \ldots, a_m)$ contained in A and such that $a_1 = a$, $a_m = b$. If a and b are ε-connected for every $\varepsilon > 0$, then we say that they are **well-chained.** These concepts enable us to make a considerable simplification in the study of compact connected sets in a metric space.

Theorem 1. *If A is a connected set having at least two distinct points a and b, then the cardinal number of A is at least $\aleph_1$.*

Proof. If A is connected and B is any set, then

$$\left.\begin{array}{l} A \cap B \neq \emptyset \\ A \cap (-B) \neq \emptyset \end{array}\right\} \text{ imply that } A \cap (\operatorname{Fr} B) \neq \emptyset,$$

[1] We have followed M. H. A. Newman (*Topology of plane sets of points*) concerning questions relating to connectivity.

for otherwise we should have

$$\begin{aligned} &A \subset \mathring{B} \cup (-\bar{B}), \\ &A \cap \mathring{B} \neq \varnothing, \\ &A \cap (-\bar{B}) \neq \varnothing, \end{aligned}$$

which contradict the fact that A is connected.

Let a and b be two distinct points of A and let ε be such that $\varepsilon < d(a, b)$. Then $A \cap \mathrm{Fr}\,[B_\varepsilon(a)] \neq \varnothing$ and so there exists a point x_ε such that

$$x_\varepsilon \in A, \ d(x_\varepsilon, a) = \varepsilon.$$

Then we have a one-one correspondence $\varepsilon \to x_\varepsilon$ between $]0, d(a, b)[$ and a subset of A and so the cardinal number of A is at least $\aleph_1$.

LEMMA 1. *In a metric space X, the set $E_\varepsilon(a)$ of points which can be linked to a by means of an ε-chain is both open and closed.*

Proof. $E_\varepsilon(a)$ is open, since

$$\begin{cases} x_0 \in E_\varepsilon(a) \\ d(x, x_0) < \varepsilon \end{cases} \text{ imply that } \quad x \in E_\varepsilon(a)$$

and $E_\varepsilon(a)$ is closed, since

$$\begin{cases} x_0 \notin E_\varepsilon(a) \\ d(x, x_0) < \varepsilon \end{cases} \text{ imply that } \quad x \notin E_\varepsilon(a).$$

LEMMA 2. *Let $(K_1, K_2, \ldots)$ be a decreasing sequence of compact sets:*

$$K_1 \supset K_2 \supset K_3 \supset \ldots$$

If each pair of points in K_i is ε-connected in K_i, then each pair in $K = \bigcap_i K_i$ is (2ε)-connected in K.

Proof. By Theorem 2 of § 6, there exists an integer n_0 such

$$n \geqq n_0 \Rightarrow K_n \subset B_{\frac{\varepsilon}{2}}(K) = \bigcup_{x \in K} B_{\frac{\varepsilon}{2}}(x).$$

If a and b are two points in K, then they are ε-connected in K_n by an ε-chain

$$(a_1 = a, a_2, \ldots a_m = b).$$

Choose $n \geqq n_0$ and let $x_i (1 < i < m)$ be a point in the (non-empty) set $B_{\varepsilon/2}(a_i) \cap K$. Then

$$(a, x_2, x_3, \ldots x_{m-1}, b)$$

is a (2ε)-chain in K, since

$$d(x_i, x_{i+1}) \leqq d(x_i, a_i) + d(a_i, a_{i+1}) + d(a_{i+1}, x_{i+1}) \leqq \frac{\varepsilon}{2} + \varepsilon + \frac{\varepsilon}{2}.$$

Theorem 2. *A necessary and sufficient condition for a compact metric space X to be connected is that any two of its points are well-chained.*

Proof. Suppose that X is connected. If $a \in X$, then $E_\varepsilon(a)$ is non-empty (since it contains a) and is both open and closed by Lemma 1; hence $E_\varepsilon(a) = X$. This holds for each strictly positive ε and so any two points in X are well-chained.

If X is not connected, then there exist open sets G_1 and G_2 such that

$$\begin{cases} X = G_1 \cup G_2, \\ G_1 \cap G_2 = \emptyset, \\ G_1 \neq \emptyset, \\ G_2 \neq \emptyset. \end{cases}$$

If X is compact, the sets $(-G_1)$ and $(-G_2)$ are compact; since they are also non-empty and disjoint, we have

$$d(-G_1, -G_2) = \min\{d(x, y) \,/\, x \in -G_1, y \in -G_2\} = \varepsilon > 0.$$

Therefore a point of G_1 cannot be linked to a point of G_2 by means of an $(\varepsilon/2)$-chain and so not every pair of points in X is well-chained.

Theorem 3. *The intersection of a decreasing sequence $(K_1, K_2, \ldots)$ of compact connected sets is a compact connected set.*

Proof. Suppose that none of the sets K_i is empty (otherwise the theorem is trivial). Then $K = \bigcap_i K_i$ is a non-empty compact set, by the finite intersection axiom. By Lemma 2, any two points in K can be linked by a (2ε)-chain in K, where ε is an arbitrary strictly positive number. Therefore K is connected.

Let K be a compact set and let a be a point of K. The **connected component** in K of a is the union of all the connected subsets of K which contain a. This union is non-empty, for $\{a\}$ is connected and is contained in K. The connected component of a is denoted by $C_K(a)$. By Theorem 1 of § 7, Chapter IV, $C_K(a)$ is connected, since it is the union of a family of connected sets whose intersection is non-empty.

The set $C_K(a)$ is closed, since $\overline{C_K(a)}$ is a connected set (by Theorem 2, § 7, Chapter IV) contained in K, and so $\overline{C_K(a)} \subset C_K(a)$, whence $\overline{C_K(a)} = C_K(a)$. Hence $C_K(a)$ is a compact set, since it is a closed subset of a compact set.

We now show that the different sets $C_K(x)$ form a partition of K. If $x \in K$, $y \in K$ and $C_K(x) \cap C_K(y) \neq \emptyset$, the set $E = C_K(x) \cup C_K(y)$ is connected and is contained in K. Therefore $E \subset C_K(x)$, whence $C_K(x) = E$. By symmetry we have

$$C_K(y) = E = C_K(x)$$

and the result follows.

Theorem 4. *Let a and b be two points in a compact set K. Then $b \in C_K(a)$ if and only if a and b are well-chained in K.*

Proof. If $b \in C_K(a)$, then a and b are well-chained in $C_K(a)$, because this set is compact and connected, and therefore are well-chained in K.

Suppose conversely that a and b are well-chained in K. By Lemma 1, the set $E_{1/n}(a)$ of points of K which can be linked to a by means of a $(1/n)$-chain in K is compact. Consider the sequence $E_{1/m}(a), E_{1/m+1}(a), E_{1/m+2}(a), \ldots$ of compact sets; this is a decreasing sequence and any two points of $E_{1/m+p}(a)$ can be linked by means of a $(1/m)$-chain. Therefore, by Lemma 2, any two points of the set

$$E = \bigcap_{n=1}^{\infty} E_{1/n}(a) = \bigcap_{n=m}^{\infty} E_{1/n}(a)$$

can be linked by means of a $(2/m)$-chain. This is true for all m and hence any two points of the compact set E are well-chained; therefore E is connected and so $E \subset C_K(a)$. Thus if a and b are well-chained in K, we have $b \in C_K(a)$.

§ 8.* Locally connected sets: curves

A metric space X is said to be **locally connected at** $a \in X$ if each neighbourhood $U(a)$ contains a connected neighbourhood $N(a)$. We say that X is **locally connected** if it is locally connected at each of its points; likewise a set A is locally connected if for each point $a \in A$ there exists a neighbourhood $N(a)$ as small as we please such that $N(a) \cap A$ is connected.

EXAMPLE. In $\mathbf{R}$, the set $A = \{0, 1, 1/2, 1/3, \ldots\}$ is not locally connected because, for each neighbourhood $N(0)$, the set

$$N(0) \cap A = \left\{0, \frac{1}{n}, \frac{1}{n+1}, \frac{1}{n+2}, \ldots\right\}$$

is not connected.

Theorem 1. *A necessary and sufficient condition for a space X to be locally connected is that, for each open set G, the connected components $C_G(x)$ in G are open sets.*

Proof. The condition is sufficient, for if it is satisfied, then in each neighbourhood G of a, $C_G(a)$ is a connected neighbourhood.

To prove that the condition is necessary, let X be a locally connected space, let G be an open set of X and let a be a point of G. If $x \in C_G(a)$ and

$N(x)$ is a connected neighbourhood of x contained in G, then $N(x) \subset C_G(a)$. Hence $C_G(a)$ is a neighbourhood of each of its points and so is an open set.

Theorem 2. *A necessary and sufficient condition for a metric space X to be compact and locally connected is that for each strictly positive number ε there exists a finite covering by compact connected sets K_i such that $\delta(K_i) \leqq \varepsilon$.*

Proof. Let X be a compact and locally connected metric space. Given $\varepsilon > 0$, we can cover X with the connected components of the open balls $\mathring{B}_{\varepsilon/2}(x)$. By Theorem 1, these components are open sets. Since X is compact, the covering contains a finite covering $(G_1, G_2, \ldots, G_n)$; the sets

$$K_i = \overline{G}_i$$

cover X and K_i is a connected compact set such that $\delta(K_i) \leqq \varepsilon$.

Suppose now that X is a metric space such that for each strictly positive ε there exists a finite covering by compact connected sets

$$(K_1, K_2, \ldots K_n)$$

such that $\delta(K_i) \leqq \varepsilon$. Since

$$X = \bigcup_{i=1}^{n} K_i$$

it follows that X is compact. Let $a \in X$ and write

$$\delta_a \begin{cases} = \min\limits_i \{d(a, K_i) \,/\, K_i \not\ni a\} & \text{if} \quad a \notin \bigcap K_i, \\ = 1 & \text{if} \quad a \in \bigcap K_i. \end{cases}$$

Let K be the union of the sets K_i which contain the point a; if $\eta < \delta_a$, then $B_\eta(a) \subset K \subset B_\varepsilon(a)$ and so K is a connected neighbourhood of a contained in $B_\varepsilon(a)$. Therefore X is locally connected.

Theorem 3. *Let σ be a single-valued continuous mapping of a metric space X into a metric space Y; if A is a compact locally connected subset of X, then the image σA is a compact locally connected subset of Y.*

Proof. Given $\varepsilon > 0$ there exists $\eta > 0$ such that if $x, x_0 \in A$, then

$$d_X(x, x_0) \leqq \eta \text{ implies that } d_Y(\sigma x, \sigma x_0) \leqq \varepsilon$$

(see Heine's Theorem, § 9). If $(K_1, K_2, \ldots, K_n)$ is a family of compact connected sets forming a covering of A and such that

$$\delta(K_i) \leqq \eta,$$

then $(\sigma K_1, \sigma K_2, \ldots, \sigma K_n)$ is a family of compact connected sets covering σA and such that $\delta(\sigma K_i) \leqq \varepsilon$. It therefore follows from Theorem 2 that σA is compact and locally connected.

These results will now be applied to the special case of curves. A set C in a metric space is called a **curve** if there exists a single-valued continuous mapping f of $[0, 1]$ into X such that

$$C = f([0, 1]).$$

Such a function f is called a **parametrisation** of the curve C; the pair (C, f) is called a **parametrised curve.** We say that x is a **multiple point** of a parametrised curve (C, f) if we have (as in figure 24)

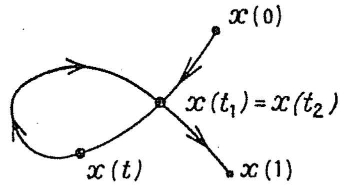


FIG. 24

$$x = f(t_1) = f(t_2); \quad t_1 \neq t_2.$$

Let g be a continuous numerical function. Then the representative curves $C_g = \{(x, g(x) \mid x \in [0, 1]\}$ are very simple examples of curves in the plane with the function f defined by $f(t) = (t, g(t))$ as parametrisation.

EXAMPLE (Hilbert). With a suitable parametrisation, the unit square can be regarded as a curve in the above sense. Let X, Y be $[0, 1]$ and $[0, 1]^2$

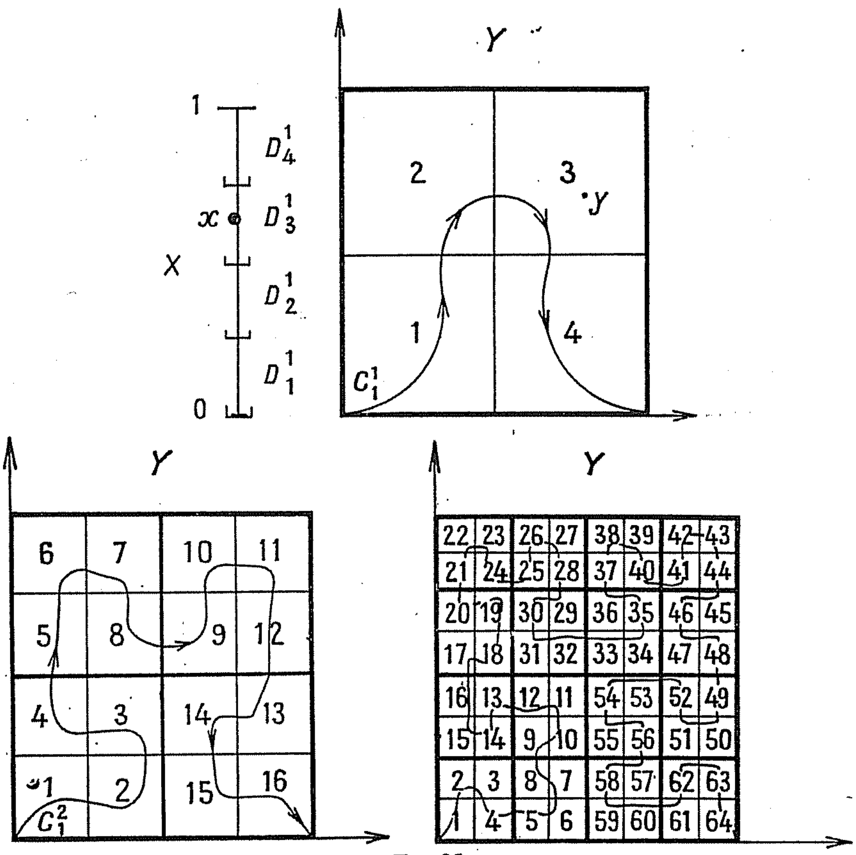


FIG. 25

respectively. Subdivide the unit square Y into four equal squares $C_1^1, C_2^1, C_3^1, C_4^1$, and the interval X into four intervals $D_1^1, D_2^1, D_3^1, D_4^1$ of equal lengths, as shown in figure 25, with D_1^1, D_2^1, D_3^1 semi-closed and D_4^1 closed. Now subdivide each square C_i^1 into four equal squares in a similar manner, and each interval D_i^1 into four intervals; we then obtain squares C_j^2 and intervals D_j^2 which correspond as in figure 25.

We continue this operation of subdividing the squares and the intervals. With each $x \in X$ we associate a point $y \in Y$ determined as follows: if x belongs to the intervals $D_{k_1}^1, D_{k_2}^2, D_{k_3}^3, \ldots$, we consider the corresponding squares $C_{k_1}^1, C_{k_2}^2, C_{k_3}^3, \ldots$; these have at least one common point y, for Y is compact and so the finite intersection axiom holds. Moreover y is unique, for the diameter $\delta(C_{k_n}^n)$ tends to zero as n tends to infinity. Thus we can define a single-valued mapping σ by writing $y = \sigma x$.

If $| x - x_0 | < \frac{1}{4^n}$, the points x and x_0 are in the same interval D^n or in adjacent intervals D^n; then the corresponding points y and y_0 are in the same square C^n or in the adjacent squares C^n, so that

$$| y - y_0 | < 3 \times \frac{1}{2^n};$$

and hence the mapping σ of X onto Y is continuous.

If f is a homeomorphism, the curve $C = f([0, 1])$ is called a **simple curve**; in the above example the unit square is not a simple curve. An immediate consequence of the definition is that a simple curve cannot have multiple points.

A curve C is called a **simple closed curve** if it is the image under a homeomorphism of a circle in the plane. For example, the boundary of a square or an ellipse is a simple closed curve (see figure 26).

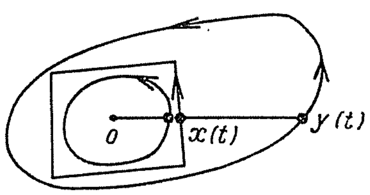


FIG. 26

Let $\tau = (0, t_1, t_2, \ldots, 1)$ be a finite sequence in $[0, 1]$ such that

$$0 < t_1 < t_2 < \ldots < 1.$$

The *length* of the curve C is defined to be

$$l(C) = \sup_{\tau} \sum_i d(f(t_i), f(t_{i+1})).$$

If $l(C) < +\infty$, we say that C is a **rectifiable** curve.

EXAMPLE. Let g be the numerical function in $[0, 1]$ defined by

$$g(t) \begin{cases} = t \sin 1/t & \text{if } t \neq 0, \\ = 0 & \text{if } t = 0. \end{cases}$$

The representative curve $C_g = \{(t, g(t)) \,/\, t \in [0, 1]\}$ is a curve in the plane $\mathbf{R}^2$ having no multiple points (see figure 27). It is not rectifiable, since

$$\sum_{k=1}^{n} d\left(f\left(\frac{1}{k\pi+\frac{\pi}{2}}\right), f\left(\frac{1}{k\pi}\right)\right) > \sum_{k=1}^{n} \frac{1}{k\pi+\frac{\pi}{2}},$$

which tends to infinity with n.

The considerations of the preceding paragraphs facilitate the study of topological properties of plane curves. We notice first of all that $C = f([0, 1])$ is a compact connected set in the plane, since $[0, 1]$ is a compact connected set and f is continuous. One consequence of this is that for any given $\varepsilon > 0$, two points of C can be linked by an ε-chain in C. Moreover C is a compact locally connected set, since $[0, 1]$ is compact and locally connected. We can, in fact, prove that the following result holds:

Hahn-Mazurkiewicz theorem. *A necessary and sufficient condition for a subset C of a metric space to be a curve is that it is compact, connected and locally connected.*

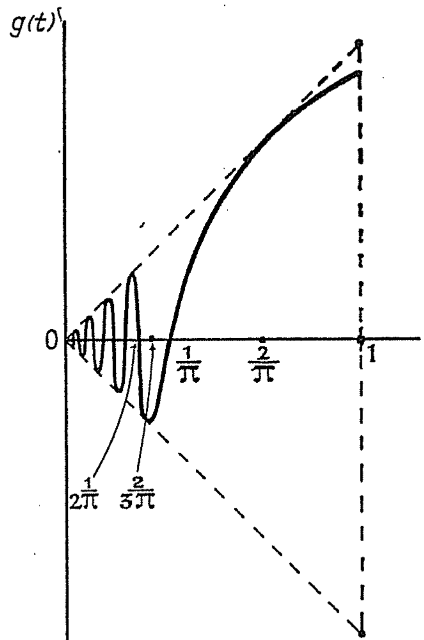


FIG. 27

We quote also the following theorem:

Jordan's theorem. *If C is a simple closed curve in the plane, its complement $-C$ is not connected and is the union of two disjoint open connected sets G_1 and G_2 whose frontiers satisfy*

$$\mathrm{Fr}\, G_1 = \mathrm{Fr}\, G_2 = C.$$

We refer the reader to works dealing specifically with the topology of plane sets for a proof of this theorem.

§ 9. Single-valued mappings of one metric space into another

Let σ be a single-valued mapping of a metric space (X, d_X) into a metric space (Y, d_Y). Then σ is continuous if

$$(\forall_X x_0)\,(\forall \varepsilon)\,(\exists \eta) : d_X(x_0, x) \leqq \eta \quad \Rightarrow \quad d_Y(\sigma x_0, \sigma x) \leqq \varepsilon$$

where the number η satisfying this condition depends in general on ε and x_0. If there exists a number η which satisfies this condition and depends on ε

but not on x_0, we say that σ is **uniformly continuous in** X; the property can be expressed as follows:

$$(\forall \varepsilon)\,(\exists \eta)\,(\forall_X x_0) : d_X(x_0, x) \leqq \eta \quad \Rightarrow \quad d_Y(\sigma x_0, \sigma x) \leqq \varepsilon.$$

Clearly a uniformly continuous mapping is continuous, but a continuous mapping need not be uniformly continuous.

Let λ be a positive number. We say that σ is **λ-contracting** (*in* X) if

$$(\forall_X x_0)\,(\forall_X x) : d_Y(\sigma x_0, \sigma x) \leqq \lambda d_X(x_0, x).$$

Clearly a λ-contracting mapping is uniformly continuous.

EXAMPLE 1. Let $X = \mathbf{R}$. The numerical function f defined by $f(x) = x^2$ is 2-contracting in $[0, 1]$, but it is not λ-contracting in $\mathbf{R}$. It is continuous in $\mathbf{R}$, but not uniformly continuous.

EXAMPLE 2. If a mapping σ of $\mathbf{R}^2$ into $\mathbf{R}^2$ is a similitude of ratio λ, then σ is λ-contracting in $\mathbf{R}^2$.

EXAMPLE 3. In a metric space X, the numerical function f defined by

$$f(x) = d(x, a)$$

is uniformly continuous, because

$$d(x, x') \leqq \varepsilon \Rightarrow d(x, a) \leqq \varepsilon + d(x', a) \Rightarrow |f(x) - f(x')| \leqq \varepsilon.$$

Heine's theorem. *If X is a compact metric space, then a single-valued continuous mapping σ of X into Y is uniformly continuous.*

Proof. Let ε be a positive number and let $\eta(x_0)$ be a positive number such that

$$d_X(x, x_0) \leqq \eta(x_0) \quad \Rightarrow \quad d_Y(\sigma x_0, \sigma x) \leqq \varepsilon.$$

(If $\inf\limits_x \eta(x) > 0$, then uniform continuity is ensured; in particular, the theorem is trivial if X is finite).

Let $\mathring{B}(x)$ be the open ball of centre x and radius $\frac{1}{2}\eta(x)$. Since X is compact, it can be covered by a finite number of these open balls:

$$\mathring{B}(x_1), \mathring{B}(x_2), \ldots \mathring{B}(x_n).$$

Write $\eta = \min \{\eta(x_1), \eta(x_2), \ldots, \eta(x_n)\}$; we propose to prove that

$$d_X(x, x') \leqq \frac{\eta}{2} \quad \Rightarrow \quad d_Y(\sigma x, \sigma x') \leqq 2\varepsilon.$$

In fact, if $d_X(x, x') \leqq \frac{1}{2}\eta$, then, for some index k between 1 and n, we have

$$\begin{cases} d_X(x', x_k) \leqq \dfrac{\eta(x_k)}{2} \leqq \eta(x_k), \\[2ex] d_X(x, x_k) \leqq d_X(x, x') + d_X(x', x_k) \leqq \dfrac{\eta(x_k)}{2} + \dfrac{\eta(x_k)}{2} = \eta(x_k), \end{cases}$$

whence

$$\begin{cases} d_Y(\sigma x, \sigma x_k) \leqq \varepsilon, \\ d_Y(\sigma x', \sigma x_k) \leqq \varepsilon, \end{cases}$$

which imply that

$$d_Y(\sigma x, \sigma x') \leqq d_Y(\sigma x, \sigma x_k) + d_Y(\sigma x_k, \sigma x') \leqq 2\varepsilon$$

and it follows that σ is uniformly continuous.

Lipschitz's theorem. *If $\lambda < 1$ and σ is a single-valued λ-contracting mapping of a complete metric space X into itself, then there exists one and only one point a in X such that $\sigma a = a$. Moreover, for each point x_0, the sequence $(\sigma^n x_0)$ converges to a.*

Proof. Let x_0 be a point in X and write $x_n = \sigma^n x_0$; then

$$d(x_n, x_{n+1}) = d(\sigma x_{n-1}, \sigma x_n) \leqq \lambda d(x_{n-1}, x_n) \leqq \lambda^n d(x_0, \sigma x_0),$$

whence

$$d(x_n, x_{n+p}) \leqq d(x_0, \sigma x_0) \sum_{k=n}^{n+p} \lambda^k.$$

This shows that the sequence (x_n) is Cauchy-convergent. Since X is complete, the sequence converges to a point a and, since σ is continuous, $(x_{n+1}) = (\sigma x_n)$ converges to σa. But (x_n) converges to a and therefore we have $a = \sigma a$. Moreover a is the only solution of $x = \sigma x$, for if $a = \sigma a$ and $b = \sigma b$ are two distinct solutions, then $d(a, b) > 0$ and so

$$\lambda d(a, b) < d(a, b) = d(\sigma a, \sigma b) \leqq \lambda d(a, b),$$

which is a contradiction.

Let (σ_n) be a sequence of single-valued mappings of a metric space X into a metric space Y. Then we say that the sequence **converges simply** in $A \subset X$ to a mapping σ_0 if, for each $x \in A$ the sequence $(\sigma_n x)$ converges to $\sigma_0 x$. We also say that the sequence **converges uniformly** in A to σ_0 if, for each ε, there exists an integer m such that

$$n \geqq m \quad \Rightarrow \quad \sup_{x \in A} d_Y(\sigma_n x, \sigma_0 x) \leqq \varepsilon.$$

EXAMPLE. Consider the numerical functions f_n in $\mathbf{R}$ defined by

$$f_n(x) = \frac{nx}{1+nx}.$$

In the interval $]0, 1]$ these functions converge simply to a function f_0 such that $f_0(x) = 1$ for all $x \in]0, 1]$. They do not converge uniformly, because if $x \in]0, 1]$, then

$$f_0(x)-f_n(x) = 1-\frac{nx}{1+nx} = \frac{1}{1+nx}$$

and so, for any n,

$$\sup_{x\in]0,\,1]} |f_n(x)-f_0(x)| = 1.$$

The following important theorem describes circumstances in which a single-valued mapping, which is a limit of single-valued continuous mappings, is itself continuous.

Weierstrass' theorem. *If a sequence (σ_n) of single-valued continuous mappings converges uniformly to a mapping σ_0, then σ_0 is also continuous.*

Proof. Let ε be an arbitrary strictly positive number. Then there exists an integer m such that

$$\sup_x d_Y(\sigma_m x, \sigma_0 x) \leqq \varepsilon.$$

On the other hand, there exists a number $\eta_m(x_0)$ such that

$$d_X(x, x_0) \leqq \eta_m(x_0) \quad \Rightarrow \quad d_Y(\sigma_m x, \sigma_m x_0) \leqq \varepsilon.$$

Therefore $d_X(x, x_0) \leqq \eta_m(x_0)$ implies that

$$d_Y(\sigma_0 x, \sigma_0 x_0) \leqq d_Y(\sigma_0 x, \sigma_m x)+d_Y(\sigma_m x, \sigma_m x_0)+d_Y(\sigma_m x_0, \sigma_0 x_0) \leqq 3\,\varepsilon$$

and so σ_0 is continuous.

COROLLARY 1. *If a sequence (σ_n) of uniformly continuous mappings converges uniformly in X to σ_0, the mapping σ_0 is also uniformly continuous.*

Proof. It is sufficient to replace $\eta_m(x_0)$ in the above argument by a number η_m independent of x_0.

COROLLARY 2. *If a sequence (σ_n) of continuous mappings converges uniformly to σ_0 on every compact set K, then the mapping σ_0 is continuous.*

Proof. Let (x_n) be a sequence converging to x_0. On the compact set

$$K = \{x_0, x_1, x_2, \ldots, x_n, \ldots\}$$

the sequence (σ_n) converges uniformly to σ_0. Then, on K, the mapping σ_0 is continuous and so $(\sigma_0 x_n) \to \sigma_0 x_0$. It follows from Theorem 4 of § 3 that the mapping σ_0 is continuous in X.

Weierstrass' theorem has the following converse:

Dini's theorem. *Let (σ_n) be a sequence of continuous mappings converging simply to a continuous mapping σ_0. If X is compact and if, for all x, $\delta_n(x) = d_Y(\sigma_n x, \sigma_0 x)$ is decreasing as it tends to zero, then (σ_n) converges uniformly to σ_0.*

Proof. We observe first of all that δ_m is a continuous numerical function, for if $(x_n) \to x_0$ in X, then $(\sigma_m x_n, \sigma_0 x_n) \to (\sigma_m x_0, \sigma_0 x_0)$ in $Y \times Y$ and so, in $\mathbf{R}$, $\delta_m(x_n) \to \delta_m(x_0)$.

To prove the theorem, we suppose that (σ_n) does not converge uniformly to σ_0. Then

$$(\exists \varepsilon)\,(\forall n)\,(\exists k;\, k > n) : \sup_{x \in X} d_Y(\sigma_k x, \sigma_0 x) > \varepsilon.$$

In particular, taking $n = 1$, there is an index k_1 and a point x_1 such that

$$d_Y(\sigma_{k_1} x_1, \sigma_0 x_1) > \varepsilon;$$

then, taking $n = k_1$, there is an index $k_2 > k_1$ and a point x_2 such that

$$d_Y(\sigma_{k_2} x_2, \sigma_0 x_2) > \varepsilon;$$

then, taking $n = k_2$, there is an index $k_3 > k_2$ and a point x_3 such that

$$d_Y(\sigma_{k_3} x_3, \sigma_0 x_3) > \varepsilon;$$

and so on. Thus we obtain a sequence $(x_n) = (x_1, x_2, x_3, \ldots)$. Since X is compact, this sequence contains a sub-sequence converging to a point $a \in X$; let $y_n = x_{h_n}$ be the general term of this sub-sequence.

Let m be an arbitrary integer; then there exists an integer n_0 such that

$$n \geqq n_0 \quad \Rightarrow \quad \delta_m(a) \geqq \delta_m(y_n) - \frac{\varepsilon}{2}.$$

Moreover, there exists an integer n_1 such that $n \geqq n_1$ implies that $k_{n'} \geqq m$, where $n' = h_n$; therefore if n is an integer greater than $\max\{n_0, n_1\}$, we have

$$\delta_m(a) \geqq \delta_m(y_n) - \frac{\varepsilon}{2} \geqq \delta_{k_{n'}}(x_{n'}) - \frac{\varepsilon}{2} > \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}.$$

But this shows that $\delta_n(a)$ does not converge to zero, which is contrary to hypothesis. Thus the theorem is proved.

REMARK 1. If the mappings σ_n are numerical functions f_n, we have a convenient criterion for uniform convergence depending upon the theory of series. In fact we do not often know the simple limit f_0 and so cannot form $\delta_n(x) = |f_n(x) - f_0(x)|$. Instead, we make use of the following result:

If $\sup_x |f_n(x) - f_{n-1}(x)| = u_n$ *is the general term of a convergent series (that is, if* $\sum_1^\infty u_n < +\infty$*), the sequence* (f_n) *converges uniformly.*

Proof. In this case, the series

$$f_n(x) = f_1(x) + (f_2(x) - f_1(x)) + \ldots + (f_n(x) - f_{n-1}(x))$$

converges to the sum $f_0(x)$ and, for all x, we have

$$\delta_n(x) = |f_n(x)-f_0(x)| \leqq \sum_{k=n}^{\infty} |f_{k+1}(x)-f_k(x)| \leqq \sum_{k=n}^{\infty} u_n.$$

Given $\varepsilon > 0$, there then corresponds a number n_0 such that

$$n \geqq n_0 \quad \Rightarrow \quad \sup_x \delta_n(x) \leqq \varepsilon.$$

REMARK 2. The preceding criterion, although very convenient, does not give a necessary and sufficient condition for uniform convergence. Instead, we sometimes use the following result:

If the sequence whose general term is $\zeta_n = \sup_p \sup_x |f_{n+p}(x)-f_n(x)|$ *converges to zero, then the sequence* (f_n) *converges uniformly.*

Proof. If $(\zeta_n) \to 0$, the sequence $(f_n(x))$ is Cauchy-convergent in $\mathbf{R}$ for all x. Since $\mathbf{R}$ is complete, $(f_n(x)) \to f_0(x)$. Moreover, given $\varepsilon > 0$, there exists a number n_0 such that

$$n \geqq n_0 \quad \Rightarrow \quad |f_{n+p}(x)-f_n(x)| \leqq \varepsilon \quad \text{(for all } x \text{ and all } p\text{).}$$

Letting p tend to $+\infty$, we get

$$n \geqq n_0 \quad \Rightarrow \quad |f_0(x)-f_n(x)| \leqq \varepsilon \quad \text{(for all } x\text{)}$$

and so (f_n) converges uniformly to f_0.

We can easily verify that this sufficient condition is also necessary.

CHAPTER VI

MAPPINGS FROM ONE TOPOLOGICAL SPACE INTO ANOTHER

§1. Semi-continuous mappings

Let Γ be a mapping of a topological space X into a topological space Y and let x_0 be a point of X. We say that Γ is **lower semi-continuous**[1] **at** x_0 if for each open set G meeting Γx_0 there is a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad \Gamma x \cap G \neq \emptyset.$$

We say that Γ is **upper semi-continuous at** x_0 if for each open set G containing Γx_0 there exists a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad \Gamma x \subset G.$$

We say that the mapping Γ is **continuous at** x_0 if it is both lower and upper semi-continuous at x_0.

If Γ is a single-valued mapping, the definition given above for lower semi-continuity coincides with the ordinary definition of continuity; the same is true for upper semi-continuity.

We say that Γ is **lower semi-continuous** in X (and abbreviate this to l.s.c. in X) if it is lower semi-continuous at each point of X. We say that Γ is **upper semi-continuous** in X (and abbreviate this to u.s.c. in X) if it is upper semi-continuous at each point of X and if, also, Γx is a compact set for each x.

If Γ is both l.s.c. in X and u.s.c. in X, then it will be called **continuous** in X.

EXAMPLE 1. Let Γ be a mapping of X into Y such that Γx is a fixed compact set K_0 of Y for each x. Then Γ is l.s.c. and u.s.c. and so it is continuous.

Theorem 1. *A necessary and sufficient condition for Γ to be l.s.c. is that for each open set G in Y, the set $\Gamma^- G$ is open.*

Proof. Suppose that Γ is l.s.c. Clearly $\Gamma^- G$ is open if it is empty. We therefore suppose that $\Gamma^- G \neq \emptyset$. If $x_0 \in \Gamma^- G$, then $\Gamma x_0 \cap G \neq \emptyset$, and so there is a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad \Gamma x \cap G \neq \emptyset,$$

[1] The two kinds of semi-continuity of a multivalued function were introduced independently by Kuratowski (*Fund. Math.* **18**, 1932, p. 148) and Bouligand (*Ens. Math.*, 1932, p. 14). In general, the definitions given by different authors do not coincide whenever we deal with non-compact spaces (at least for upper semi-continuity; which is the more important from the point of view of applications). The definitions adopted here, which we have developed elsewhere (C. Berge, *Mém. Sc. Math.* 138), enable us to include the case when the image of a point x can be empty.

whence

$$U(x_0) \subset \Gamma^- G.$$

Therefore $\Gamma^- G$ is a neighbourhood of each of its points and so is open.

Suppose now that $\Gamma^- G$ is an open set for each open set G in Y. Let G be an open set meeting Γx_0; then $\Gamma^- G$ is an open neighbourhood of x_0 and we have

$$x \in \Gamma^- G \quad \Rightarrow \quad \Gamma x \cap G \neq \varnothing.$$

Therefore Γ is l.s.c.

Theorem 2. *A necessary and sufficient condition for Γ to be u.s.c. is that the set Γx is compact for each x and that for each open set G in Y the set $\Gamma^+ G$ is open*(1).

Proof. Suppose that Γ is u.s.c. If $\Gamma^+ G = \varnothing$ then it is open. Suppose therefore that $\Gamma^+ G \neq \varnothing$ and let $x_0 \in \Gamma^+ G$. There exists a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad \Gamma x \subset G.$$

Then $U(x_0) \subset \Gamma^+ G$ and so the set $\Gamma^+ G$ is a neighbourhood of each one of its points and therefore is an open set.

Suppose now that, for each open set G in Y, the set $\Gamma^+ G$ is open and that Γx is compact for each x. Let $x_0 \in X$ and let G be an open set containing Γx_0. Then $\Gamma^+ G$ is an open neighbourhood of x_0 and

$$x \in \Gamma^+ G \quad \Rightarrow \quad \Gamma x \subset G.$$

Therefore Γ is u.s.c.

Mappings of X into Y which are u.s.c. have the following fundamental property:

Theorem 3. *If Γ is u.s.c., the image ΓK of a compact subset K of X is also compact.*

Proof. Let $\{G_i \mid i \in I\}$ be an open covering of ΓK. If $x \in K$, the set Γx, which is compact, can be covered by a finite number of the G_i; let G_x denote the union of the sets in such a finite family. Then $(\Gamma^+ G_x \mid x \in K)$ is an open covering of K and so it contains a finite covering $\Gamma^+ G_{x_1}, \Gamma^+ G_{x_2}, \ldots, \Gamma^+ G_{x_n}$. The sets $G_{x_1}, G_{x_2}, \ldots G_{x_n}$, cover ΓK and so ΓK can be covered by a finite number of the G_i.

(1) (Translator's note.) In dealing with u.s.c. mappings, we define $\Gamma^+ G$ by

$$\Gamma^+ G = \{x / x \in X : \Gamma x \subset G\}.$$

This is a slightly different definition from that given in Chapter II. In particular,

$$\Gamma^+ \varnothing = \{x / x \in X : \Gamma x = \varnothing\}.$$

In addition to the two types of semi-continuity, it is sometimes convenient to consider a third topological property. We say that Γ is a **closed mapping** of X into Y if whenever $x_0 \in X$, $y_0 \in Y$, $y_0 \notin \Gamma x_0$ there exist two neighbourhoods $U(x_0)$ and $V(y_0)$ such that $x \in U(x_0) \Rightarrow \Gamma x \cap V(y_0) = \emptyset$.

Consider the graphical representation $\sum_{x \in X} \Gamma x$ of Γ in $X \times Y$; this is a closed set if and only if Γ is a closed mapping, for the above condition is equivalent to

$$U(x_0) \times V(y_0) \subset -\sum_{x \in X} \Gamma x.$$

We observe that an immediate consequence of the definition is that if Γ is a closed mapping then the set Γx is closed in Y.

EXAMPLE. If f is a continuous numerical function in $X \times Y$, the mapping defined by $\Gamma x = \{y \mid y \in Y, f(x, y) \leqq 0\}$ is a closed mapping of X into Y, for, in $X \times Y$, the graphical representation

$$\sum_{x \in X} \Gamma x = \{(x, y) \mid f(x, y) \leqq 0\}$$

is a closed set.

In particular, if λ is a continuous numerical function in a metric space (X, d) the mapping

$$\Gamma x = B_{\lambda(x)}(x) = \{y \mid y \in X, d(x, y) - \lambda(x) \leqq 0\}$$

is closed.

Theorem 4. *If Γ is a closed mapping, then*

$$\left.\begin{array}{l} (x_n) \to x_0 \\ (y_n) \to y_0 \\ (\forall n) : y_n \in \Gamma x_n \end{array}\right\} \Rightarrow y_0 \in \Gamma x_0.$$

Proof. The graphical representation of Γ is a closed set, $((x_n, y_n)) \to (x_0, y_0)$ and $(x_n, y_n) \in \sum_{x \in X} \Gamma x$, whence $(x_0, y_0) \in \sum_{x \in X} \Gamma x$.

Theorem 5. *If $(\Gamma_i \mid i \in I)$ is a family of closed mappings of X into Y, then $\Gamma = \bigcap_{i \in I} \Gamma_i$ is also a closed mapping.*

Proof. If $y_0 \notin \Gamma x_0$, then there exists an index i_0 such that

$$y_0 \notin \Gamma_{i_0} x_0.$$

Therefore there exist neighbourhoods $U(x_0)$ and $V(y_0)$ such that

$$\Gamma_{i_0} U(x_0) \cap V(y_0) = \emptyset$$

and so

$$\Gamma U(x_0) \cap V(y_0) = \emptyset;$$

whence Γ is a closed mapping.

Theorem 6. *Every u.s.c. mapping is closed.*

Proof. Let Γ be a u.s.c. mapping of X into Y and suppose that $y_0 \notin \Gamma x_0$. Since Γx_0 is compact, there exists an open set G in Y containing Γx_0 and a neighbourhood $V(y_0)$ such that

$$G \cap V(y_0) = \varnothing.$$

Since Γ is u.s.c., there exists a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad \Gamma x \subset G.$$

Then

$$x \in U(x_0) \quad \Rightarrow \quad \Gamma x \cap V(y_0) = \varnothing$$

and Γ is therefore closed.

Theorem 7. *If Γ_1 is a closed mapping of X into Y and Γ_2 is a u.s.c. mapping of X into Y, the mapping $\Gamma = \Gamma_1 \cap \Gamma_2$ is u.s.c.*

Proof. For each x, the set Γx is compact, because it is a closed set contained in a compact set $\Gamma_2 x$. Let G be an open set such that $\Gamma x_0 = \Gamma_1 x_0 \cap \Gamma_2 x_0 \subset G$. We shall prove that there exists a neighbourhood $U(x_0)$ such that $\Gamma U(x_0) \subset G$. If $G \supset \Gamma_2 x_0$, there is nothing to prove; suppose then that $\Gamma_2 x_0 \cap (-G) = K \neq \varnothing$. Let y be a point of K; then there exist neighbourhoods $V(y)$ and $U_y(x_0)$ such that

$$\Gamma_1 U_y(x_0) \cap V(y) = \varnothing.$$

Since the set K is compact, there exist elements $y_1, y_2, \ldots, y_n$ in K such that $V(y_1), V(y_2), \ldots, V(y_n)$ cover K. Write $V(K) = \bigcup_{i=1}^{n} V(y_i)$. Then there exists a neighbourhood $U'(x_0)$ such that

$$x \in U'(x_0) \quad \Rightarrow \quad \Gamma_2 x \subset G \cup V(K).$$

Putting $U(x_0) = U_{y_1}(x_0) \cap U_{y_2}(x_0) \cap \ldots \cap U_{y_n}(x_0) \cap U'(x_0)$, we have

$$\begin{cases} \Gamma_1 U(x_0) \cap V(K) = \varnothing, \\ \Gamma_2 U(x_0) \subset G \cup V(K), \end{cases}$$

whence

$$(\Gamma_1 \cap \Gamma_2)\, U(x_0) \subset G.$$

Therefore Γ is u.s.c.

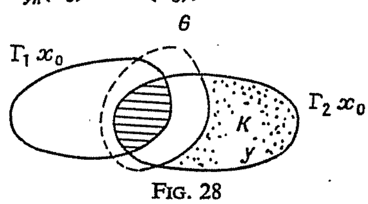


FIG. 28

COROLLARY. *If Y is a compact space, a mapping of X into Y is closed if and only if it is u.s.c.*

Proof. If Γ is closed, and Δ is the mapping such that $\Delta x = Y$ for each x, then $\Gamma = \Gamma \cap \Delta$ is u.s.c., because Δ is u.s.c. The other half of the result follows from Theorem 6.

Theorem 8. *If X is a compact space and Γ is a u.s.c. mapping of X into itself such that, for each x, $\Gamma x \neq \emptyset$, then there exists a compact non-empty subset K of X such that $\Gamma K = K$.*

Proof. The sets $X, \Gamma X, \Gamma^2 X, \ldots$ form a sequence of non-empty compact sets; if any of these is equal to its successor, it is a set of the required kind and so the desired result is proved. We therefore assume that the sets are distinct; since $X \supset \Gamma X$, we have $\Gamma X \supset \Gamma^2 X$, $\Gamma^2 X \supset \Gamma^3 X$ and so on. Thus the sequence of sets $\Gamma^n X$ is decreasing. By the finite intersection axiom, we have

$$K = \bigcap_{n=1}^{\infty} \Gamma^n X \neq \emptyset.$$

For each n, we have $K \subset \Gamma^{n-1} X$ and so $\Gamma K \subset \Gamma^n X$, whence $\Gamma K \subset K$.

We now prove that $K \subset \Gamma K$. Let a be a point of K; then, for each n, there exists a point x_n in $\Gamma^n X$ such that $a \in \Gamma x_n$. The sequence (x_n) admits a cluster point x_0 and a sub-sequence (x_{k_n}) converges to x_0. Since at most $n-1$ points of this sub-sequence are outside $\Gamma^n X$ and $\Gamma^n X$ is compact, we have $x_0 \in \Gamma^n X$ and so $x_0 \in K$. Also $((x_{k_n}, a)) \to (x_0, a)$ and $(x_{k_n}, a) \in \sum_{x \in X} \Gamma x$ and so, since Γ is a closed mapping, $a \in \Gamma x_0$; therefore $a \in \Gamma K$. It follows that $K \subset \Gamma K$ and hence that $\Gamma K = K$.

§2. Properties of the two types of semi-continuity

In this section we compare the properties of lower semi-continuity with those of upper semi-continuity.

Theorem 1. *If Γ_1 is an l.s.c. mapping of X into Y and Γ_2 is an l.s.c. mapping of Y into Z, the composition product $\Gamma = \Gamma_2 \cdot \Gamma_1$ is an l.s.c. mapping of X into Z.*

Proof. If G is an open set in Z, then

$$\Gamma^- G = \{x \,/\, \Gamma_2 \cdot \Gamma_1 x \cap G \neq \emptyset\} = \{x \,/\, \Gamma_1 x \cap \Gamma_2^- G \neq \emptyset\} = \Gamma_1^-(\Gamma_2^- G).$$

Therefore, if Γ_1 and Γ_2 are l.s.c., $\Gamma^-\, G$ is open and so Γ is l.s.c.

Theorem 1′. *If Γ_1 is a u.s.c. mapping of X into Y and Γ_2 is a u.s.c. mapping of Y into Z, the composition product $\Gamma = \Gamma_2 \cdot \Gamma_1$ is a u.s.c. mapping of X into Z.*

Proof. By Theorem 3 of § 1, the set $\Gamma x = \Gamma_2(\Gamma_1 x)$ is compact for all x; moreover, if G is an open set of Z, then

$$\Gamma^+ G = \{x \,/\, \Gamma_2 \cdot \Gamma_1 x \subset G\} = \{x \,/\, \Gamma_1 x \subset \Gamma_2^+ G\} = \Gamma_1^+(\Gamma_2^+ G).$$

Therefore if Γ_1 and Γ_2 are u.s.c., $\Gamma^+\, G$ is open and so Γ is u.s.c.

Theorem 2. *The union* $\Gamma = \bigcup_{i \in I} \Gamma_i$ *of a family of l.s.c. mappings* Γ_i *of* X *into* Y *is also an l.s.c. mapping of* X *into* Y.

Proof. If G is an open set of Y, we have

$$\Gamma^- G = \{x \,/\, \bigcup_{i \in I} \Gamma_i x \cap G \neq \emptyset\} = \bigcup_{i \in I} \Gamma_i^- G.$$

Therefore the set $\Gamma^- G$ is open and so Γ is l.s.c.

Theorem 2′. *The intersection* $\Gamma = \bigcap_{i \in I} \Gamma_i$ *of a family of u.s.c. mappings* Γ_i *of* X *into* Y *is also a u.s.c. mapping of* X *into* Y.

Proof. By Theorem 5 of § 1, the mapping $\Phi = \bigcap_{i \neq i_0} \Gamma_i$ is closed; therefore, by Theorem 7 of § 1, the mapping $\Gamma = \Gamma_{i_0} \cap \Phi$ is u.s.c.

Theorem 3′. *The union* $\Gamma = \bigcup_{i=1}^{n} \Gamma_i$ *of a finite family of u.s.c. mappings* Γ_i *of* X *into* Y *is also u.s.c.*

Proof. The set $\Gamma x = \bigcup_{i=1}^{n} \Gamma_i x$ is compact for all x, by Theorem 4 of § 6, Chapter IV. Moreover, if G is an open set in Y, we have

$$\Gamma^+ G = \{x \,/\, \bigcup_{i=1}^{n} \Gamma_i x \subset G\} = \bigcap_{i=1}^{n} \Gamma_i^+ G.$$

Therefore $\Gamma^+ G$ is an open set and so Γ is u.s.c.

In this case the analogous property for l.s.c. mappings does not hold.

Theorem 4. *The Cartesian product* $\Gamma = \prod_{i=1}^{n} \Gamma_i$ *of a finite number of l.s.c. mappings* Γ_i *of* X *into* Y_i *is an l.s.c. mapping of* X *into* $Y = \prod_{i=1}^{n} Y_i$.

Proof. Let $G = \bigcup_k E^k$ be the union of the elementary open sets E^k in Y. Since $E^k = \prod_{i=1}^{n} G_i^k$, where G_i^k is an open set in Y_i, we have

$$\Gamma^- E^k = \{x \,/\, \prod_{i=1}^{n} \Gamma_i x \cap \prod_{i=1}^{n} G_i^k \neq \emptyset\} = \bigcap_{i=1}^{n} \Gamma_i^- G_i^k.$$

Then $\Gamma^- E^k$ is an open set and therefore so is $\Gamma^- G = \bigcup_k \Gamma^- E^k$. Hence Γ is l.s.c.

Theorem 4′. *The Cartesian product* $\Gamma = \prod_{i=1}^{n} \Gamma_i$ *of a finite number of u.s.c. mappings* Γ_i *of* X *into* Y_i *is a u.s.c. mapping of* X *into* $Y = \prod_{i=1}^{n} Y_i$.

Proof. To fix our ideas, we consider the case $n = 2$. Since $\Gamma_1 x$ and $\Gamma_2 x$ are compact, so is $\Gamma x = \Gamma_1 x \times \Gamma_2 x$, by Tychonoff's Theorem, § 9, Chapter IV.

Let G be an open set in Y and let a be an element of $\Gamma^+ G$. Since Γa is compact and is contained in G, it can be covered by a finite number of elementary open sets contained in G; let $E^1, E^2, \ldots, E^n$ be such a finite covering. Suppose that $(y_1, y_2) \in \Gamma_1 a \times \Gamma_2 a$; let $E(y_1)$ denote the union of the E^k meeting the set $\{y_1\} \times \Gamma_2 a$ and let $E(y_2)$ be the union of the E^k meeting the set $\Gamma_1 a \times \{y_2\}$. The projection $\pi_2 E(y_1)$ of the set $E(y_1)$ on the space Y_2 is open, by Theorem 2 of § 9, Chapter IV. As y_1 varies the sets $\pi_2 E(y_1)$ are finite in number and therefore their intersection is open; similarly the intersection of the sets $\pi_1 E(y_2)$ is open. We write

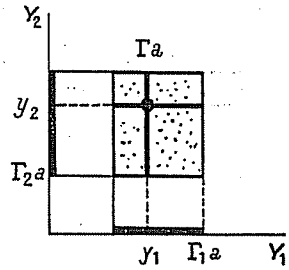


FIG. 29

$$G_1 = \bigcap_{y_2 \in \Gamma_2 a} \pi_1 E(y_2),$$

$$G_2 = \bigcap_{y_1 \in \Gamma_1 a} \pi_2 E(y_1).$$

Then $E = G_1 \times G_2$ is an elementary open set of Y and

$$\Gamma^+ E = \{x \,/\, \Gamma_1 x \times \Gamma_2 x \subset G_1 \times G_2\} = (\Gamma_1^+ G_1 \cap \Gamma_2^+ G_2) \cup \Gamma_1^+ \emptyset \cup \Gamma_2^+ \emptyset.$$

Therefore $\Gamma^+ E$ is an open set of X; since $\Gamma a \subset E \subset G$, we also have

$$a \in \Gamma^+ E \subset \Gamma^+ G.$$

Therefore the set $\Gamma^+ G$ is a neighbourhood of each of its points and so is open; hence Γ is u.s.c.

§ 3. Maximum theorem

We shall make frequent use of the following results.

Theorem 1. *If ϕ is a lower semi-continuous numerical function in $X \times Y$ and Γ is an l.s.c. mapping of X into Y such that, for each x, $\Gamma x \neq \emptyset$, the numerical function M defined by*

$$M(x) = \sup \{\phi(x, y) \,/\, y \in \Gamma x\}$$

is lower semi-continuous.

Proof. Suppose that $x_0 \in X$ and let y_0 be such that

$$y_0 \in \Gamma x_0; \ \phi(x_0, y_0) \geqq M(x_0) - \varepsilon.$$

There exist neighbourhoods $U(x_0)$ and $V(y_0)$ such that

$$(x, y) \in U(x_0) \times V(y_0) \Rightarrow \phi(x, y) \geqq \phi(x_0, y_0) - \varepsilon \geqq M(x_0) - 2\varepsilon;$$

there exists a neighbourhood $U'(x_0)$ such that

$$x \in U'(x_0) \Rightarrow \Gamma x \cap V(y_0) \neq \emptyset.$$

Therefore

$$x \in U(x_0) \cap U'(x_0) \Rightarrow M(x) \geqq M(x_0) - 2\varepsilon.$$

Theorem 2. *If ϕ is an upper semi-continuous numerical function in $X \times Y$ and Γ is a u.s.c. mapping of X into Y such that, for each x, $\Gamma x \neq \emptyset$, the numerical function M defined by*

$$M(x) = \max \{\phi(x, y) \mid y \in \Gamma x\}$$

is upper semi-continuous.

Proof. Suppose that $x_0 \in X$; to each y in Γx_0 there correspond neighbourhoods $U_y(x_0)$ and $V(y)$ such that

$$(x, z) \in U_y(x_0) \times V(y) \quad \Rightarrow \quad \phi(x, z) \leqq \phi(x_0, y) + \varepsilon.$$

Since Γx_0 is compact, it can be covered by a finite number of neighbourhoods of the form $V(y)$, say $V(y_1), V(y_2), \ldots, V(y_n)$. Putting $U'(x_0) = \bigcap_{i=1}^{n} U_{y_i}(x_0)$ and $V(\Gamma x_0) = \bigcup_{i=1}^{n} V(y_i)$, we have

$$x \in U'(x_0),\ y \in V(\Gamma x_0) \Rightarrow \phi(x, y) \leqq \max_i \phi(x_0, y_i) + \varepsilon \leqq M(x_0) + \varepsilon.$$

Moreover there exists a neighbourhood $U(x_0)$ such that

$$x \in U(x_0) \quad \Rightarrow \quad \Gamma x \subset V(\Gamma x_0)$$

and so

$$x \in U(x_0) \cap U'(x_0) \quad \Rightarrow \quad M(x) = \max_{y \in \Gamma x} \phi(x, y) \leqq M(x_0) + \varepsilon.$$

Maximum theorem. *If ϕ is a continuous numerical function in Y and Γ is a continuous mapping of X into Y such that, for each x, $\Gamma x \neq \emptyset$, then the numerical function M defined by $M(x) = \max \{\phi(y) \mid y \in \Gamma x\}$ is continuous in X and the mapping Φ defined by $\Phi x = \{y \mid y \in \Gamma x,\ \phi(y) = M(x)\}$ is a u.s.c. mapping of X into Y.*

Proof. The function ϕ is continuous in $X \times Y$ and so M is a continuous numerical function; moreover the mapping Δ given by

$$\Delta x = \{y \mid M(x) - \phi(y) \leqq 0\}$$

is closed (by the example following Theorem 3 of § 1) and hence $\Phi = \Gamma \cap \Delta$ is u.s.c.

Let ϕ be a continuous numerical function defined in a topological space Y. A family of compact sets $\mathscr{K} = (K_i \mid i \in I)$ in Y is called **selective** (with respect to ϕ) if for each i there exists one and only one a_i such that

$$a_i \in K_i; \ \phi(a_i) = \max \{\phi(y) \mid y \in K_i\}.$$

In other words, the maximum of ϕ is attained at only one point of the set K_i. For example, in $\mathbf{R}^n$ every family of balls is selective, with respect to $\phi(y) = \pi_i y$; in $\mathbf{R}$, every family of compact sets is selective with respect to $\phi(y) = y$.

Theorem 3. *Let Γ be a continuous mapping of X into Y such that, for each x, $\Gamma x \neq \emptyset$. If the family of sets $(\Gamma x \mid x \in X)$ is selective, there is a single-valued continuous mapping σ of X into Y such that, for each x, $\sigma x \in \Gamma x$.*

Proof. Let ϕ be a continuous numerical function in Y for which $(\Gamma x \mid x \in X)$ is selective; if $\Phi x = \{y \mid y \in \Gamma x,\ \phi(y) = M(x)\}$ then Φ is a single-valued mapping of X into Y. Moreover, it is u.s.c. by the preceding theorem and so is continuous (since it is single-valued). The mapping $\sigma x = \Phi x$ satisfies

$$(\forall x) : \sigma x \in \Gamma x.$$

as required.

COROLLARY. *If Γ is a continuous mapping of X into $\mathbf{R}$ such that, for each x, $\Gamma x \neq \emptyset$, there exists a continuous single-valued mapping σ such that, for each x, $\sigma x \in \Gamma x$.*

Proof. It is sufficient to take $\phi(y) = y$, whence $\sigma x = \max \Gamma x$.

§ 4. Fixed points of a mapping of R into R

We can now prove an important result which will be generalised in Chapters VIII and IX.

Kakutani's theorem (weak form). *Let $[a, b]$ be a closed interval in $\mathbf{R}$ and let Γ be a u.s.c. mapping of $[a, b]$ into $[a, b]$. If Γx is a closed interval for each x, then there exists an x_0 such that $x_0 \in \Gamma x_0$.*

Proof. (1) We first show that if ε is a positive number, then there exists a point x_0 in $[a, b]$ such that $B_\varepsilon(x_0) \cap \Gamma x_0 \neq \emptyset$.

Suppose that this is not the case. Then, for each x, we have $B_\varepsilon(x) \cap \Gamma x = \emptyset$ and Γx (being a closed interval) can only be to the right of x or to the left of x in $[a, b]$. Let A be the set of $x \in \mathbf{R}$ such that Γx is to the right of x and let B be the set such that Γx is to the left of x. Since $a \in A$ and $b \in B$, the sets A and B are both non-empty. Also $A \cap B = \emptyset$ and $A \cup B = [a, b]$.

Moreover, if $x_1 \in A$, then, since Γx_1 is compact,

$$x_1 + \varepsilon < \min \Gamma x_1.$$

Let λ be a number such that

$$x_1 + \varepsilon < \lambda < \min \Gamma x_1.$$

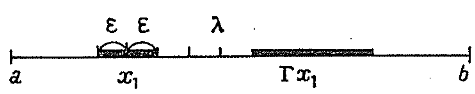


FIG. 30

Since Γ is u.s.c., there is a neighbourhood $V(x_1)$ such that

$$x \in V(x_1) \quad \Rightarrow \quad \Gamma x \subset]\lambda, b]$$

and so

$$x \in V(x_1) \cap B_\varepsilon(x_1) \quad \Rightarrow \quad x \in A.$$

Therefore A is open; by symmetry B is also open. But this implies that A and B determine an open partition of $[a, b]$, which is not possible since $[a, b]$ is connected. Therefore we have reached a contradiction.

(2) It follows that, for each integer n, the set

$$K_n = \{x \mid B_{\frac{1}{n}}(x) \cap \Gamma x \neq \emptyset\}$$

is non-empty. Since the mapping Δ such that $\Delta x = B_{\frac{1}{n}}(x) \cap \Gamma x$ is u.s.c., the set $\Delta^+ \emptyset$ is open and therefore $K_n = -\Delta^+ \emptyset$ is closed. Moreover $K_1 \supset K_2 \supset K_3 \supset \ldots$ and so, by the finite intersection axiom, there exists a point x_0 of $[a, b]$ belonging to all the K_n. Then

$$(\forall n) : B_{\frac{1}{n}}(x_0) \cap \Gamma x_0 \neq \emptyset.$$

Hence x_0 is a cluster point of Γx_0 and so, since Γx_0 is closed, we have $x_0 \in \Gamma x_0$.

COROLLARY 1. **Brouwer's fixed point theorem** (weak form). *If σ is a continuous single-valued mapping of $[a, b]$ into $[a, b]$, there exists a point x_0 such that $\sigma x_0 = x_0$.*

Proof. The mapping σ satisfies the conditions of Kakutani's Theorem, $\sigma x = [y, y] = \{y\}$ being a closed interval.

COROLLARY 2. *If Γ is a continuous mapping of $[a, b]$ into $[a, b]$ such that, for each x, $\Gamma x \neq \emptyset$, there exists a point x_0 in $[a, b]$ such that*

$$x_0 \in \Gamma x_0; \; x_0 = \max \Gamma x_0.$$

Proof. The mapping $\sigma x = \max \Gamma x$ is continuous, by the Maximum Theorem, and so we can apply Corollary 1 to show that there exists an x_0 such that

$$x_0 = \sigma x_0 = \max \Gamma x_0.$$

§ 5.* Limits of a family of sets

Let $(A_1, A_2, \ldots) = (A_n)$ be a sequence of sets[1] in a topological space. We say that x_0 is a **limit point** of (A_n) if to each neighbourhood $U(x_0)$ there corresponds an integer n such that

$$k \geqq n \quad \Rightarrow \quad A_k \cap U(x_0) \neq \emptyset.$$

[1] The results on sequences of sets are due to Kuratowski (*Topologie*, 1). Certain extensions to filtered families of sets have been made by G. Choquet (*Ann. Univ. Grenoble*, 1947, vol. 23, p. 57).

We say that x_0 is a **cluster point** of (A_n) if for each neighbourhood $U(x_0)$ and for each positive integer n, we have

$$(\exists k;\ k \geqq n) : A_k \cap U(x_0) \neq \emptyset.$$

To extend these ideas to any family $(A_i \mid i \in I)$, we associate with it a filter base $\mathscr{B} = (S, S', \ldots)$ on the index set I; we then obtain a **filtered family of sets**, which is denoted by $(A_i) = (A_i \mid i \in I, \mathscr{B})$. Unless otherwise stated, a sequence (A_n) will always be considered to have the Fréchet base consisting of the sections $S_n = \{k \mid k \in \mathrm{N},\ k \geqq n\}$. For a filtered family $(A_i) = (A_i \mid i \in I, \mathscr{B})$ we say that x_0 is a **limit point** if

$$(\forall U(x_0))\,(\exists S;\ S \in \mathscr{B})\,(\forall_S\, i) : A_i \cap U(x_0) \neq \emptyset.$$

We say that x_0 is a **cluster point** if

$$(\forall U(x_0))\,(\forall S)\,(\exists_S\, i) : A_i \cap U(x_0) \neq \emptyset.$$

The set of limit points of (A_i) is called the **lower limit** of (A_i) and is denoted by $\underline{\mathrm{Lim}}\,(A_i)$; the set of cluster points of (A_i) is called the **upper limit** of (A_i) and is denoted by $\overline{\mathrm{Lim}}\,(A_i)$. In all cases $\underline{\mathrm{Lim}}\,(A_i) \subset \overline{\mathrm{Lim}}\,(A_i)$. If $\underline{\mathrm{Lim}}\,(A_i) = \overline{\mathrm{Lim}}\,(A_i) = A_0$, we say that the family (A_i) **converges** to A_0 and write $(A_i) \to A_0$, or that A_0 is the **limit** of (A_i) and write $A_0 = \mathrm{Lim}\,(A_i)$.

EXAMPLE 1. Let X be a set with the discrete topology. Then $x_0 \in \underline{\mathrm{Lim}}\,(A_i)$ is equivalent to

$$(\exists S)\,(\forall_S\, i) : x_0 \in A_i.$$

selon

$\underline{\mathrm{Lim}}\,(A_i)$ is then the **principal lower limit** of (A_i) with respect to $\mathscr{B}$, which we have already encountered (Chapter I, § 9) and which we write $\underline{\mathrm{Lim}}_{\mathscr{B}}\,(A_i)$.

Further, $x_0 \in \overline{\mathrm{Lim}}\,(A_i)$ is equivalent to

$$(\forall S)\,(\exists_S\, i) : x_0 \in A_i.$$

$\overline{\mathrm{Lim}}\,(A_i)$ is then the **principal upper limit** of (A_i) with respect to $\mathscr{B}$, which we write $\overline{\mathrm{Lim}}_{\mathscr{B}}\,(A_i)$.

EXAMPLE 2. Let (A_n) be a sequence of sets in a metric space X. Then $x_0 \in \underline{\mathrm{Lim}}\,(A_i)$ is equivalent to

$$(\forall B_\varepsilon(x_0))\,(\exists n) : k \geqq n \text{ implies that } A_k \cap B_\varepsilon(x_0) \neq \emptyset.$$

In other words, $x_0 \in \underline{\mathrm{Lim}}\,(A_i)$ if the sequence whose general term is $d(x_0, A_n)$ converges to 0.

Further, $x_0 \in \overline{\mathrm{Lim}}\,(A_i)$ is equivalent to

$$(\forall B_\varepsilon(x_0))\,(\forall n)\,(\exists k;\ k \geqq n) : A_k \cap B_\varepsilon(x_0) \neq \emptyset.$$

In other words, $x_0 \in \overline{\mathrm{Lim}}\,(A_i)$ if the sequence whose general term is $d(x_0, A_n)$ admits 0 as a cluster point.

EXAMPLE 3. Let Γ be a mapping of X into Y such that, for each x, $\Gamma x \neq \emptyset$; if $x_0 \in X$, the lower limit of the filtered family $(\Gamma x \mid x \in X; \mathscr{V}(x_0))$ is written $\underline{\operatorname{Lim}}_{x \to x_0} \Gamma x$ and the upper limit is written $\overline{\operatorname{Lim}}_{x \to x_0} \Gamma x$.

In a compact set, the sets $\underline{\operatorname{Lim}}_{x \to x_0} \Gamma x$ and $\overline{\operatorname{Lim}}_{x \to x_0} \Gamma x$ can be used to define lower and upper semi-continuity at x_0 of a mapping Γ.

Theorem 1. *If $\ddot{\mathscr{B}}$ is the grill of the filter base $\mathscr{B}$, the lower limit of the family $(A_i) = (A_i \mid i \in I, \mathscr{B})$ is given by*

$$\underline{\operatorname{Lim}}\,(A_i) = \bigcap_{H \in \ddot{\mathscr{B}}} \overline{\Big[\bigcup_{i \in H} A_i\Big]}.$$

Proof. We recall that the grill $\ddot{\mathscr{B}}$ consists of the sets H which meet all the sets S of $\mathscr{B}$. If $x_0 \in \underline{\operatorname{Lim}}\,(A_i)$, then to each neighbourhood $U(x_0)$ there corresponds a set S_U of $\mathscr{B}$ such that

$$i \in S_U \text{ implies that } A_i \cap U(x_0) \neq \emptyset.$$

Then for all H we have

$$\bigcup_{i \in H} A_i \cap U(x_0) \neq \emptyset,$$

since H belongs to the grill $\ddot{\mathscr{B}}$ and so meets S_U.

Conversely, suppose that $x_0 \notin \underline{\operatorname{Lim}}\,(A_i)$. Then

$$(\exists U(x_0))\,(\forall S)\,(\exists_S\, i(S))) : A_{i(S)} \cap U(x_0) = \emptyset$$

and so, if $H = \{i(S) \mid S \in \mathscr{B}\}$, we have

$$\bigcup_{i \in H} A_i \cap U(x_0) = \emptyset.$$

To sum up, we have

$$x_0 \in \underline{\operatorname{Lim}}\,(A_i) \Leftrightarrow (\forall H)\,(\forall U(x_0)) : \bigcup_{i \in H} A_i \cap U(x_0) \neq \emptyset \Leftrightarrow (\forall H) : x_0 \in \overline{\bigcup_{i \in H} A_i},$$

whence

$$\underline{\operatorname{Lim}}\,(A_i) = \bigcap_{H \in \ddot{\mathscr{B}}} \overline{\Big[\bigcup_{i \in H} A_i\Big]}.$$

Theorem 1′. *The upper limit of the family $(A_i \mid i \in I, \mathscr{B})$ is*

$$\overline{\operatorname{Lim}}\,(A_i) = \bigcap_{S \in \mathscr{B}} \overline{\Big[\bigcup_{i \in S} A_i\Big]}.$$

Proof. We have

$$x_0 \in \overline{\operatorname{Lim}}\,(A_i) \Leftrightarrow (\forall S)\,(\forall U(x_0)) : \bigcup_{i \in S} A_i \cap U(x_0) \neq \emptyset \Leftrightarrow (\forall S) : x_0 \in \overline{\bigcup_{i \in S} A_i}$$

and the result follows.

These formulae generalise those which give the limits of a family of sets with respect to $\mathscr{B}$ (Chapter I, § 9).

COROLLARY 1. *$\underline{\mathrm{Lim}}\,(A_i)$ and $\overline{\mathrm{Lim}}\,(A_i)$ are closed sets.*

This is immediate, since $\underline{\mathrm{Lim}}\,(A_i)$ and $\overline{\mathrm{Lim}}\,(A_i)$ have been proved to be intersections of closed sets.

COROLLARY 2. *If $A_i \subset B_i$ for all i, then*

$$\underline{\mathrm{Lim}}\,(A_i) \subset \underline{\mathrm{Lim}}\,(B_i),$$
$$\overline{\mathrm{Lim}}\,(A_i) \subset \overline{\mathrm{Lim}}\,(B_i).$$

This is an immediate consequence of Theorems 1 and 1′.

COROLLARY 3. *We have*

$$\underline{\mathrm{Lim}}\,(\bar{A}_i) = \underline{\mathrm{Lim}}\,(A_i),$$
$$\overline{\mathrm{Lim}}\,(\bar{A}_i) = \overline{\mathrm{Lim}}\,(A_i).$$

Proof. We have $\overline{\bigcup_{i \in J} A_i} = \overline{\bigcup_{i \in J} \bar{A}_i}$, since $\overline{\bigcup_{i \in J} A_i}$ is the intersection of the closed sets containing all the A_i, or, what amounts to the same thing, all the $\bar{A}_i$.

Theorem 2. *If $(A'_i \,/\, i \in I, \mathscr{B}')$ is a sub-family of*

$$(A_i) = (A_i \,/\, i \in I, \mathscr{B})$$

(that is, if $A_i = A'_i$, $\mathscr{B}' \vdash \mathscr{B}$), we have

$$\underline{\mathrm{Lim}}\,(A_i) \subset \underline{\mathrm{Lim}}\,(A'_i).$$

Proof. If $x_0 \in \underline{\mathrm{Lim}}\,(A_i)$, then for each $U(x_0)$ there exists a set $S \in \mathscr{B}$ such that $i \in S$ implies that $A_i \cap U(x_0) \neq \emptyset$. There exists a set S' in $\mathscr{B}'$ contained in S and for this set we have

$$i \in S' \quad \Rightarrow \quad A_i \cap U(x_0) \neq \emptyset.$$

Hence $x_0 \in \underline{\mathrm{Lim}}\,(A'_i)$.

Theorem 2′. *If (A'_i) is a sub-family of (A_i), then*

$$\overline{\mathrm{Lim}}\,(A_i) \supset \overline{\mathrm{Lim}}\,(A'_i).$$

Proof. If $x_0 \in \overline{\mathrm{Lim}}\,(A'_i)$, then for all $U(x_0)$ and for all S' in $\mathscr{B}'$,

$$\bigcup_{i \in S'} A_i \cap U(x_0) \neq \emptyset.$$

But for each S in $\mathscr{B}$, there exists an S' in $\mathscr{B}'$ contained in S and so we have

$$\bigcup_{i \in S} A_i \cap U(x_0) \neq \emptyset.$$

Hence $x_0 \in \overline{\mathrm{Lim}}\,(A_i)$.

COROLLARY. *If* (A'_i) *is a sub-family of* (A_i) *and if* $(A_i) \to A_0$, *then* $(A'_i) \to A_0$.

Proof. We have

$$A_0 = \underline{\mathrm{Lim}}\,(A_i) \subset \underline{\mathrm{Lim}}\,(A'_i) \subset \overline{\mathrm{Lim}}\,(A'_i) \subset \overline{\mathrm{Lim}}\,(A_i) = A_0$$

and so

$$\underline{\mathrm{Lim}}\,(A'_i) = \overline{\mathrm{Lim}}\,(A'_i) = A_0.$$

Theorem 3. *If a filtered family is such that* $A_i = A_0$ *for all* i, *then* $(A_i) \to \bar{A}_0$.

Proof. We shall show that

$$\bar{A}_0 \overset{(1)}{\subset} \underline{\mathrm{Lim}}\,(A_i) \overset{(2)}{\subset} \overline{\mathrm{Lim}}\,(A_i) \overset{(3)}{\subset} \bar{A}_0.$$

The inclusion (2) is immediate. For the inclusion (1), we have

$$x_0 \in \bar{A}_0 \Rightarrow (\forall U(x_0)) : A_i \cap U(x_0) \neq \varnothing \Rightarrow x_0 \in \underline{\mathrm{Lim}}\,(A_i).$$

For the inclusion (3), we have

$$x_0 \in \overline{\mathrm{Lim}}\,(A_i) \Rightarrow (\forall U(x_0))\,(\forall S) : \bigcup_{i \in S} A_i \cap U(x_0) \neq \varnothing \Rightarrow x_0 \in \bar{A}_0.$$

Thus the theorem is proved.

The definitions of lower and upper limits can be formulated rather differently as follows. Let (A_i) be a filtered family and let Γ be the mapping of X into I defined by

$$\Gamma x = \{i \,/\, x \in A_i\}.$$

Suppose that $U \in \mathscr{V}(x_0)$. We have

$$\Gamma U = \{i \,/\, A_i \cap U \neq \varnothing\};$$

then $x_0 \in \underline{\mathrm{Lim}}\,(A_i)$ is equivalent to

$$(\forall U(x_0))\,(\exists S) : S \subset \Gamma U(x_0).$$

Similarly, $x_0 \in \overline{\mathrm{Lim}}\,(A_i)$ is equivalent to

$$(\forall U(x_0))\,(\forall S) : S \cap \Gamma U(x_0) \neq \varnothing.$$

These remarks help us to make the proofs of the following theorems somewhat easier.

Theorem 4. *If* (A_i^α) *is a sub-family of* (A_i), *then*

$$\underline{\mathrm{Lim}}\,(A_i) = \bigcap_\alpha \overline{\mathrm{Lim}}\,(A_i^\alpha).$$

Proof. If $x_0 \in \underline{\mathrm{Lim}}\,(A_i)$, then, for any sub-family (A_i^α),

$$x_0 \in \underline{\mathrm{Lim}}\,(A_i) \subset \underline{\mathrm{Lim}}\,(A_i^\alpha) \subset \overline{\mathrm{Lim}}\,(A_i^\alpha),$$

whence

$$x_0 \in \bigcap_\alpha \overline{\mathrm{Lim}}\,(A_i^\alpha).$$

Conversely, if $x_0 \notin \underline{\mathrm{Lim}}\,(A_i)$, there exists a neighbourhood U of x_0 such that

$$(\forall S) : S \cap (-\Gamma U) \neq \varnothing.$$

Then, by Theorem 3 of § 6, Chapter I, the sets $S \cap (-\Gamma U)$ form a filter base $\mathscr{B}'$ with $\mathscr{B}' \vdash \mathscr{B}$; every set A'_i of the sub-family of (A_i) defined by means of $\mathscr{B}'$ is disjoint from U and so

$$x_0 \notin \overline{\mathrm{Lim}}\,(A'_i).$$

Theorem 4′. *If (A^α_i) is a sub-family of (A_i), then*

$$\overline{\mathrm{Lim}}\,(A_i) = \bigcup_\alpha \underline{\mathrm{Lim}}\,(A^\alpha_i).$$

Proof. If $x_0 \in \underline{\mathrm{Lim}}\,(A^\alpha_i)$, we have

$$x_0 \in \underline{\mathrm{Lim}}\,(A^\alpha_i) \subset \overline{\mathrm{Lim}}\,(A^\alpha_i) \subset \overline{\mathrm{Lim}}\,(A_i)$$

and so $x_0 \in \overline{\mathrm{Lim}}\,(A_i)$.

Conversely, if $x_0 \in \overline{\mathrm{Lim}}\,(A_i)$, the sets $S \cap \Gamma U$ (where S runs through $\mathscr{B}$ and U runs through $\mathscr{V}(x_0)$), form a filter base $\mathscr{B}'$ with $\mathscr{B}' \vdash \mathscr{B}$. Writing $(A'_i) = (A_i \,/\, i \in I, \mathscr{B}')$, we have

$$(\forall U(x_0))\,(\exists S';\ S' \in \mathscr{B}') : S' \subset \Gamma U(x_0).$$

Therefore $x_0 \in \underline{\mathrm{Lim}}\,(A'_i)$ and so the proof of the theorem is complete.

Theorem 5. *Let (A_i) be a filtered family of sets; if every sub-family (A'_i) has a sub-family (A''_i) converging to A_0, then*

$$(A_i) \to A_0.$$

Proof. We shall show that

$$A_0 \overset{(1)}{\subset} \underline{\mathrm{Lim}}\,(A_i) \overset{(2)}{\subset} \overline{\mathrm{Lim}}\,(A_i) \overset{(3)}{\subset} A_0.$$

The inclusion (2) is immediate. For the inclusion (1), we have, for any sub-family (A'_i),

$$A_0 = \overline{\mathrm{Lim}}\,(A''_i) \subset \overline{\mathrm{Lim}}\,(A'_i).$$

Then, by Theorem 4,

$$A_0 \subset \bigcap_\alpha \overline{\mathrm{Lim}}\,(A^\alpha_i) = \underline{\mathrm{Lim}}\,(A_i).$$

To prove the inclusion (3), we suppose that $x_0 \in \overline{\mathrm{Lim}}\,(A_i)$; then, by Theorem 4′, there exists a sub-family (A'_i) such that $x_0 \in \underline{\mathrm{Lim}}\,(A'_i)$ and so

$$x_0 \in \underline{\mathrm{Lim}}\,(A'_i) \subset \underline{\mathrm{Lim}}\,(A''_i) = A_0.$$

REMARK. We note that limits of sequences (F_n) of non-empty closed sets have the following properties:

(1) If (F_n') is a sub-sequence of (F_n) and if $F_0 = \mathrm{Lim}\,(F_n)$, then

$$F_0 = \mathrm{Lim}\,(F_n').$$

(2) If $F_n = F_0$ for all n, the sequence (F_n) is convergent and $F_0 = \mathrm{Lim}\, F_n$.

(3) If every sub-sequence (F_n') contains a sub-sequence which converges to F_0, the sequence (F_n) is convergent and $F_0 = \mathrm{Lim}\,(F_n)$.

Thus the collection $\mathscr{F}$ of non-empty closed sets forms an L^*-space (cf. § 2 of Chapter IV). Unfortunately, the filtered families (F_i) do not in general possess the fourth characteristic property of filtered families in a topological space (Theorem 5, § 4, Chapter IV) and so it is not possible to define the convergence of the (F_i) by means of a topological structure.

Theorem 6. *We have*

$$\overline{\mathrm{Lim}}\,(A_i \cup B_i) = \overline{\mathrm{Lim}}\,(A_i) \cup \overline{\mathrm{Lim}}\,(B_i).$$

Proof. Since $A_i \cup B_i \supset A_i$, then $\overline{\mathrm{Lim}}\,(A_i \cup B_i) \supset \overline{\mathrm{Lim}}\,(A_i)$ whence, by symmetry,

$$\overline{\mathrm{Lim}}\,(A_i \cup B_i) \supset \overline{\mathrm{Lim}}\,(A_i) \cup \overline{\mathrm{Lim}}\,(B_i).$$

We now prove the opposite inclusion. Let x_0 be a point of

$$\overline{\mathrm{Lim}}\,(A_i \cup B_i),$$

and, for each neighbourhood U of x_0, write

$$\Gamma U = \{i \,/\, A_i \cap U \neq \varnothing\};\ \Delta U = \{i \,/\, B_i \cap U \neq \varnothing\}.$$

We must show that at least one of the following relations holds:

$$(\forall U)(\forall S) : S \cap \Gamma U \neq \varnothing,$$
$$(\forall U)(\forall S) : S \cap \Delta U \neq \varnothing.$$

Suppose that this is not so; then there exist sets S and T in $\mathscr{B}$ and sets U and V in $\mathscr{V}(x_0)$ such that

$$S \cap \Gamma U = \varnothing, \quad T \cap \Delta V = \varnothing.$$

If $S_0 \subset S \cap T$ and $U_0 \subset U \cap V$, then

$$S_0 \cap \Gamma U_0 = \varnothing,$$
$$S_0 \cap \Delta U_0 = \varnothing,$$

whence

$$S_0 \cap [\Gamma U_0 \cup \Delta U_0) = S_0 \cap \{i \,/\, (A_i \cup B_i) \cap U_0 \neq \varnothing\} = \varnothing.$$

Since $x_0 \in \overline{\mathrm{Lim}}\,(A_i \cup B_i)$, we have a contradiction.

Theorem 6′. *We have*

$$\underline{\mathrm{Lim}}\,(A_i \cup B_i) \supset \underline{\mathrm{Lim}}\,(A_i) \cup \underline{\mathrm{Lim}}\,(B_i),$$
$$\underline{\mathrm{Lim}}\,(A_i \cup B_i) \subset \underline{\mathrm{Lim}}\,(A_i) \cup \underline{\mathrm{Lim}}\,(B_i) \cup [\overline{\mathrm{Lim}}\,(A_i) \cap \overline{\mathrm{Lim}}\,(B_i)].$$

Proof. The first relation is immediate. To prove the second, suppose that x_0 is such that

$$x_0 \in \underline{\mathrm{Lim}}\,(A_i \cup B_i),$$
$$x_0 \notin \overline{\mathrm{Lim}}\,(A_i) \cap \overline{\mathrm{Lim}}\,(B_i),$$
$$x_0 \notin \underline{\mathrm{Lim}}\,(A_i),$$
$$x_0 \notin \underline{\mathrm{Lim}}\,(B_i).$$

We can suppose that $x_0 \notin \overline{\mathrm{Lim}}\,(A_i)$ (if not, then we interchange the families (A_i) and (B_i)). Then there exist a set U_0 in $\mathscr{V}(x_0)$ and a set S_0 in $\mathscr{B}$ such that

$$i \in S_0 \quad \Rightarrow \quad A_i \cap U_0 = \varnothing.$$

For any neighbourhood U of x_0, there exists a set S_1 such that

$$\begin{cases} i \in S_0 & \Rightarrow \quad A_i \cap (U \cap U_0) = \varnothing, \\ i \in S_1 & \Rightarrow \quad (A_i \cup B_i) \cap (U \cap U_0) \neq \varnothing. \end{cases}$$

Then, if $S \subset S_0 \cap S_1$, $S \in \mathscr{B}$, we have

$$i \in S \quad \Rightarrow \quad B_i \cap (U \cap U_0) \neq \varnothing \quad \Rightarrow \quad B_i \cap U \neq \varnothing.$$

Since $x_0 \notin \underline{\mathrm{Lim}}\,(B_i)$, we have a contradiction, and so the theorem is proved.

COROLLARY 1. *If the family (B_i) converges, then*

$$\overline{\mathrm{Lim}}\,(A_i \cup B_i) = \overline{\mathrm{Lim}}\,(A_i) \cup \mathrm{Lim}\,(B_i),$$
$$\underline{\mathrm{Lim}}\,(A_i \cup B_i) = \underline{\mathrm{Lim}}\,(A_i) \cup \mathrm{Lim}\,(B_i).$$

This follows immediately from Theorems 6 and 6′.

COROLLARY 2. *If the family (A_i) converges to A_0 and the family (B_i) converges to B_0, the family $(A_i \cup B_i)$ converges to $A_0 \cup B_0$.*

This follows immediately from the preceding corollary.

COROLLARY 3. *If the family $(A_i \cup B_i)$ converges, and if*

$$\overline{\mathrm{Lim}}\,(A_i) \cap \overline{\mathrm{Lim}}\,(B_i) = \varnothing,$$

then the families (A_i) and (B_i) converge.

Proof. In this case, we have

$$\underline{\mathrm{Lim}}\,(A_i) \cup \underline{\mathrm{Lim}}\,(B_i) = \underline{\mathrm{Lim}}\,(A_i \cup B_i) = \overline{\mathrm{Lim}}\,(A_i \cup B_i) = \overline{\mathrm{Lim}}\,(A_i) \cup \overline{\mathrm{Lim}}\,(B_i).$$

Since $\underline{\mathrm{Lim}}\,(A_i) \subset \overline{\mathrm{Lim}}\,(A_i)$, $\underline{\mathrm{Lim}}\,(B_i) \subset \overline{\mathrm{Lim}}\,(B_i)$, we have

$$\underline{\mathrm{Lim}}\,(A_i) = \overline{\mathrm{Lim}}\,(A_i),$$
$$\underline{\mathrm{Lim}}\,(B_i) = \overline{\mathrm{Lim}}\,(B_i).$$

§ 6.* Hausdorff metrics

Some of the ideas which we have just studied take on a special significance in the case in which X is a metric space.

Let A and B be two non-empty closed sets in a metric space X, and write

$$\rho(A, B) = \sup \{d(x, B) \,/\, x \in A\},$$
$$\rho(B, A) = \sup \{d(y, A) \,/\, y \in B\},$$

where $d(x, B)$ is the distance from the point x to the set B (cf. § 1, Chapter V). The numerical function δ defined by

$$\delta(A, B) = \max \{\rho(A, B), \rho(B, A)\}$$

is called a **Hausdorff metric.** We shall prove that δ satisfies the necessary properties of a metric for the family $\mathscr{F}'$ of non-empty closed sets.

(1) $\delta(A, B) \geqq 0$.
(2) $\delta(A, B) = 0 \Rightarrow \rho(A, B) = 0 \Rightarrow A \subset B$, and so, by symmetry, $A = B$;
$A = B \Rightarrow \rho(A, B) = 0,\ \rho(B, A) = 0 \Rightarrow \delta(A, B) = 0$.
(3) $\delta(A, B) = \delta(B, A)$.
(4) To prove the triangular inequality, we observe that if $x \in A$ and $\varepsilon > 0$, there exist points $y \in B$ and $z \in C$ such that

$$d(x, B)+\delta(B, C) \geqq d(x, y)-\varepsilon+d(y, C) \geqq d(x, y)+d(y, z)-2\varepsilon$$
$$\geqq d(x, z)-2\varepsilon \geqq d(x, C)-2\varepsilon.$$

Since this inequality is satisfied for all $\varepsilon > 0$, we have

$$d(x, C) \leqq d(x, B)+\delta(B, C),$$

whence

$$\rho(A, C) = \sup_{x \in A} d(x, C) \leqq \delta(A, B)+\delta(B, C).$$

In this inequality, we can interchange A and C without changing the right-hand side, and so

$$\delta(A, C) \leqq \delta(A, B)+\delta(B, C).$$

Thus we can regard $\mathscr{F}'$ as a metric space, with δ as the metric.

Theorem 1. *Let X and Y be two metric spaces, $\mathscr{K}'$ the family of non-empty compact sets in Y, Γ a mapping of X into Y such that, for each x, $\Gamma x \neq \emptyset$. Then Γ is a continuous mapping of X into Y if and only if it is a single-valued continuous mapping of X into $\mathscr{K}'$.*

Proof. To begin with, we note that

$$\delta(\Gamma x, \Gamma x_0) \leqq \varepsilon \Leftrightarrow \left\{ \begin{array}{l} \Gamma x \subset B_\varepsilon(\Gamma x_0) \\ \Gamma x_0 \subset B_\varepsilon(\Gamma x) \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \Gamma x \subset B_\varepsilon(\Gamma x_0) \\ (\forall_{\Gamma x_0} y) : B_\varepsilon(y) \cap \Gamma x \neq \emptyset. \end{array} \right.$$

If Γ is a continuous mapping of X into Y, then to each $\varepsilon > 0$, there corresponds a number η such that

$$d(x, x_0) \leqq \eta \text{ implies that } \Gamma x \subset \mathring{B}_\varepsilon(\Gamma x_0) \subset B_\varepsilon(\Gamma x_0).$$

Moreover, since Γx_0 is compact, it contains points $y_1, y_2, \ldots, y_n$ such that the balls $B_{\varepsilon/2}(y_i)$ cover Γx_0. Therefore there is a number η' such that

$$d(x, x_0) \leqq \eta' \Rightarrow (\forall i) : B_{\frac{\varepsilon}{2}}(y_i) \cap \Gamma x \neq \varnothing \Rightarrow (\forall_{\Gamma x_0} y) : B_\varepsilon(y) \cap \Gamma x \neq \varnothing$$

and so, finally, we have

$$d(x, x_0) \leqq \min\{\eta, \eta'\} \quad \Rightarrow \quad \delta(\Gamma x, \Gamma x_0) \leqq \varepsilon.$$

Conversely, suppose that Γ is a single-valued continuous mapping of X into $\mathscr{K}'$. Let G be an open subset of Y containing Γx_0. Then by Theorem 3 of § 1, Chapter V, there exists a number ε such that $B_\varepsilon(\Gamma x_0) \subset G$ and so there exists a number η such that

$$x \in B_\eta(x_0) \quad \Rightarrow \quad \Gamma x \subset B_\varepsilon(\Gamma x_0) \subset G.$$

Hence Γ is u.s.c.

Moreover, if G is an open set meeting Γx_0, there exists a point y_0 in $\Gamma x_0 \cap G$ and an ε such that $B_\varepsilon(y_0) \subset G$ and hence there exists a number η such that

$$x \in B_\eta(x_0) \Rightarrow \Gamma x \cap B_\varepsilon(y_0) \neq \varnothing \Rightarrow \Gamma x \cap G \neq \varnothing.$$

Hence Γ is l.s.c.

COROLLARY. *If Γ is a continuous mapping defined in a compact metric space X, Γ is uniformly continuous: that is, to each $\varepsilon > 0$ there corresponds a number η such that, for each pair (x, x'), we have*

$$d(x, x') \leqq \eta \quad \Rightarrow \quad \delta(\Gamma x, \Gamma x') \leqq \varepsilon.$$

To prove this, it is sufficient to combine the result of the previous theorem with Heine's Theorem (cf. § 9, Chapter V).

REMARK. The Hausdorff metric enables us to topologise the family $\mathscr{F}'$ of non-empty closed sets in X. In the case in which X is not a metric space, this suggests the following more general problem: can we associate with $\mathscr{F}'$ a topological structure such that

(1) the study of the continuity of a mapping Γ with values in X is reduced to that of a single-valued mapping with values in $\mathscr{F}'$,

(2) the study of the convergence of a sequence (F_n) of closed sets of X is reduced to that of a sequence of elements of $\mathscr{F}'$?

In fact there exist several methods of topologising $\mathscr{F}'$, amongst which are one due to Vietoris and another due to Bourbaki. Nevertheless we should note the following points:

(1) In adopting this point of view, we can only hope for partial results concerning mappings; we must assume that $\Gamma x \neq \varnothing$ for each x and that Γ is semi-continuous in both senses; also the differences between elements and sets are not taken into account.

(2) Regarding the study of convergence of sequences (F_n), we have already indicated that this does not in general possess the characteristics of convergence of sequences of elements in a proper topological space. The introduction of a 'pseudo-topology' appears to lead to inextricable complications.

CHAPTER VII

MAPPINGS OF ONE VECTOR SPACE INTO ANOTHER

§ 1. Vector spaces

We now study certain special types of operations whose nature is of particular interest to algebraists.

I. An operation in a set X is called an **addition** if it is a correspondence in which to each pair (x, y) of $X \times X$ there corresponds an element of X, written $x+y$, such that

(1) $x+(y+z) = (x+y)+z$ (associativity),

(2) there exists an element 0 of X, called the **neutral element**, such that, for all x,

$$x+0 = 0+x = x,$$

(3) to each x there corresponds an element $-x$ of X, called the **inverse** of x, such that $x+(-x) = (-x)+x = 0$,

(4) $x+y = y+x$ (commutativity).

A pair consisting of a set X and an addition $+$ on X is called an **abelian group** (if condition (4) is omitted, then we call the pair a non-abelian group). We refer to the element $x+y$ as the *sum* of x and y.

The set X of integers, positive, negative or zero, together with ordinary addition $+$ forms an abelian group, with zero as the neutral element. Similarly the set X of strictly positive real numbers together with multiplication forms an abelian group having 1 as the neutral element. The set of displacements in the plane forms a non-abelian group, with the 'unit' displacement, which leaves everything unchanged, as the neutral element.

REMARK. An abelian group cannot contain more than one neutral element, for if e and e' are neutral elements, then $e' = e+e' = e$. Similarly an element x cannot have more than one inverse, for if a and b are inverses of x, we have

$$a = 0+a = (b+x)+a = b+(x+a) = b+0 = b.$$

II. In an abelian group $(X, +)$ a **scalar multiplication** is an operation in which to each pair consisting of a real number λ and an element x of X there corresponds an element of X (written λx) such that

(1) $\lambda(x+y) = \lambda x+\lambda y$,
(2) $(\lambda+\mu)x = \lambda x+\mu x$,
(3) $(\lambda\mu)x = \lambda(\mu x)$,
(4) $1x = x$.

A set X together with an addition and a scalar multiplication is called a **vector space** (over the real numbers). If addition and scalar multiplication are fixed and no confusion is possible, we shall refer to X itself as the vector space. The points of X will be called *vectors*.

EXAMPLE 1. $\mathbf{R}^n$ is a vector space, if we define addition and scalar multiplication by

$$x+y = (x^1, x^2, \ldots, x^n)+(y^1, y^2, \ldots y^n) = (x^1+y^1, x^2+y^2, \ldots, x^n+y^n),$$
$$\lambda x = \lambda(x^1, x^2, \ldots, x^n) = (\lambda x^1, \lambda x^2, \ldots, \lambda x^n)$$

and write

$$0 = (0, 0, \ldots, 0),$$
$$-x = -(x^1, x^2, \ldots, x^n) = (-x^1, -x^2, \ldots, -x^n).$$

It is easily verified that all the requirements for a vector space are fulfilled.

EXAMPLE 2. The set Φ of numerical functions defined on the segment $[0, 1]$ forms a vector space if, for $f, g \in \Phi$ we write

$$[f+g](x) = f(x)+g(x),$$
$$[\lambda f](x) = \lambda \cdot f(x),$$
$$[0](x) = 0,$$
$$[-f](x) = -f(x).$$

EXAMPLE 3. The space L_2 of numerical sequences (x_n), such that

$$\sum_{n=1}^{\infty} |x_n|^2 < +\infty$$

(real Hilbert space) is a vector space if we write

$$(x_n)+(y_n) = (x_n+y_n),$$
$$\lambda(x_n) = (\lambda x_n),$$
$$0 = (0, 0, \ldots),$$
$$-(x_n) = (-x_n).$$

We have $(\lambda x_n) \in L_2$, for

$$\sum_{n=1}^{\infty} |\lambda x_n|^2 = |\lambda|^2 \sum |x_n|^2 < +\infty;$$

also $(-x_n) \in L_2$, for

$$\sum |-x_n|^2 = \sum |x_n|^2 < +\infty.$$

To show that $x+y \in L_2$, we observe that

$$|x_n|^2 + |y_n|^2 - 2|x_n| \times |y_n| = (|x_n| - |y_n|)^2 \geqq 0.$$

Then

$$|x_n+y_n|^2 \leqq (|x_n|+|y_n|)^2 \leqq |x_n|^2+|y_n|^2+2|x_n|\times|y_n| \leqq 2|x_n|^2+2|y_n|^2$$

and so

$$\sum_{n=1}^{\infty}|x_n+y_n|^2 \leqq 2\sum_{n=1}^{\infty}|x_n|^2+2\sum_{n=1}^{\infty}|y_n|^2 < +\infty.$$

REMARK 1. $0x = 0$. For

$$x+0x = 1x+0x = (1+0)x = 1x = x$$

and so

$$0x = (-x+x)+0x = -x+(x+0x) = -x+x = 0.$$

REMARK 2. $(-1)x = -x$. For

$$(-1)x+x = (-1+1)x = 0x = 0$$

and so

$$(-1)x = (-1)x+[x+(-x)] = [(-1)x+x]+(-x) = 0+(-x) = -x.$$

REMARK 3. $\lambda x = 0$, $\lambda \neq 0$ are equivalent to $x = 0$, $\lambda \neq 0$. For if $x = 0$ and $\lambda \neq 0$, then, by Remark 2,

$$\lambda x = \lambda 0+\lambda[y+(-1)y] = (\lambda-\lambda)y = 0.$$

Conversely, if $\lambda x = 0$ and $\lambda \neq 0$, we have

$$x = \left(\frac{1}{\lambda}\lambda\right)x = \frac{1}{\lambda}(\lambda x) = \frac{1}{\lambda}0 = 0.$$

Let A and B be two non-empty subsets of a vector space X. Put

$$A+B = \{a+b \,/\, a \in A,\ b \in B\},$$
$$\lambda A = \{\lambda a \,/\, a \in A\},$$

where λ is a real number. Then $A+B$ and λA are non-empty subsets of X. We also write

$$A+\varnothing = \varnothing,$$
$$\lambda\varnothing = \varnothing,$$
$$-A = (-1)A,$$
$$A-B = A+(-B).$$

Theorem 1. *We have*

(1) $A+B = B+A$,
(2) $A+(B+C) = (A+B)+C$,
(3) $A+\{0\} = A$,
(4) $A+(-A) \supset \{0\}$ *if* $A \neq \emptyset$,
(5) $\lambda(A+B) = \lambda A+\lambda B$,
(6) $(\lambda+\mu)A \subset \lambda A+\mu A$,
(7) $\lambda\mu A = \lambda(\mu A)$,
(8) $1 \cdot A = A$.

Proof. These formulae are trivial if A or B is empty and are easily verified if $A \neq \emptyset$ and $B \neq \emptyset$.

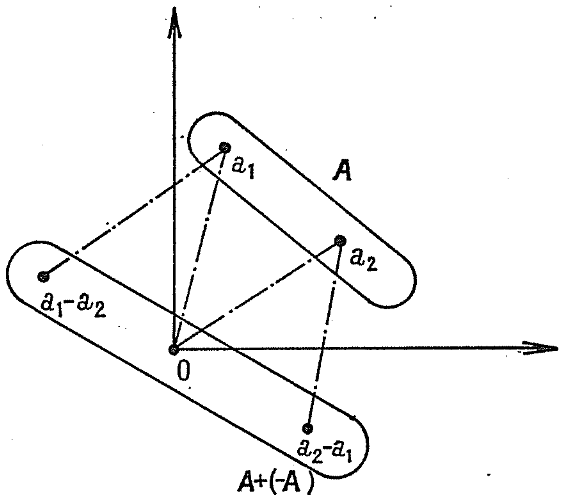


FIG. 31

In general, the set $\mathscr{P}(X)$ of subsets of X does not form a vector space; the inclusions (4) and (6) are not necessarily equalities, as we can see from figure 31, in which $X = \mathbf{R}^2$ and $A = \{a_1, a_2\}$. We have

$$A+(-A) \neq \{0\}$$

or

$$1A+(-1)A \neq (1-1)A$$

Theorem 2. $(A+B) \cap C = \emptyset$ *is equivalent to* $A \cap (C-B) = \emptyset$.

Proof. If $A \cap (C-B) \neq \emptyset$, we have

$$(\exists_A\, a)\, (\exists_B\, b)\, (\exists_C\, c) : a = c-b.$$

Then $a+b = c$ and $(A+B) \cap C \neq \emptyset$; thus we have shown that

$$(A+B) \cap C = \emptyset \Rightarrow A \cap (C-B) = \emptyset.$$

Replacing B by $-B$, A by C and C by A in this formula, we obtain

$$(C-B)\cap A=\varnothing \Rightarrow C\cap(A+B)=\varnothing.$$

Hence

$$(A+B)\cap C=\varnothing \Leftrightarrow A\cap(C-B)=\varnothing.$$

Theorem 3. *$A+B\subset C$ implies that $A\subset C-B$.*

Proof. If $A+B\subset C$, we have

$$(\forall_A a)\,(\forall_B b)\,(\exists_C c)\,:\,a+b=c$$

and so

$$(\forall_A a)\,(\exists_B b)\,(\exists_C c)\,:\,a=c-b,$$

whence

$$A\subset C-B.$$

§2. Linear mappings

Let X and Y be two vector spaces and let Γ be a mapping of X into Y such that $\Gamma x\neq\varnothing$ for at least one $x\in X$. We say that Γ is **linear** if

$$\begin{array}{lll}(1) & \left.\begin{array}{l}y\in\Gamma x\\ y'\in\Gamma x'\end{array}\right\} & \Rightarrow\quad y+y'\in\Gamma(x+x'),\\ (2) & y\in\Gamma x & \Rightarrow\quad \lambda y\in\Gamma(\lambda x).\end{array}$$

If Γ is a single-valued mapping f, these conditions become

$$\begin{array}{ll}(1') & f(x)+f(x')=f(x+x'),\\ (2') & \lambda f(x)=f(\lambda x).\end{array}$$

A numerical function which is linear is often called a **linear form.**

Example 1. Let $X=\mathbf{R}^2$, $Y=\mathbf{R}$ and consider the projection π_1 which maps the point (x^1,x^2) of $\mathbf{R}^2$ into the point x^1 of $\mathbf{R}$. We have

$$\pi_1 x+\pi_1 x'=x^1+x'^1=\pi_1(x+x'),$$
$$\lambda\pi_1 x=\lambda x^1=\pi_1(\lambda x).$$

Therefore π_1 is a single-valued linear mapping.

Example 2. Let Φ be the vector space formed by the numerical functions defined on $[0,1]$ and let Γ be the mapping of Φ into Φ which maps each function ϕ of Φ into the set $\Gamma\phi$ of its indefinite integrals; then

$$\Gamma\phi(t)=\left\{\int_t^a \phi(\xi)\,\mathrm{d}\xi+\lambda\;/\;\lambda\in\mathbf{R}\right\}\quad\text{if } \phi \text{ has an indefinite integral,}$$

$$\Gamma\phi(t)=\varnothing\qquad\text{otherwise.}$$

We can verify that this mapping is linear.

Theorem 1. *A mapping Γ is linear if and only if*

$$\begin{array}{ll} (1) & \Gamma(x+x') = \Gamma x+\Gamma x', \\ (2) & \Gamma(\lambda x) = \lambda\Gamma x \quad \text{if} \quad \lambda \neq 0, \\ (3) & 0 \in \Gamma 0. \end{array}$$

Proof. Clearly if Γ satisfies these conditions it is linear.

Conversely, suppose that Γ is a linear mapping. We shall prove the properties in the order (3), (2), (1).

To prove (3), we suppose that x is such that $\Gamma x \neq \emptyset$ and let y be an element of Γx. Then

$$0 = 0y \in \Gamma(0x) = \Gamma 0.$$

To prove (2), let λ be a real number not equal to zero and let x be an element of X; if $\Gamma x \neq \emptyset$, then it follows from the definition of linearity that

$$\lambda\Gamma x \subset \Gamma(\lambda x)$$

and this inclusion is also satisfied if $\Gamma x = \emptyset$. To show that the opposite inclusion also holds, we replace λ by $1/\lambda$ and x by λx in this formula and get

$$\frac{1}{\lambda}\Gamma(\lambda x) \subset \Gamma\left(\frac{1}{\lambda}\lambda x\right) = \Gamma x.$$

It follows that

$$\Gamma(\lambda x) \subset \lambda\Gamma x.$$

Therefore

$$\Gamma(\lambda x) = \lambda\Gamma x.$$

To prove (1), let x and x' be elements of X; if $\Gamma x \neq \emptyset$, $\Gamma x' \neq \emptyset$, then it follows from the definition of linearity that

$$\Gamma x+\Gamma x' \subset \Gamma(x+x').$$

If $\Gamma x = \emptyset$ or $\Gamma x' = \emptyset$, this inclusion is still valid. To show that the opposite inclusion also holds, we replace x by $x+x'$ and x' by $-x'$ in this formula and get

$$\Gamma(x+x')+\Gamma(-x') \subset \Gamma x.$$

Using (2) and Theorem 3 of § 1, we now get

$$\Gamma(x+x') \subset \Gamma x-\Gamma(-x') = \Gamma x+(-1)^2\,\Gamma x' = \Gamma x+\Gamma x'.$$

Therefore

$$\Gamma x+\Gamma x' = \Gamma(x+x').$$

COROLLARY. *A necessary and sufficient condition for a linear mapping Γ to be single-valued is $\Gamma 0 = \{0\}$.*

Proof. By the theorem, we have $0 \in \Gamma 0$; if Γ is single-valued, then clearly $\Gamma 0 = \{0\}$.

Conversely, if $\Gamma 0 = \{0\}$, then, by the theorem,

$$\Gamma x - \Gamma x = \Gamma(x-x) = \Gamma 0 = \{0\}.$$

If a and b are two elements of Γx, then $a-b = 0$ and so $a = b$; hence Γ is single-valued.

Theorem 2. *The inverse Γ^- of a linear mapping Γ of X into Y is a linear mapping of Y into X.*

Proof. If Γ is linear, then

$$\left.\begin{matrix} x \in \Gamma^- y \\ x' \in \Gamma^- y' \end{matrix}\right\} \Rightarrow \left.\begin{matrix} y \in \Gamma x \\ y' \in \Gamma x' \end{matrix}\right\} \Rightarrow y+y' \in \Gamma(x+x') \Rightarrow x+x' \in \Gamma^-(y+y')$$

and also

$$x \in \Gamma^- y \Rightarrow y \in \Gamma x \Rightarrow \lambda y \in \Gamma(\lambda x) \Rightarrow \lambda x \in \Gamma^-(\lambda y).$$

Theorem 3. *If Γ_1 is a linear mapping of X into Y and Γ_2 is a linear mapping of Y into Z, the composition product $\Gamma = \Gamma_2 \cdot \Gamma_1$ is a linear mapping of X into Z.*

Proof. We have

$$\begin{aligned} &\left\{\begin{matrix} z \in \Gamma_2 \cdot \Gamma_1 x \\ z' \in \Gamma_2 \cdot \Gamma_1 x' \end{matrix}\right. \Rightarrow \left\{\begin{matrix} z \in \Gamma_2 y,\ y \in \Gamma_1 x \\ z' \in \Gamma_2 y',\ y' \in \Gamma_1 x' \end{matrix}\right. \Rightarrow \\ &\qquad \Rightarrow \left\{\begin{matrix} z+z' \in \Gamma_2(y+y') \\ y+y' \in \Gamma_1(x+x') \end{matrix}\right. \Rightarrow z+z' \in \Gamma_2 \cdot \Gamma_1(x+x') \end{aligned}$$

and also

$$z \in \Gamma_2 \cdot \Gamma_1 x \Rightarrow z \in \Gamma_2 y,\ y \in \Gamma_1 x \Rightarrow \left\{\begin{matrix} \lambda z \in \Gamma_2(\lambda y) \\ \lambda y \in \Gamma_1(\lambda x) \end{matrix}\right. \Rightarrow \lambda z \in \Gamma_2 \cdot \Gamma_1(\lambda x).$$

Theorem 4. *If Γ_1 and Γ_2 are two linear mappings of X into Y, then $\Gamma = \Gamma_1 \cap \Gamma_2$ is a linear mapping of X into Y.*

Proof. We have

$$\begin{aligned} &\left\{\begin{matrix} z \in \Gamma x \\ z' \in \Gamma x' \end{matrix}\right. \Rightarrow \left\{\begin{matrix} z \in \Gamma_1 x,\ z \in \Gamma_2 x \\ z' \in \Gamma_1 x',\ z' \in \Gamma_2 x' \end{matrix}\right. \Rightarrow \\ &\qquad \Rightarrow \left\{\begin{matrix} z+z' \in \Gamma_1(x+x') \\ z+z' \in \Gamma_2(x+x') \end{matrix}\right. \Rightarrow z+z' \in \Gamma(x+x') \end{aligned}$$

and also

$$z \in \Gamma x \Rightarrow \left\{\begin{matrix} z \in \Gamma_1 x \\ z \in \Gamma_2 x \end{matrix}\right. \Rightarrow \left\{\begin{matrix} \lambda z \in \Gamma_1(\lambda x) \\ \lambda z \in \Gamma_2(\lambda x) \end{matrix}\right. \Rightarrow \lambda z \in \Gamma(\lambda x).$$

Theorem 5. *If Γ_1 and Γ_2 are two linear mappings of X into Y, the mapping $\Gamma = \Gamma_1+\Gamma_2$ (defined by $\Gamma x = \Gamma_1 x+\Gamma_2 x$) is linear.*

Proof. We have

$$\begin{cases} z \in \Gamma x \\ z' \in \Gamma x' \end{cases} \Rightarrow \begin{cases} z = y_1+y_2;\ y_1 \in \Gamma_1 x\ ;\ y_2 \in \Gamma_2 x \\ z' = y_1'+y_2';\ y_1' \in \Gamma_1 x';\ y_2' \in \Gamma_2 x' \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} z+z' = (y_1+y_1')+(y_2+y_2'), \\ y_1+y_1' \in \Gamma_1(x+x'), \\ y_2+y_2' \in \Gamma_2(x+x'), \end{cases} \Rightarrow z+z' \in \Gamma(x+x')$$

and also

$$z \in \Gamma x \Rightarrow \begin{cases} z = y_1+y_2 \\ y_1 \in \Gamma_1 x \\ y_2 \in \Gamma_2 x \end{cases} \Rightarrow \begin{cases} \lambda z = \lambda y_1+\lambda y_2 \\ \lambda y_1 \in \Gamma_1(\lambda x_1) \\ \lambda y_2 \in \Gamma_2(\lambda x_2) \end{cases} \Rightarrow \lambda z \in \Gamma(\lambda x).$$

§ 3. Linear varieties, cones, convex sets

DEFINITION 1. A subset E of a vector space X is called a **vector subspace** if

$$\begin{array}{lll} (1) & x \in E,\ y \in E & \Rightarrow \quad x+y \in E, \\ (2) & x \in E,\ \lambda \in \mathbf{R} & \Rightarrow \quad \lambda x \in E. \end{array}$$

It follows at once from the definition that $\varnothing$, $\{0\}$ and X are vector subspaces; moreover, by putting $\lambda = 0$ in (2), we see that every non-empty vector subspace contains the point 0. Conditions (1) and (2) can be replaced by the single condition

$$(1') \begin{cases} x, y \in E \\ \lambda, \mu \in \mathbf{R} \end{cases} \Rightarrow \quad \lambda x+\mu y \in E.$$

Two vector subspaces E_1 and E_2 are said to be **supplementary** if

$$E_1+E_2 = X, \qquad E_1 \cap E_2 = \{0\}.$$

If E_1 and E_2 are two supplementary subspaces, each point x of X can be written in the form $x = x_1+x_2$, where $x_1 \in E_1$ and $x_2 \in E_2$. Moreover, this decomposition is unique, for

$$\begin{cases} x = x_1+x_2 = y_1+y_2 \\ x_1, y_1 \in E_1 \\ x_2, y_2 \in E_2 \end{cases} \Rightarrow \begin{cases} x_1-y_1 = x_2-y_2 \\ x_1-y_1 \in E_1 \\ x_2-y_2 \in E_2 \end{cases} \Rightarrow$$

$$\Rightarrow x_1-y_1 = x_2-y_2 = 0 \Rightarrow \begin{cases} x_1 = y_1 \\ x_2 = y_2. \end{cases}$$

A set of the form

$$D_a = \{x \mid x = \lambda a,\ \lambda \in \mathbf{R}\}$$

where a is a point of X different from 0, is called a **straight line through** 0; clearly every straight line through 0 is a vector subspace. If P is a vector subspace supplementary to a straight line D_a through 0, then P is called a **plane through** 0.

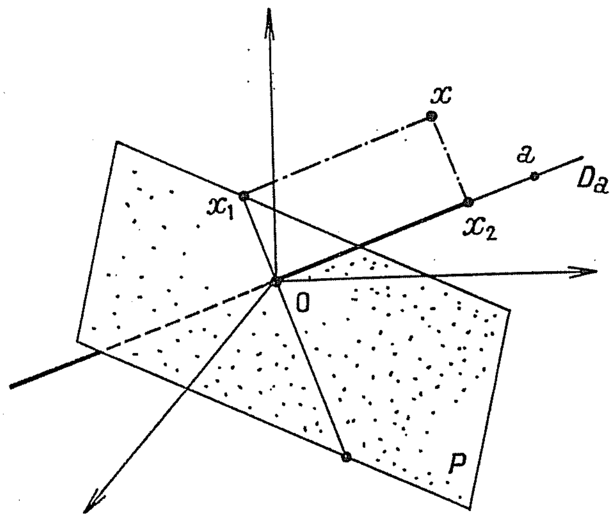


FIG. 32

Theorem 1. *If f is a numerical linear function which is not identically zero, then the set $E_f = \{x \mid f(x) = 0\}$ is a plane through 0; conversely, every plane through 0 is a set of this form.*

Proof. We first show that E_f is a vector subspace, as follows:

(1) $x \in E_f,\ y \in E_f \Rightarrow f(x+y) = f(x)+f(y) = 0+0 = 0 \Rightarrow x+y \in E_f,$
(2) $x \in E_f \Rightarrow f(\lambda x) = \lambda f(x) = \lambda 0 = 0 \Rightarrow \lambda x \in E_f.$

Since f is not identically zero in X, there exists a point a of X such that $f(a) \neq 0$; we now show that the straight line D_a through 0 satisfies

$$D_a + E_f = X, \qquad D_a \cap E_f = \{0\}.$$

The second condition follows at once from the fact that $f(\lambda a) = \lambda f(a) = 0$ only if $\lambda = 0$. To prove the first, we put $\lambda = f(x) / f(a)$; then we have

$$f(x-\lambda a) = f(x) - \lambda f(a) = 0 \Rightarrow x - \lambda a = b \in E_f \Rightarrow x = \lambda a + b,\ b \in E_f.$$

Hence E_f is a plane through 0.

Conversely, suppose that P is a plane through 0 and that D_a is a straight line through 0 supplementary to P. Every point x of X can be written in the form $x = \lambda a + b$, where $b \in P$. Writing $\lambda = f(x)$, we obtain a numerical function of x; and this function is linear, for

(1) $x + x' = [f(x)+f(x')]a + b + b' = f(x+x')a + b''; \quad b, b', b'' \in P,$
(2) $\lambda x = [\lambda f(x)]a + \lambda b = f(\lambda x)a + b'; \quad b, b' \in P.$

Thus we have $f(x+x') = f(x)+f(x')$ and $f(\lambda x) = \lambda f(x)$. Moreover, f is not identically zero, since $f(a) = 1$; finally we have

$$P = \{x \,/\, f(x) = 0\}.$$

There are several different generalisations of the idea of vector subspace.

DEFINITION 2. A subset A of a vector space X is called a **linear variety** if

$$\begin{cases} x, x' \in A, \\ \lambda, \lambda' \in \mathbf{R}, \\ \lambda+\lambda' = 1, \end{cases} \Rightarrow \quad \lambda x+\lambda' x' \in A.$$

The set $\varnothing$, the set X and the set $\{a\}$ consisting of a single point a, are all linear varieties.

Theorem 2. *If E is a vector subspace and x_0 is a point of X, the set $E+x_0$ is a linear variety. Conversely, every linear variety is of the form $E+x_0$ for some vector subspace E and some point x_0.*

Proof. $E+x_0$ is a linear variety, for if λ and λ' are real numbers such that $\lambda+\lambda' = 1$, we have

$$x, x' \in E+x_0 \;\Rightarrow\; \begin{cases} x-x_0 \in E \\ x'-x_0 \in E \end{cases} \Rightarrow$$

$$\Rightarrow\; (\lambda x+\lambda' x')-x_0 = \lambda(x-x_0)+\lambda'(x'-x_0) \in E \;\Rightarrow\; \lambda x+\lambda' x' \in E+x_0.$$

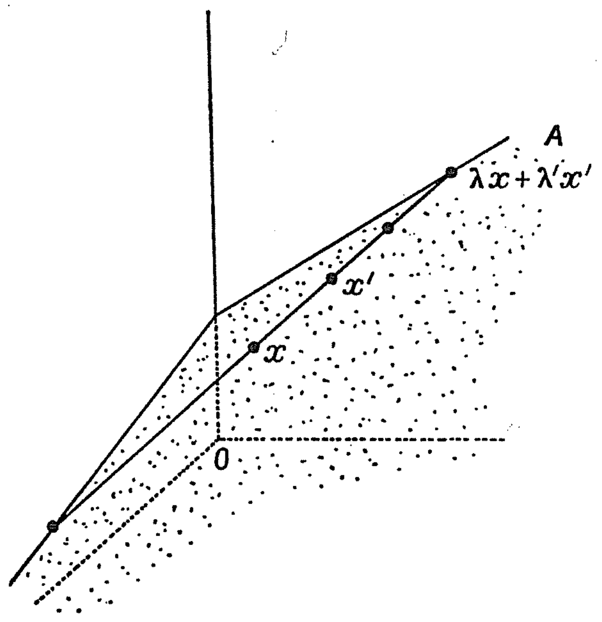


FIG. 33

Suppose now that A is any linear variety; we shall prove that there exists a vector subspace E and a point x_0 such that $A = E+x_0$.

If $A = \varnothing$, then clearly $A = \varnothing = \varnothing+x_0$; if $A \neq \varnothing$, let x_0 be a point in A and write $E = A-x_0$. Then E is a vector subspace, for

$$(1)\quad x \in E \;\Rightarrow\; \begin{cases} x = a-x_0 \\ a \in A \end{cases} \Rightarrow$$
$$\Rightarrow\; \lambda x = \lambda a-\lambda x_0 = \lambda a+(1-\lambda)x_0-x_0 \in A-x_0 = E,$$

$$(2)\quad x, x' \in E \;\Rightarrow\; \begin{cases} x = a-x_0 \\ x' = a'-x_0 \\ a, a' \in A \end{cases} \Rightarrow\; \frac{x+x'}{2} = \frac{a+a'}{2}-x_0 \in A-x_0$$
$$\Rightarrow\; x+x' \in E.$$

Thus $A = E+x_0$, where E is a vector subspace.

If A is a linear variety such that $A = E_1+x_1$ and $A = E_2+x_2$, then $E_1 = E_2$. For $x_1-x_2 \in E_2$, since

$$x_1 = 0+x_1 \in A = E_2+x_2.$$

Then

$$E_1 = E_2+(x_2-x_1) \subset E_2+E_2 \subset E_2.$$

By symmetry, we also have $E_2 \subset E_1$ and so $E_1 = E_2$.

DEFINITION 3. The space E such that $A = E+x_0$ (necessarily unique, as has just been proved) is called the **subspace parallel to** A. If two varieties A and B have the same parallel subspace E, they are said to be **parallel** to one another.

A linear variety D which has a straight line D_a through 0 as a parallel subspace is called a **straight line**; a linear variety P which has a plane through 0 as a parallel subspace is called a **plane**.

REMARK 1. *Two varieties A and B which are parallel to one another are either coincident or disjoint.*

Proof. Suppose that $A \cap B \neq \varnothing$; if $A = E+a$ and $B = E+b$ $(a \in A;$ $b \in B)$ there exist points e_1 and e_2 in E such that $e_1+a = e_2+b$; then

$$a = (e_2-e_1)+b \in E+b = B.$$

Since a is an arbitrary point of A, we have $A \subset B$; by symmetry $B \subset A$ and so $A = B$.

REMARK 2. *Every plane is of the form $E_f^\alpha = \{x \,/\, f(x) = \alpha\}$ for some $\alpha \in \mathbf{R}$ and some numerical linear function f not identically zero; conversely, every set of this form is a plane.*

Proof. We have

$$x \in E_f + x_0 \Leftrightarrow x - x_0 \in E_f \Leftrightarrow f(x) = f(x_0)$$

and so

$$E_f + x_0 = \{x \,/\, f(x) = f(x_0)\}.$$

Conversely, let E_f^α be the set $\{x \,/\, f(x) = \alpha\}$ and let a be such that $f(a) \neq 0$.

Putting $x_0 = \dfrac{\alpha}{f(a)} a$, we get

$$x \in E_f^\alpha \Leftrightarrow f(x) - f(x_0) = 0 \Leftrightarrow x - x_0 \in E_f$$

and therefore $E_f^\alpha = E_f + x_0$ is a plane.

REMARK 3. *Two distinct points determine one and only one straight line.*

Proof. If $a \neq b$, the points a and b determine the straight line

$$D(a, b) = \{x \,/\, x = \lambda a + \mu b \,/\, \lambda, \mu \in \mathbf{R}, \lambda + \mu = 1\}.$$

Geometrically, we can see that a set A is a linear variety if, for each pair a, b of points of A, the straight line $D(a, b)$ is contained in A.

DEFINITION 4. We say that a set A is a **cone** if

$$x \in A, \lambda \geqq 0 \Rightarrow \lambda x \in A.$$

Every vector subspace is a cone.

EXAMPLE 1. The set $D = \{x \,/\, x = \lambda a, \lambda \geqq 0\}$, where $a \in X$ and $a \neq 0$, is called the **half-line** through a.

Geometrically, we can see that a set A is a cone if, for each point a in A, the half-line D_a is contained in A.

EXAMPLE 2. A set of the form

$$H_f = \{x \,/\, f(x) \geqq 0\},$$

where f is a linear numerical function not identically zero, is a cone; such a set is called a **half-space**.

EXAMPLE 3. Let I be a subset of $\{1, 2, \ldots, n\}$ and let Q_I be the subset of $\mathbf{R}^n$ defined by $Q_I = \{x \,/\, x = (x^1, x^2, \ldots, x^n), x^i \geqq 0 \text{ if } i \in I, x^i \leqq 0 \text{ if } i \notin I\}$. Then Q_I is a cone. There are 2^n such sets in $\mathbf{R}^n$.

DEFINITION 5. A set C is said to be **convex** if

$$(1)\quad \left.\begin{array}{l} x, x' \in C \\ \lambda, \lambda' \geqq 0 \\ \lambda+\lambda' = 1 \end{array}\right\} \Rightarrow \lambda x+\lambda' x' \in C.$$

Every linear variety is convex; in particular, $\emptyset$, $\{a\}$, X are convex.

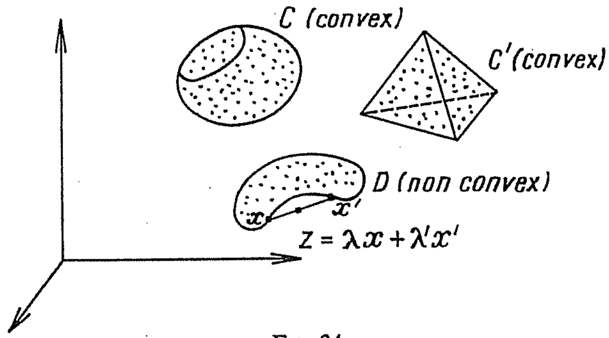


FIG. 34

EXAMPLE 1. Let f be a numerical linear function not identically zero. We can see immediately that the following sets are convex:

the *plane* $E_f^\alpha = \{x \,/\, f(x) = \alpha\}$,
the *open half-space* $H_f'^\alpha = \{x \,/\, f(x) < \alpha\}$,
the *closed half-space* $H_f^\alpha = \{x \,/\, f(x) \leqq \alpha\}$.

EXAMPLE 2. The following sets, which are called **linear intervals**, are convex:

the **linear segment** $[a, b] = \{\lambda a+\mu b \,/\, \lambda, \mu \geqq 0, \lambda+\mu = 1\}$,
$[a, b[\; = \{\lambda a+\mu b \,/\, \lambda > 0, \mu \geqq 0, \lambda+\mu = 1\}$,
$]a, b] = \{\lambda a+\mu b \,/\, \lambda \geqq 0, \mu > 0, \lambda+\mu = 1\}$,
$]a, b[\; = \{\lambda a+\mu b \,/\, \lambda > 0, \mu > 0, \lambda+\mu = 1\}$.

Geometrically we see that a set C is convex if for each pair (a, b) of points of C, the linear segment $[a, b]$ is contained in C.

EXAMPLE 3. A set K such that

$$\left.\begin{array}{r} x, x' \in K \\ \lambda \geqq 0,\ \lambda' \geqq 0 \end{array}\right\} \Rightarrow \lambda x+\lambda' x' \in K$$

is called a **convex cone.** A convex cone is a cone and is a convex set; conversely if a set K is a cone and is convex, it is a convex cone, for

$$\left.\begin{array}{l} x, x' \in K \\ \lambda, \lambda' > 0 \end{array}\right\} \Rightarrow \frac{\lambda}{\lambda+\lambda'} x+\frac{\lambda'}{\lambda+\lambda'} x' = y \in K \Rightarrow \lambda x+\lambda' x' = (\lambda+\lambda')\, y \in K.$$

EXAMPLE 4. (Calculus of probabilities.) The probability distributions $(p_i \,/\, i \in I)$ on a given set $A = \{a_i \,/\, i \in I\}$ form a convex set in the linear space of functions defined on A.

REMARK. We shall denote by $\mathbf{P}_n$ the set formed by the elements $p = (p_1, p_2, \ldots, p_n)$ of $\mathbf{R}^n$ such that

$$p_1, p_2, \ldots, p_n \geqq 0, \quad p_1+p_2+\ldots+p_n = 1.$$

Condition (1) defining a convex set C is equivalent to the following:

(1′) *For any positive integer n, we have*

$$\left.\begin{array}{l} x_1, x_2, \ldots, x_n \in C \\ (p_1, p_2, \ldots, p_n) \in \mathbf{P}_n \end{array}\right\} \Rightarrow p_1x_1+p_2x_2+\ldots+p_nx_n \in C.$$

Clearly (1′) $\Rightarrow$ (1), for (1′) reduces to (1) when $n = 2$. We now show that (1) $\Rightarrow$ (1′). If (1) is satisfied, then clearly (1′) is also satisfied for $n = 1$. Suppose that (1′) is satisfied for $n = k-1$; then

$$\sum_{i=1}^{k} p_ix_i = \sum_{i=1}^{k-1} p_ix_i + p_kx_k =$$

$$= \Big(\sum_{i=1}^{k-1} p_i\Big)\left(\frac{p_1}{\sum_1^{k-1} p_i}\, x_1 + \frac{p_2}{\sum_1^{k-1} p_i}\, x_2 + \ldots + \frac{p_{k-1}}{\sum_1^{k-1} p_i}\, x_{k-1}\right) + p_kx_k \in C$$

and so (1′) is satisfied for $n = k$. Therefore, by induction, (1′) is satisfied for all positive integers n.

Theorem 3. *If A and B are two convex sets (resp. vector subspaces, linear varieties, cones) the intersection $A \cap B$ is a convex set (resp. a vector subspace, a linear variety, a cone).*

Proof. Suppose that A and B are two convex sets. If $\lambda, \lambda' \geqq 0$ and $\lambda+\lambda' = 1$, we have

$$\left\{\begin{array}{l} x \in A \cap B \\ x' \in A \cap B \end{array}\right. \Rightarrow \left\{\begin{array}{l} x \in A,\ x' \in A \\ x \in B,\ x' \in B \end{array}\right. \Rightarrow \left\{\begin{array}{l} \lambda x+\lambda' x' \in A \\ \lambda x+\lambda' x' \in B \end{array}\right. \Rightarrow \lambda x+\lambda' x' \in A \cap B$$

and so $A \cap B$ is a convex set.

In the other cases, analogous proofs can easily be constructed.

Theorem 4. *If A and B are two convex sets (resp. vector subspaces, linear varieties, cones) their sum $A+B$ is a convex set (resp. a vector subspace, a linear variety, a cone).*

Proof. Suppose that A and B are two convex sets. If $\lambda, \lambda' \geqq 0$ and $\lambda+\lambda' = 1$, we have

$$\begin{cases} x \in A+B \\ x' \in A+B \end{cases} \Rightarrow \begin{cases} x = a+b \\ x' = a'+b' \\ a, a' \in A \\ b, b' \in B \end{cases} \Rightarrow \begin{cases} \lambda x+\lambda' x' = (\lambda a+\lambda' a')+(\lambda b+\lambda' b') \\ \lambda a+\lambda' a' \in A \\ \lambda b+\lambda' b' \in B \end{cases}$$

and so $A+B$ is a convex set.

In the other cases, analogous proofs can easily be constructed.

Theorem 5. *Let Γ be a linear mapping of X into Y; if A is a convex subset (resp. vector subspace, linear variety, cone) in X, the image ΓA is a convex subset (resp. vector subspace, linear variety, cone) in Y.*

Proof. Suppose that A is a convex set. If $\lambda, \lambda' \geqq 0$ and $\lambda+\lambda' = 1$, we have

$$\begin{cases} y \in \Gamma A \\ y' \in \Gamma A \end{cases} \Rightarrow \begin{cases} y \in \Gamma a, \ a \in A \\ y' \in \Gamma a', a' \in A \end{cases} \Rightarrow \begin{cases} \lambda y \in \Gamma(\lambda a) \\ \lambda' y' \in \Gamma(\lambda' a') \end{cases} \Rightarrow$$
$$\Rightarrow \begin{cases} \lambda y+\lambda' y' \in \Gamma(\lambda a+\lambda' a') \\ \lambda a+\lambda' a' \in A \end{cases} \Rightarrow \lambda y+\lambda' y' \in \Gamma A$$

and so ΓA is a convex set.

In the other cases, analogous proofs can easily be constructed.

COROLLARY. *If $\lambda \in \mathbf{R}$ and A is a convex set (resp. vector subspace, linear variety, cone), then λA is a convex set (resp. vector subspace, linear variety, cone).*

This follows immediately when we apply Theorem 5 to the linear mapping Γ defined by $\Gamma x = \{\lambda x\}$.

Theorem 6. *Let Γ be a linear mapping of X into Y; if A is a convex subset (resp. vector subspace, linear variety, cone) in Y, then $\Gamma^- A$ and $\Gamma^+ A$ are convex subsets (resp. vector subspaces, linear varieties, cones) in X.*

Proof. Let A be a convex set; if Γ is a linear mapping of X into Y, then Γ^- is a linear mapping of Y into X and so, by Theorem 5, $\Gamma^- A$ is a convex set.

If $\lambda, \lambda' > 0$ and $\lambda+\lambda' = 1$, then

$$\begin{cases} x \in \Gamma^+ A \\ x' \in \Gamma^+ A \end{cases} \Rightarrow \begin{cases} \Gamma x \subset A \\ \Gamma x' \subset A \end{cases} \Rightarrow \begin{cases} \Gamma(\lambda x) = \lambda \Gamma x \subset \lambda A \\ \Gamma(\lambda' x') = \lambda' \Gamma x' \subset \lambda' A. \end{cases} \Rightarrow$$
$$\Rightarrow \Gamma(\lambda x+\lambda' x') \subset \lambda A+\lambda' A \subset A \Rightarrow \lambda x+\lambda' x' \in \Gamma^+ A.$$

If $\lambda+\lambda' = 1$ and λ or λ' is zero, the same property holds trivially. Therefore $\Gamma^+ A$ is a convex set.

In the other cases, analogous proofs can easily be constructed.

§ 4. Dimension of a convex set

Let A be a subset of a vector space X. The intersection of the convex sets containing A is a convex set $[A]$, called the **convex closure** (or **convex cover**) of A. The intersection of the vector subspaces containing A is a vector subspace s $[A]$, called the **spatial closure** of A. The intersection of the linear varieties containing A is a linear variety lin $[A]$, called the **linear closure** of A. The intersection of the cones containing A is a cone k $[A]$, called the **conical closure** of A. The intersection of the convex cones containing A is a convex cone kc $[A]$, called the **convex conical closure** of A.

The correspondence $A \to [A]$ is a closure operation (see Chapter I, § 7), for

$$\begin{aligned}&[A] \supset A,\\ &A \supset B \Rightarrow [A] \supset [B],\\ &[[A]] = [A],\\ &[\emptyset] = \emptyset.\end{aligned}$$

It can easily be verified that similar properties hold for s $[A]$, lin $[A]$, k $[A]$, kc $[A]$.

Theorem 1. *Let A be the set $\{a_i \,/\, i \in I\}$. Then*

$$\begin{aligned}[A] &= \Big\{\sum_{i\in J} p_i a_i \,/\, J \text{ finite} \subset I;\ \sum_{i\in J} p_i = 1,\ p_i \geqq 0 \text{ for all } i \text{ in } J\Big\}\\ \text{s}\,[A] &= \Big\{\sum_{i\in J} \lambda_i a_i \,/\, J \text{ finite} \subset I;\ \lambda_i \in \mathbf{R} \text{ for all } i\Big\},\\ \text{lin}\,[A] &= \Big\{\sum_{i\in J} \lambda_i a_i \,/\, J \text{ finite} \subset I;\ \sum \lambda_i = 1\Big\},\\ \text{k}\,[A] &= \{\lambda a_i \,/\, \lambda \geqq 0,\ i \in I\} = \bigcup_{\lambda \geqq 0} \lambda A,\\ \text{kc}\,[A] &= \Big\{\sum_{i\in J} \lambda_i a_i \,/\, J \text{ finite} \subset I;\ \lambda_i \geqq 0 \text{ for all i in } J\Big\}.\end{aligned}$$

Proof. We shall prove the first formula, the proofs of the others being similar. We write

$$B = \Big\{\sum_{i\in J} p_i a_i \,/\, J \text{ finite} \subset I,\ \sum_{i\in J} p_i = 1,\ p_i \geqq 0 \text{ for all } i\Big\}.$$

If C is a convex set containing A, then, by the remark in § 3, we have $C \supset B$ and so

$$[A] \supset B.$$

To prove that $[A] = B$, it is therefore sufficient to verify that B is a convex set containing A. By definition $B \supset A$ and, since

$$\begin{cases} b = \sum_{i \in J} p_i a_i \in B, \\ b' = \sum_{i \in K} p'_i a'_i \in B, \\ \lambda, \lambda' \geqq 0, \ \lambda + \lambda' = 1, \end{cases}$$

imply that

$$\begin{cases} \lambda b + \lambda' b' = \sum_{i \in J} (\lambda p_i) a_i + \sum_{i \in K} (\lambda' p'_i) a'_i, \\ \sum_{i \in J} (\lambda p_i) + \sum_{i \in K} (\lambda' p'_i) = \lambda + \lambda' = 1, \end{cases}$$

we have $\lambda b + \lambda' b' \in B$ and so B is convex.

Given a set A, a plane E_f^α such that $f(x) \geqq \alpha$ for all x in A and $f(x) = \alpha$ for at least one x in A, is called a **plane of support** of A. We shall prove the following corollary to Theorem 1.

COROLLARY. *If P is a plane of support of A, then $P \cap [A] = [P \cap A]$.*

Proof. We have $f(a) \geqq \alpha$ for all a in A; if $x \in P \cap [A]$, then

$$\begin{cases} x = p_1 a_1 + p_2 a_2 + \dots + p_n a_n \in P, \\ a_1, a_2, \dots, a_n \in A, \\ (p_1, p_2, \dots, p_n) \in \mathbf{P}_n. \end{cases}$$

For all i, we have $f(a_i) = \alpha$; otherwise we should have $f(a_i) > \alpha$ for at least one i and then

$$f(x) = p_1 f(a_1) + p_2 f(a_2) + \dots + p_n f(a_n) > p_1 \alpha + p_2 \alpha + \dots + p_n \alpha = \alpha$$

so that x would not belong to P. Hence all the a_i belong to $P \cap A$; therefore $x \in [P \cap A]$.

Conversely, it is easily proved that if $x \in [P \cap A]$, then $x \in P \cap [A]$. Hence $P \cap [A] = [P \cap A]$.

Theorem 2. *We have*

$$\begin{aligned} &[\mathrm{lin}\,[A]] = \mathrm{lin}\,[[A]] = \mathrm{lin}\,[A], \\ &\mathrm{k}\,[[A]] = [\mathrm{k}\,[A]] = \mathrm{kc}\,[A]. \end{aligned}$$

Proof. We shall prove only that $\mathrm{k}\,[[A]] = \mathrm{kc}\,[A]$, the other identities being immediate. Because of Theorem 4 of § 7, Chapter I, it is sufficient to show that $\mathrm{k}\,[[A]]$ is a convex set. Let $x = \lambda a$ and $x' = \lambda' a'$ be two points of $\mathrm{k}\,[[A]]$, where $\lambda, \lambda' \geqq 0$ and $a, a' \in [A]$; if $(p, p') \in \mathbf{P}_2$, then

$$px+p'x' = p\lambda a+p'\lambda'a' = (p\lambda+p'\lambda')\left(\frac{p\lambda}{p\lambda+p'\lambda'}a+\frac{p'\lambda'}{p\lambda+p'\lambda'}a'\right) \in \mathrm{k}[[A]]$$

and so k $[[A]]$ is convex.

We say that the points $x_1, x_2, \ldots, x_n$ of a vector space X are **linearly independent** if

$$\lambda_1 x_1+\lambda_2 x_2+\ldots+\lambda_n x_n = 0 \text{ implies that } \lambda_1 = \lambda_2 = \ldots = \lambda_n = 0.$$

If $x_1, x_2, \ldots, x_n$ are not linearly independent, then we say that they are **linearly dependent.**

A vector subspace E is said to have **dimension** n if in E there exist n linearly independent points and if any set of more than n distinct points of E is linearly dependent. We say that a linear variety V has dimension n if the vector subspace E to which it is parallel has dimension n and we say that a convex set C has dimension n if the linear variety lin $[C]$ has dimension n.

Theorem 3. *Let $e_1, e_2, \ldots, e_n$ be linearly independent points in a vector subspace E; then a necessary and sufficient condition for E to have dimension n is that each point of E can be expressed in the form*

$$x = \lambda_1 e_1+\lambda_2 e_2+\ldots+\lambda_n e_n;\ \lambda_1, \lambda_2, \ldots \lambda_n \in \mathbf{R}.$$

Proof. If E has dimension n, there exist $n+1$ numbers $\mu, \mu_1, \mu_2, \ldots \mu_n$, not all zero, such that

$$\mu x+\mu_1 e_1+\mu_2 e_2+\ldots+\mu_n e_n = 0.$$

Since the e_i are linearly independent, μ is not zero.

Putting $\lambda_i = -\dfrac{\mu_i}{\mu}$, we have

$$x = -\frac{\mu_1}{\mu}e_1-\frac{\mu_2}{\mu}e_2-\ldots-\frac{\mu_n}{\mu}e_n = \lambda_1 e_1+\lambda_2 e_2+\ldots+\lambda_n e_n.$$

Suppose conversely that each point of E can be expressed in this form; we shall prove that E has dimension n. Assuming that this is not the case, there is an integer $k > 0$ such that there exist $n+k$ linearly independent points $a_1, a_2, \ldots, a_{n+k}$. By hypothesis, we have $a_1 = \lambda_1 e_1+\lambda_2 e_2+\ldots+\lambda_n e_n$ for some λ_i, at least one of which is non-zero. Without loss of generality we can suppose that $\lambda_1 \neq 0$; then

$$e_1 = \frac{1}{\lambda_1}a_1-\frac{\lambda_2}{\lambda_1}e_2-\ldots-\frac{\lambda_n}{\lambda_1}e_n$$

and so each point x of E can be expressed linearly in terms of $a_1, e_2, e_3, \ldots, e_n$. By a similar argument we can show that each point x of E can be expressed linearly in terms of $a_1, a_2, e_3, \ldots, e_n$ and we can continue this process until each of the e_i has been replaced by an a_i. Thus each point x can be expressed linearly in terms of $a_1, a_2, \ldots, a_n$. In particular this is true for $a_{n+1}, a_{n+2}, \ldots, a_{n+k}$; but, by hypothesis, $a_1, a_2, \ldots, a_{n+k}$ are linearly independent and so we have a contradiction. It follows that there is no integer $k > 0$ such that there exists a set of $n+k$ linearly independent points. Hence E has dimension n.

REMARK. If E has dimension n, then an n-tuple of linearly independent points $(e_1, e_2, \ldots, e_n)$ is called a **basis** for E. Each point of E can then be expressed in the form $x = \lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_n e_n$, where the coefficients λ_i are unique, since

$$x = \sum \lambda_i e_i = \sum \lambda_i' e_i \Rightarrow \sum (\lambda_i - \lambda_i') e_i = 0 \Rightarrow (\forall i) : \lambda_i - \lambda_i' = 0.$$

Theorem 4. *If a subset A of a vector space X is contained in a linear variety of dimension n, then each point x of $[A]$ can be expressed in the form*

$$\begin{gathered} x = p_1 a_1 + p_2 a_2 + \ldots + \ldots + p_{n+1} a_{n+1}; \\ (p_1, p_2, \ldots, p_{n+1}) \in \mathbf{P}_{n+1}, \\ a_1, a_2, \ldots, a_{n+1} \in A. \end{gathered}$$

Proof. We must show that each point of the form

$$x = p_1 a_1 + p_2 a_2 + \ldots + p_k a_k$$

where $(p_1, p_2, \ldots, p_k) \in \mathbf{P}_k$ and k is arbitrary, can also be written

$$x = p_1' a_1' + p_2' a_2' + \ldots + p_{n+1}' a_{n+1}',$$

where $(p_1', p_2', \ldots, p_{n+1}') \in \mathbf{P}_{n+1}$.

This is trivial if $k \leqq n+1$ and so we suppose that $k > n+1$ and that all the p_i are non-zero. The points $a_1 - a_k, a_2 - a_k, \ldots, a_{k-1} - a_k, 0$ are contained in a subspace E of dimension $n < k-1$ and so there exist numbers $\lambda_1, \lambda_2, \ldots, \lambda_{k-1}$, not all zero, such that

$$\lambda_1(a_1 - a_k) + \lambda_2(a_2 - a_k) + \ldots + \lambda_{k-1}(a_{k-1} - a_k) = 0.$$

Writing $\lambda_k = -\lambda_1 - \lambda_2 \ldots - \lambda_{k-1}$, we have

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k = 0; \sum_{i=1}^{k} \lambda_i = 0.$$

The numbers λ_i / p_i are not all zero; let μ be the greatest. Then $\mu > 0$ and

$$x = x-0 = \sum_{i=1}^{k}\left(p_i-\frac{\lambda_i}{\mu}\right)a_i;\ \sum_{i=1}^{k}\left(p_i-\frac{\lambda_i}{\mu}\right) = \sum_{i=1}^{k}p_i-\frac{1}{\mu}\sum_{i=1}^{k}\lambda_i = 1,$$

$$(\forall i) : p_i-\frac{\lambda_i}{\mu} \geqq 0,\quad (\exists i) : \left(p_i-\frac{\lambda_i}{\mu}\right) = 0.$$

Thus we have reduced the number of the a_i which appear in the expression for x; repeating the operation as often as is necessary, we can express x in terms of at most $n+1$ of the points a_i.

Theorem 5. *If σ is a single-valued linear mapping of X into Y and C is a convex subset of X having dimension n, then the convex set σC has dimension less than or equal to n.*

Proof. Let $E+x_0$ be the linear variety of dimension n containing C; its image

$$\sigma(E+x_0) = \sigma E+\sigma x_0$$

is a linear variety containing σC. To prove the theorem, it is sufficient to show that the subspace σE has dimension less than or equal to n. If $(e_1, e_2, \ldots, e_n)$ is a basis for E, then each point of σE can be written

$$y = \sigma(\lambda_1 e_1+\lambda_2 e_2+ \ldots +\lambda_n e_n) = \lambda_1\sigma e_1+\lambda_2\sigma e_2+ \ldots +\lambda_n\sigma e_n.$$

Let $(\sigma e_{i_1}, \sigma e_{i_2}, \ldots, \sigma e_{i_k})$ be a k-tuple of linearly independent points extracted from the σe_i, with k as large as possible. This is a basis for σE and so σE has dimension k; therefore $k \leqq n$.

§ 5. The gauge of a convex set

In this section we study 'local' properties of a convex set: that is, properties which relate to a particular point x_0; we suppose, for convenience, that $x_0 = 0$.

Let C be a convex set. We say that a half-line D^+ issuing from 0 is **privileged** (with respect to 0) if to each point x of D^+ there corresponds a number $\eta > 0$ such that

$$\eta x \in C.$$

Similarly we say that a line D passing through 0 is privileged if to each point x in D there corresponds a number $\eta > 0$ such that

$$\eta x \in C, \qquad -\eta x \in C.$$

A convex set C is said to be **symmetric** (with respect to 0) if

$$x \in C \text{ implies that } -x \in C;$$

in other words, $C = -C$.

A convex set C is said to be **semi-bounded** (with respect to 0) if every half-line issuing from 0 contains at least one point not belonging to C. In $\mathbf{R}^n$, a bounded set is semi-bounded, but the converse is not true: in $\mathbf{R}^2$ the convex set

$$C = \{(x, y) \,/\, x \geqq 0,\ y > 0,\ y \leqq 1\} \cup [(0, 0), (1, 0)[$$

is semi-bounded but is not bounded (see figure 35).

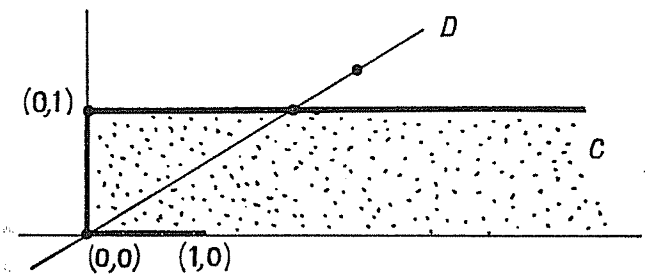


FIG. 35

Let 0 be a point of a convex set C. If there are no privileged lines with respect to 0, then 0 is called an **extreme** point of C. If, on the other hand, every line through 0 is privileged with respect to 0, then 0 is called an **internal** point of C. The set of extreme points of C is denoted by $\ddot{C}$; this is called the **profile** of C. We say that a point x of X is **frontal** (with respect to 0) if

$$\begin{aligned}&(1)\quad]0, x[\ \subset C,\\ &(2)\quad D_x^+ - [0, x] \subset -C,\end{aligned}$$

where D_x^+ is the half-line issuing from 0.

EXAMPLE. In a plane convex polygon, the extreme points are the vertices of the polygon; the idea of extreme point is a natural generalisation of that of vertex.

In the above figure, the extreme points are (0, 0) and (0, 1); the internal points are the points (x, y) such that

$$x > 0;\ y < 1;\ y > 0$$

The frontal points are the points (x, y) such that $x \geqq 0, y = 1$, and the point (1, 0).

Theorem 1. *If a convex set C admits a privileged line with respect to a point* 0, *then the union of all the privileged lines with respect to* 0 *is a vector subspace.*

Proof. Let A be the union of the privileged lines with respect to 0. We first show that if $a_1, a_2 \in A$, then $a_1+a_2 \in A$. If $a_1 = 0$ or $a_2 = 0$, this is trivial. Suppose therefore that $a_1 \neq 0$ and $a_2 \neq 0$. There exists a positive number η such that

$$\begin{aligned} +\eta a_1 &= c_1 \in C, & +\eta a_2 &= c_2 \in C, \\ -\eta a_1 &= -c_1 \in C, & -\eta a_2 &= -c_2 \in C. \end{aligned}$$

Then

$$\begin{cases} \dfrac{\eta}{2}(a_1+a_2) = \dfrac{c_1+c_2}{2} \in C, \\[2ex] -\dfrac{\eta}{2}(a_1+a_2) = \dfrac{(-c_1)+(-c_2)}{2} \in C \end{cases}$$

and so a_1+a_2 belongs to a privileged line.

We now show that if $a \in A$ then $\lambda a \in A$. If $\lambda = 0$, this is trivial. Suppose therefore that $\lambda \neq 0$. Then

$$\frac{\eta}{\lambda}(\lambda a) = \eta a \in C, \quad -\frac{\eta}{\lambda}(\lambda a) = -\eta a \in C$$

and therefore λa belongs to a privileged line.

Theorem 2. *A necessary and sufficient condition for* 0 *to be an extreme point of C is that $C-\{0\}$ is a convex set.*

Proof. If $C-\{0\}$ is not convex, there exist points x_1 and x_2 such that

$$x_1, x_2 \in C-\{0\}; \; p_1x_1+p_2x_2 \notin C-\{0\}; \; (p_1, p_2) \in \mathbf{P}_2.$$

Since $p_1x_1+p_2x_2 \in C$, we have $p_1x_1+p_2x_2 = 0$. This implies that the straight line joining x_1 and x_2 passes through 0 and is privileged; therefore 0 is not an extreme point.

Conversely, suppose that 0 is not an extreme point; we shall show that $C-\{0\}$ is not convex. Since 0 is not an extreme point, there exists a privileged line; if x is a point of this line there exists a positive number η such that

$$\eta x \in C-\{0\}; \quad -\eta x \in C-\{0\}.$$

Then $C-\{0\}$ cannot be convex, since

$$\frac{1}{2}\eta x+\frac{1}{2}(-\eta x) = 0 \notin C-\{0\}.$$

If C is a convex set containing 0, then the generalised numerical function j defined by

$$j(x) \begin{cases} = \inf\{t \mid t > 0,\ tC \ni x\} & \text{if } tC \ni x \text{ for at least one } t > 0, \\ = +\infty & \text{otherwise} \end{cases}$$

is called the **gauge** of C (with respect to 0). If $x \in C$, then $j(x) \leqq 1$ and if $x \notin C$, then $j(x) \geqq 1$.

Theorem 3. *The gauge j of a convex set C satisfies the following conditions*:

$$\begin{aligned} &(1)\quad j(x) \geqq 0 \textit{ for all } x \neq 0;\ j(0) = 0, \\ &(2)\quad j(\lambda x) = \lambda j(x) \textit{ if } \lambda > 0 \\ &(3)\quad j(x+y) \leqq j(x)+j(y). \end{aligned}$$

Proof. Condition (1) is immediate. If $\lambda > 0$, then

$$\begin{cases} j(x) = +\infty \Rightarrow j(\lambda x) = +\infty = \lambda j(x), \\ j(x) < +\infty \Rightarrow j(\lambda x) = \inf\{t \mid tC \ni \lambda x\} = \\ \qquad\qquad\qquad = \inf\{\lambda s \mid \lambda s\, C \ni \lambda x\} = \lambda j(x) \end{cases}$$

and so (2) is satisfied. Condition (3) is immediate if $j(x) = +\infty$ or $j(y) = +\infty$; if $j(x) < +\infty$ and $j(y) < +\infty$, then, for all $\varepsilon > 0$,

$$\begin{cases} s = j(x)+\varepsilon \\ t = j(y)+\varepsilon \end{cases} \Rightarrow \begin{cases} j\left(\dfrac{x}{s}\right) < 1 \\ j\left(\dfrac{y}{t}\right) < 1 \end{cases} \Rightarrow \begin{cases} \dfrac{x}{s} \in C \\ \dfrac{y}{t} \in C \end{cases} \Rightarrow$$

$$\Rightarrow \frac{1}{s+t}(x+y) = \frac{s}{s+t}\left(\frac{x}{s}\right)+\frac{t}{s+t}\left(\frac{y}{t}\right) \in C.$$

Hence we have

$$(s+t)C = (j(x)+j(y)+2\varepsilon)C \ni (x+y)$$

and therefore

$$j(x+y) \leqq j(x)+j(y)+2\varepsilon.$$

Since ε is arbitrary, condition (3) follows.

REMARK. Condition (2) cannot be replaced by

$$(2')\quad j(\lambda x) = \lambda j(x) \text{ if } \lambda \geqq 0,$$

because it is necessary to take into account the case $j(x) = +\infty$, $\lambda = 0$.

Theorem 4. *Given a function j satisfying properties* (1), (2) *and* (3) *of Theorem* 3, *there exists a convex set*

$$C_j = \{x \,/\, x \in X,\ j(x) \leqq 1\}$$

which has j as gauge (this set is called the **representative** of j).

Proof. The set C_j is convex, for if $(p_1, p_2) \in \mathbf{P}_2$, then

$$x_1, x_2 \in C_j \Rightarrow j(p_1x_1+p_2x_2) \leqq p_1 j(x_1)+p_2 j(x_2) \leqq p_1+p_2 = 1$$
$$\Rightarrow p_1x_1+p_2x_2 \in C_j.$$

Furthermore, $0 \in C_j$, since $j(0) = 0$.

Let j' be the gauge of C_j and let $a \in X$ be such that $a \neq 0$. We shall prove that $j'(a) = j(a)$.

Case (1). If $j(a) = +\infty$, the only point of C_j on the half-line D_a^+ is 0 and so $j'(a) = +\infty = j(a)$.

Case (2). If $j(a) = 0$, then $D_a^+ \subset C_j$ and so $j'(a) = 0 = j(a)$.

Case (3). If $j(a) > 0$ and $j(a) < +\infty$, the intersection of C_j and D_a^+ is a linear segment $[0, x_0]$; we have

$$j(x_0) = 1;\ j'(x_0) = 1$$

and therefore

$$j(a) = j(\lambda x_0) = \lambda j(x_0) = \lambda = \lambda j'(x_0) = j'(\lambda x_0) = j'(a).$$

REMARK. We observe that if a convex set C, containing 0, has gauge j, then C_j is the union of C and its frontal points. Thus a set which contains all its frontal points is the representative of its gauge.

Theorem 5. *If $C_1, C_2, \ldots, C_n$ are convex sets having gauges $j_1, j_2, \ldots, j_n$ respectively, the convex set $C = \bigcap C_i$ has gauge j given by*

$$j(x) = \max_i j_i(x).$$

Proof. Case (1). If $tC \not\ni x (\forall t > 0)$, there exists a C_i such that $tC_i \not\ni x$ $(\forall t > 0)$ and then

$$j(x) = \max_i j_i(x) = +\infty.$$

Case (2). If $tC \ni x$ for at least one $t > 0$, then

$$j(x) = \max_i j_i(x) = \inf \{t \,/\, t > 0,\ tC_i \ni x\ (\forall i \leqq n)\} = \inf \{t \,/\, t > 0,\ tC \ni x\}.$$

Therefore j is the gauge of $C = \bigcap_{i=1}^{n} C_i$.

Theorem 6. *If C has gauge j, the convex set $C' = -C$ has gauge j', where* $j'(x) = j(-x)$.

Proof. *Case* (1). If $t(-C) \not\ni x$ ($\forall t > 0$), then $tC \not\ni -x$ ($\forall t > 0$) and so

$$j'(x) = j(-x) = +\infty.$$

Case (2). If $t(-C) \ni x$ for some $t > 0$, then

$$j'(x) = j(-x) = \inf \{t \mid tC \ni -x\} = \inf \{t \mid t(-C) \ni x\}.$$

Therefore j' is the gauge of $C' = -C$.

Theorem 7. *A convex set C containing* 0 *is symmetric with respect to* 0 (*except for frontal points*) *if and only if its gauge j satisfies $j(x) = j(-x)$; C has* 0 *as an internal point if and only if its gauge j satisfies $j(x) < +\infty$ for all x; C is semi-bounded if and only if its gauge j satisfies $j(x) > 0$ for all $x \neq 0$.*

This follows immediately from the preceding theorems.

In a vector space X, a gauge which satisfies $j(x) = j(-x)$ is called a **semi-norm.** If, further, $j(x) > 0$ for all $x \neq 0$, j is called a **norm**; if also $j(x) < +\infty$ for all x, then j is called a **proper norm.**

A proper norm j satisfies the following characteristic properties

$$\begin{aligned}
&(1)\quad j(x) \geqq 0,\\
&(2)\quad j(x) = 0 \iff x = 0,\\
&(3)\quad j(\lambda x) = |\lambda|\, j(x) \text{ (for all } \lambda),\\
&(4)\quad j(x_1+x_2) \leqq j(x_1)+j(x_2).
\end{aligned}$$

EXAMPLE 1. In the space $\mathbf{R}^n$, put

$$d(x, 0) = \sqrt{(x^1)^2+(x^2)^2+\ldots+(x^n)^2} = \| x \|.$$

Then the function j defined by $j(x) = \| x \|$ is a norm, for

$$\begin{aligned}
&\| x \| \geqq 0,\\
&\| x \| = 0 \iff x = 0,\\
&\| \lambda x \| = |\lambda| \cdot \| x \|,\\
&\| x_1+x_2 \| \leqq \| x_1 \| + \| x_2 \|.
\end{aligned}$$

EXAMPLE 2. Let f be a numerical linear function defined on a vector space X. The function j defined by $j(x) = |f(x)|$ is a semi-norm, because

$$\begin{aligned}
&(1)\quad |f(x)| \geqq 0; \quad |f(0)| = 0,\\
&(2)\quad |f(\lambda x)| = \lambda\, |f(x)| \text{ if } \lambda > 0,\\
&(3)\quad |f(x+y)| = |f(x)+f(y)| \leqq |f(x)| + |f(y)|,\\
&(4)\quad |f(x)| = |f(-x)|.
\end{aligned}$$

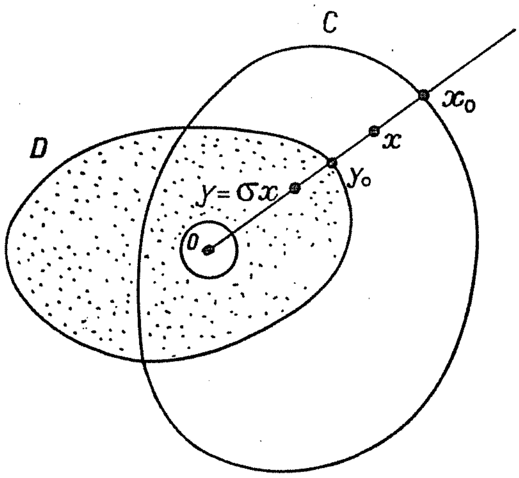


FIG. 36

Let C and D be two convex sets in a vector space X, having j and k respectively as gauges. Suppose that C and D each have 0 as an internal point, that they are semi-bounded and that they each contain their frontal points. We can define a single-valued mapping of X into X by writing

$$y = \sigma x = \left(\frac{j(x)}{k(x)}\right) x.$$

By means of this mapping we can set up a correspondence in which to each point x of C there corresponds a point y of D such that

$$k(y) = k\left[\frac{j(x)}{k(x)} x\right] = \frac{j(x)}{k(x)} k(x) = j(x).$$

In other words, the *mapping* σ *establishes a one-one correspondence between the points of C and those of D and between the frontal points of C and those of D.* This mapping is called the **radial projection** of C on D. We shall study it from a topological point of view in the following chapter.

§ 6. The Hahn-Banach theorem

We say that a plane $\{x \,/\, f(x) = 1\}$ **separates** two sets A and B if

$$x \in A \;\Rightarrow\; f(x) \leqq 1,$$
$$x \in B \;\Rightarrow\; f(x) \geqq 1.$$

In many problems in analysis we encounter two disjoint convex sets to be separated by a plane; in this section we prove the fundamental theorem concerning separation.

LEMMA 1 (Kakutani). *If A and B are two disjoint convex sets, there exist two convex sets C_0 and D_0 such that $C_0 \supset A$, $D_0 \supset B$, $C_0 \cap D_0 = \emptyset$, $C_0 \cup D_0 = X$.*

Proof. Let $\mathscr{C}$ be the family of pairs (C, D) consisting of two disjoint convex sets C and D such that $C \supset A$, $D \supset B$; $\mathscr{C}$ is not empty, since $(A, B) \in \mathscr{C}$. Write $(C, D) \leqq (C', D')$ if $C \subset C'$ and $D \subset D'$; then $\mathscr{C}$ is ordered by the ordering relation $\leqq$. Since every totally ordered sub-family has an upper bound in $\mathscr{C}$, then, by Zorn's theorem (Chapter III, § 6) there exists a maximal element (C_0, D_0) in $\mathscr{C}$:

$$\left.\begin{array}{r} C \supset C_0,\ D \supset D_0 \\ (C, D) \in \mathscr{C} \end{array}\right\} \Rightarrow C = C_0,\ D = D_0.$$

If $x_0 \in X$, then the following two conditions cannot be satisfied simultaneously:

$$\begin{array}{l} (\exists_{C_0} c)\,(\exists_{D_0} d) : c \in [x_0, d] \\ (\exists_{C_0} c_0)\,(\exists_{D_0} d_0) : d_0 \in [x_0, c_0]. \end{array}$$

(Otherwise $[c, c_0]$ and $[d, d_0]$ intersect and so C_0 and D_0 intersect). Suppose, for example, that the first condition is not satisfied, so that

$$(\forall_{C_0} c)\,(\forall_{D_0} d) : c \notin [x_0, d].$$

If we write $D = [D_0 \cup \{x_0\}]$, we have $C_0 \cap D = \emptyset$ and so $(C_0, D) \in \mathscr{C}$. Since $D \supset D_0$, it follows that $D = D_0$ and therefore $x_0 \in D_0$. A similar argument shows that, if the second condition is not satisfied, then $x_0 \in C_0$. Therefore either $x_0 \in D_0$ or $x_0 \in C_0$, and hence $C_0 \cup D_0 = X$.

LEMMA 2 (Ghouila-Houri). *Let C and D be non-empty disjoint convex sets in a vector space X such that $C \cup D = X$. If C has an internal point, there exists a linear non-constant numerical function f such that*

$$\begin{array}{lcl} x \in C & \Rightarrow & f(x) \leqq 1, \\ x \in D & \Rightarrow & f(x) \geqq 1. \end{array}$$

Proof. By making a translation (if necessary) we can choose the origin 0 to be an internal point of C. Then, given a point $x \in X$, there exists a real number $\lambda > 0$ such that $x \in \lambda C$.

If there exists a $\lambda > 0$ such that $x \notin \lambda C$, we write

$$f(x) = \inf \{\lambda \,/\, \lambda > 0;\ x \in \lambda C\} > 0.$$

If there exists a $\lambda < 0$ such that $x \notin \lambda C$, we write

$$f(x) = -f(-x) < 0.$$

If $x \in \lambda C$ for all $\lambda \neq 0$, we write

$$f(x) = 0.$$

The function f defined in this way is not identically zero and satisfies

$$\begin{aligned} x \in C &\Rightarrow f(x) \leqq 1, \\ x \in D &\Rightarrow f(x) \geqq 1. \end{aligned}$$

It is clear that if $\alpha \in \mathbf{R}$ and $x \in X$, then $f(\alpha x) = \alpha f(x)$. It remains to be shown that for any two given points x and y we have

$$f(x+y) = f(x)+f(y).$$

Case 1. $f(x) = 0 = f(y)$. In this case $\frac{1}{2}(x+y) \in \lambda C$ for all $\lambda > 0$ and so

$$f(x+y) = 0 = f(x)+f(y).$$

Case 2. $f(x) = 0$ and $f(y) > 0$. If $\lambda > 0$, $\frac{1}{\lambda} y \in C$ and $0 < \varepsilon < \frac{1}{\lambda}$, then

$$\left(\frac{1}{\lambda}-\varepsilon\right)(x+y) = \varepsilon\lambda\left(\frac{1-\varepsilon\lambda}{\varepsilon\lambda^2}\right)x+(1-\varepsilon\lambda)\frac{1}{\lambda}\,y \in C.$$

If $\lambda > 0$, $\frac{1}{\lambda}(x+y) \in C$ and $0 < \varepsilon < \frac{1}{\lambda}$, then

$$\left(\frac{1}{\lambda}-\varepsilon\right)y = \varepsilon\lambda\left(\frac{\varepsilon\lambda-1}{\varepsilon\lambda^2}\right)x+(1-\varepsilon\lambda)\frac{1}{\lambda}(x+y) \in C.$$

Hence

$$f(x+y) = f(x)+f(y).$$

Case 3. $f(x) = 0$ and $f(y) < 0$. We have

$$f(x+y) = -f(-x-y) = -f(-x)-f(-y) = f(x)+f(y).$$

Case 4. $f(x) > 0$ and $f(y) > 0$. If $\lambda > 0$, $\mu > 0$, $\frac{1}{\lambda} x \in C$ and $\frac{1}{\mu} y \in C$, then

$$\frac{1}{\lambda+\mu}(x+y) = \left(\frac{\lambda}{\lambda+\mu}\right)\frac{1}{\lambda}x+\left(\frac{\mu}{\lambda+\mu}\right)\frac{1}{\mu}\,y \in C.$$

Similarly, if $\lambda > 0$, $\mu > 0$, $\frac{1}{\lambda} x \in D$ and $\frac{1}{\mu} y \in D$, then $\frac{1}{\lambda+\mu}(x+y) \in D$.

Therefore

$$f(x+y) = \inf\left\{\lambda+\mu \,/\, \frac{x+y}{\lambda+\mu} \in C\right\} = f(x)+f(y).$$

Case 5. $f(x) < 0$ and $f(y) < 0$. We have

$$f(x+y) = -f(-x-y) = -f(-x)-f(-y) = f(x)+f(y).$$

Case 6. $f(x) > 0$ and $f(y) < 0$. If $f(x+y) \geqq 0$, then

$$f(x) = f(x+y)+f(-y) = f(x+y)-f(y).$$

If $f(x+y) < 0$, then

$$f(y) = f(x+y)+f(-x) = f(x+y)-f(x).$$

This completes the proof of Lemma 2.

Hahn-Banach theorem. *Let A and B be two disjoint non-empty convex sets in a vector space X. If A has an internal point, there exists a plane separating A and B.*

Proof. By Lemma 1, there exist two disjoint convex sets C and D such that $A \subset C$, $B \subset D$ and $C \cup D = X$. The theorem then follows immediately from Lemma 2.

This proof(1), unlike the usual proofs, makes no appeal to topological concepts.

(1) This proof is due to Ghouila-Houri and has been added to the present (English) edition with the kind of permission of its author.

CHAPTER VIII

CONVEX SETS AND CONVEX FUNCTIONS IN THE SPACE $\mathbf{R}^n$

§ 1. Topological properties of convex sets

In the preceding chapter we made no use of topological ideas in studying the vector space X. In the present chapter, we consider the space $X = \mathbf{R}^n$ and regard it both as a vector space and a topological space[(1)]; we obtain a number of interesting theorems by using the metric topology of $\mathbf{R}^n$.

Let $x^1, x^2, \ldots, x^n$ be real numbers; we shall denote the corresponding point of $\mathbf{R}^n$ by x, so that $x = (x^1, x^2, \ldots, x^n)$. The symbol δ_j^i (known as the Kronecker symbol) represents the number 0 if $i \neq j$ and the number 1 if $i = j$. The points

$$\delta_i = (\delta_i^1, \delta_i^2, \ldots, \delta_i^n) = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$$

form a basis for $\mathbf{R}^n$, because for any point x we have

$$x = x^1 \delta_1 + x^2 \delta_2 + \ldots + x^n \delta_n.$$

Let a and x be points of $\mathbf{R}^n$. We write

$$\langle a, x\rangle = a^1x^1 + a^2x^2 + \ldots + a^nx^n.$$

The number $\langle a, x\rangle$ is called the **scalar product** of the points a and x; we have

$$\begin{aligned} &(1) \quad \langle a, x\rangle = \langle x, a\rangle, \\ &(2) \quad \langle a, \lambda x\rangle = \lambda\langle a, x\rangle, \\ &(3) \quad \langle a, x_1+x_2\rangle = \langle a, x_1\rangle + \langle a, x_2\rangle. \end{aligned}$$

Thus, for fixed a, the correspondence $x \to \langle a, x\rangle$ determines a numerical linear function in $\mathbf{R}^n$. We note that

$$x^k = \langle x, \delta_k\rangle.$$

Two vectors x and y are said to be **orthogonal** if $\langle x, y\rangle = 0$.

We write

$$\| x \| = \sqrt{\langle x, x\rangle} = \sqrt{(x^1)^2 + (x^2)^2 + \ldots + (x^n)^2}$$

(see Example 1, § 5, Chapter VII); we call $\| x \|$ the **Euclidean norm** of x and

[(1)] We obtain certain results often used in the calculus of variations, linear programming, the theory of games, etc. Some of the results of the present chapter are extended to more general spaces in Chapter IX.

in what follows we shall denote the distance between x and y by $\| x-y \|$ in preference to $d(x, y)$.

REMARK 1. We have

$$|\langle x, y\rangle| \leqq \| x \| \cdot \| y \| \quad \text{(the Cauchy-Schwartz inequality).}$$

Proof. The expression

$$\| \lambda x+y \|^2 = \langle \lambda x+y, \lambda x+y\rangle = \lambda^2 \| x \|^2+2\lambda \langle x, y\rangle+\| y \|^2$$

is a quadratic form in λ with no negative values and so its discriminant is not strictly positive: that is,

$$\Delta' = (\langle x, y\rangle)^2-\| x \|^2 \cdot \| y \|^2 \leqq 0,$$

which proves the inequality.

REMARK 2. We have

$$\| x+y \| \leqq \| x \|+\| y \| \quad \text{(the triangular inequality).}$$

Proof. Using the Cauchy-Schwartz inequality, we have

$$\| x+y \|^2 = \langle x+y, x+y\rangle = \| x \|^2+\| y \|^2+2\langle x, y\rangle$$
$$\ldots \leqq \|x\|^2+\|y\|^2+2\,|\langle x, y\rangle| \leqq \|x\|^2+\|y\|^2+2\,\|x\| \cdot \|y\| = (\|x\|+\|y\|)^2$$

and the result follows.

REMARK 3. The numerical function j defined by $j(x) = \| x \|$ is a norm in the sense defined above (see p. 153); in fact

(1) $\| x \| \geqq 0$, $\| x \| \in \mathbf{R}$,

(2) $\| x \| = 0$ is equivalent to $x = 0$.

(3) $\| \lambda x \| = |\lambda| \cdot \| x \|$,

(4) $\| x+y \| \leqq \| x \|+\| y \|$.

The convex set representing this norm is the ball

$$B_1(0) = \{x \,/\, \| x \| \leqq 1\}$$

(which is thus a symmetric convex set admitting 0 as an internal point).

Theorem 1. *If f is a numerical linear function in $\mathbf{R}^n$, there exists a fixed point $a \in \mathbf{R}^n$ such that $f(x) = \langle a, x\rangle$; furthermore, f is continuous.*

Proof. We have

$$f(x) = f(\Sigma x^i\delta_i) = \Sigma x^i f(\delta_i).$$

Then, writing $a = (f(\delta_1), f(\delta_2), \ldots, f(\delta_n))$, we have $f(x) = \langle a, x\rangle$. Moreover, f is continuous in $\mathbf{R}^n$, for $\| x-x_0 \| \leqq \dfrac{\varepsilon}{\Sigma |a^i|}$ implies that

$$|f(x)-f(x_0)| = |\Sigma a^i(x^i-x_0^i)| \leqq \Sigma |a^i| \cdot |x^i-x_0^i| \leqq \frac{\varepsilon}{\Sigma |a^i|} \Sigma |a^i| = \varepsilon.$$

COROLLARY 1. *The plane* $E_f^\alpha = \{x \,/\, f(x) = \alpha\}$ *and the closed half-space*

$$H_f^\alpha = \{x \,/\, f(x) \geqq \alpha\}$$

are closed sets; the open half-space

$$H_f'^\alpha = \{x \,/\, f(x) > \alpha\}$$

is an open set.

Proof. Since the numerical linear function f is both upper and lower semi-continuous, the sets $\{x \,/\, f(x) \geqq \alpha\}$ and $\{x \,/\, f(x) \leqq \alpha\}$ are closed and the sets $\{x \,/\, f(x) > \alpha\}$ and $\{x \,/\, f(x) < \alpha\}$ are open.

COROLLARY 2. *If a set* A *is situated entirely on one side of the plane* $f(x) = \alpha$ *(that is, if* $\inf\limits_{x \in A} f(x) \geqq \alpha$ *or* $\sup\limits_{x \in A} f(x) \leqq \alpha$*) then so is its closure* $\bar{A}$.

Proof. If $\inf\limits_{x \in A} f(x) \geqq \alpha$, we have $H_f^\alpha \supset A$; then, since H_f^α is closed,

$$H_f^\alpha = \bar{H}_f^\alpha \supset \bar{A}.$$

A similar argument can be used if $\sup\limits_{x \in A} f(x) \leqq \alpha$.

Theorem 2. *Let* $X = \mathbf{R}^n$*; then the single-valued mapping* σ *defined by*

$$\sigma(x, y) = x + y$$

is a continuous mapping of $X \times X$ *into* X*; the single-valued mapping* τ *defined by*

$$\tau(\lambda, x) = \lambda x$$

is a continuous mapping of $\mathbf{R} \times X$ *into* X.

Proof. The mapping σ is continuous at (x_0, y_0), for

$$d[(x, y), (x_0, y_0)] = \max\{\|x - x_0\|, \|y - y_0\|\} \leqq \varepsilon$$

implies that

$$\begin{aligned} \|(x+y) - (x_0+y_0)\| &= \|(x-x_0)+(y-y_0)\| \\ &\leqq \|x-x_0\| + \|y-y_0\| \leqq \varepsilon+\varepsilon = 2\varepsilon. \end{aligned}$$

Since (x_0, y_0) is arbitrary, σ is continuous in $X \times X$.

The mapping τ is continuous at (λ_0, x_0), for

$$d[(\lambda, x), (\lambda_0, x_0)] = \max\{|\lambda - \lambda_0|, \|x - x_0\|\} \leqq \varepsilon$$

implies that

$$\|\lambda x - \lambda_0 x_0\| = \|\lambda x - \lambda x_0 + \lambda x_0 - \lambda_0 x_0\|$$

$$\leqq \|\lambda(x-x_0)\| + \|(\lambda-\lambda_0)x_0\| \leqq |\lambda|\varepsilon + \varepsilon\|x_0\| = (\|x_0\| + |\lambda|)\,\varepsilon.$$

For ε sufficiently small, we have

$$\|\lambda x - \lambda_0 x_0\| \leqq (\|x_0\| + |\lambda_0| + 1)\,\varepsilon.$$

Since (λ_0, x_0) is arbitrary, τ is continuous in $\mathbf{R} \times X$.

COROLLARY 1. *If A and G are two sets in X and if G is an open set, then the set $A+G$ is also open.*

Proof. The mapping σ_a defined by $\sigma_a x = x+a$ is single-valued and continuous and so is its inverse σ_a^{-1}, where $\sigma_a^{-1}(x) = x-a$. Hence σ_a is a homeomorphism and therefore $G+a = \sigma_a G$ is an open set. Therefore the set

$$G+A = \bigcup_{a\in A}(G+a)$$

is also open.

COROLLARY 2. *If K and K' are two compact sets in X, the set $K+K'$ is compact.*

Proof. By Tychonoff's theorem, $K\times K'$ is a compact set in $X\times X$; since $\sigma(x, y) = x+y$ defines a continuous single-valued mapping of $X\times X$ into the separated space X, the image $\sigma(K\times K') = K+K'$ is a compact set.

COROLLARY 3. *If F is a closed set in X and K is a compact set in X, then the set $K+F$ is a closed set in X.*

Proof. Let $(x_n) = (y_n+z_n)$ be a convergent sequence in $K+F$, where $y_n \in K$ and $z_n \in F$. The sequence (y_n) contains a convergent sub-sequence (y_{k_n}). If $(y_{k_n}) \to y_0$ and $(x_{k_n}) \to x_0$, we have

$$(z_{k_n}) = (x_{k_n}-y_{k_n}) \to x_0-y_0 = z_0.$$

Since $z_{k_n} \in F$ and F is closed, we have $z_0 \in F$ and so

$$x_0 = y_0+z_0 \in K+F.$$

But if, in a metric space, every convergent sequence of elements of a set F' converges to a point of F', then we know that this set is closed (Theorem 2, § 3, Chapter V). Therefore $K+F$ is closed.

COROLLARY 4. *If the sets G, F, K are respectively open, closed and compact, then the sets λG, λF, λK are respectively open, closed and compact (except when $\lambda = 0$ in the first case).*

Proof. If $\lambda \neq 0$, the mapping defined by $\sigma_\lambda x = \lambda x$ is single-valued and continuous and the same is true of its inverse, given by $\sigma_\lambda^{-1} x = \frac{1}{\lambda} x$; therefore σ_λ is a homeomorphism and so $\lambda G = \sigma_\lambda G$ is open, $\lambda F = \sigma_\lambda F$ is closed and $\lambda K = \sigma_\lambda K$ is compact. In the cases of closed sets and compact sets, the theorem is true trivially if $\lambda = 0$.

Let E_f^α be a plane and let A and B be two non-empty sets; we say that E_f^α **separates** A and B (see § 6, Chapter VII) if

$$\sup_{x\in A} f(x) \leqq \alpha \leqq \inf_{y\in B} f(y).$$

We also say that E_f^α separates A and B strictly if

$$\sup_{x \in A} f(x) < \alpha < \inf_{y \in B} f(y).$$

LEMMA 1. *If C is a non-empty closed convex set not containing the origin, there exists a linear function f and a positive number α such that*

$$x \in C \Rightarrow f(x) > \alpha.$$

Proof. Let $B_\lambda(0)$ be a closed ball with centre the origin and meeting C; the set $C \cap B_\lambda(0)$ is compact. The continuous function defined by $g(x) = \| x \|$ therefore attains its infimum for this set at a point x_0 of the set. Since $x_0 \in C$ we have $\| x_0 \| > 0$; furthermore, we have

$$x \in C \cap B_\lambda(0) \Rightarrow \| x \| \geqq \| x_0 \|,$$

whence

$$x \in C \Rightarrow \| x \| \geqq \| x_0 \|.$$

We shall prove that, at each point x of C,

$$\langle x_0, x \rangle \geqq \| x_0 \|^2.$$

This is sufficient to prove the lemma, for the function defined by $f(x) = \langle x_0, x \rangle$ is linear.

Suppose that $y \in C$ and that $\langle x_0, y \rangle = \| x_0 \|^2 - \varepsilon$, where $\varepsilon > 0$. Let t be a number between 0 and 1. Then

$$\begin{aligned} \| (1-t)x_0 + ty \|^2 &= (1-t)^2 \| x_0 \|^2 + 2t(1-t) \langle x_0, y \rangle + t^2 \| y \|^2 \\ &= \| x_0 \|^2 - 2t \| x_0 \|^2 + t^2 \| x_0 \|^2 + 2t(1-t) [\| x_0 \|^2 - \varepsilon] + t^2 \| y \|^2, \\ &= \| x_0 \|^2 - 2t(1-t)\varepsilon + t^2 [\| y \|^2 - \| x_0 \|^2]. \end{aligned}$$

We can choose t to be such that

$$\frac{t}{1-t} [\| y \|^2 - \| x_0 \|^2] < 2\varepsilon;$$

this implies that

$$\| (1-t)x_0 + ty \|^2 - \| x_0 \|^2 < -2t(1-t)\varepsilon + t \,.\, 2\varepsilon(1-t) = 0.$$

But, since $(1-t)x_0 + ty \in C$, we also have

$$\| (1-t)x_0 + ty \| \geqq \| x_0 \|$$

which gives a contradiction. Therefore it is false that $\langle x_0, y \rangle = \| x_0 \|^2 - \varepsilon$ and so we have $\langle x_0, x \rangle \geqq \| x_0 \|^2$ as required.

Thus the lemma is established.

LEMMA 2. *If C is a non-empty convex set not containing the origin, there exists a linear form f, not identically zero, such that*

$$x \in C \Rightarrow f(x) \geqq 0.$$

Proof. With each point $x \in C$ we associate the set

$$A_x = \{y \,/\, \| y \| = 1 \text{ and } \langle y, x \rangle \geqq 0\}.$$

Let $x_1, x_2, \ldots, x_k$ be any finite family of points in C. The set of points x of the form

$$x = \sum p_i x_i, \text{ where } \sum p_i = 1 \text{ and } p_i \geqq 0$$

is a closed convex set not containing the origin; therefore, by Lemma 1, there exists a vector y such that $\langle y, x_i \rangle > 0$ for all i. We can choose y to be such that $\| y \| = 1$; then we have

$$\bigcap_{i=1}^{k} A_{x_i} \neq \varnothing.$$

By definition the sets A_x are contained in the compact set $\{y \,/\, \| y \| = 1\}$ and therefore, by the finite intersection property, we have

$$\bigcap_{x \in C} A_x \neq \varnothing.$$

Choose a to be a point of $\bigcap_{x \in C} A_x$; then the function f defined by $f(x) = \langle a, x \rangle$ satisfies the desired conditions.

First separation theorem. *If C and C' are two non-empty disjoint convex sets, there exists a plane E_f^α which separates them.*

Proof. The set $C - C' = C + (-C')$ is convex and does not contain the origin. Therefore, by Lemma 2, there exists a linear function f, not identically zero, such that

$$c \in C \text{ and } c' \in C' \;\Rightarrow\; f(c) - f(c') = f(c - c') \geqq 0.$$

We then have

$$\inf_{c \in C} f(c) \geqq \sup_{c' \in C'} f(c').$$

Second separation theorem. *If C and C' are two non-empty disjoint convex sets and if C is compact and C' is closed, there exists a plane E_f^α which separates C and C' strictly: that is,*

$$\sup_{x \in C'} f(x) < \alpha < \inf_{x \in C} f(x).$$

Proof. The set $(-C')$ is convex and closed. Therefore, since C is convex and compact, the set $C - C' = C + (-C')$ is convex and closed, by Corollary 3 to Theorem 2. Moreover $0 \notin C - C'$, since $C \cap C' = \varnothing$. Therefore, by Lemma 1, there exists a linear form f and a number $\lambda > 0$ such that

$$x \in C - C' \;\Rightarrow\; f(x) > \lambda.$$

Therefore

$$c \in C \text{ and } c' \in C' \;\Rightarrow\; f(c) - f(c') > \lambda > 0,$$

whence

$$\inf_{c \in C} f(c) \geqq \sup_{c' \in C'} f(c') + \lambda > \sup_{c' \in C'} f(c').$$

Thus there exists a number α such that

$$\sup_{c' \in C'} f(c') < \alpha < \inf_{c \in C} f(c)$$

and E_f^{α} is then a plane with the required properties.

An important application of this result to Economics is obtained from the following:

Farkas' corollary. *Let $a_1, a_2, \dots, a_p, b$ be points of $\mathbf{R}^n$ other than 0 and suppose that each solution x of the system*

$$\langle a_i, x \rangle \geqq 0 \qquad (i = 1, 2, \dots, p)$$

satisfies $\langle b, x \rangle \geqq 0$; then there exist coefficients $\lambda_1, \lambda_2, \dots, \lambda_p$ all $\geqq 0$, such that

$$b = \lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_p a_p.$$

Proof. Let A be the convex cone generated by the points $a_1, a_2, \dots, a_p$; by Theorem 1 of § 4, Chapter VII, every point b^k can be written

$$b^k = \lambda_1^k a_1 + \lambda_2^k a_2 + \dots + \lambda_p^k a_p,$$

where

$$\lambda_1^k, \lambda_2^k, \dots, \lambda_p^k \geqq 0.$$

In particular, we deduce that the set A is closed, for if a sequence of points b^k converges to a point b^0, then this point also satisfies the above conditions and so is in A.

Suppose that the point b does not belong to A. Then the set $\{b\}$ is a compact convex set and A is a closed convex set not meeting $\{b\}$. Therefore they can be separated strictly by a plane with equation $\langle t, x \rangle = \alpha$, such that $\langle t, b \rangle < \alpha < \inf_{x \in A} \langle t, x \rangle$. Since $0 \in A$, we have $\alpha < 0$; but no point a of the cone A satisfies $\langle t, a \rangle < 0$ and we can therefore choose α as close to 0 as we please; hence we have $\langle a_i, t \rangle \geqq 0$ $(i = 1, 2, \dots, p)$, but $\langle b, t \rangle < 0$, contrary to hypothesis. It follows that $b \in A$ and so the result is proved.

Intersection theorem[(1)]. *In the space $\mathbf{R}^n$ let $C_1, C_2, \dots, C_m$ be compact convex sets whose union is a convex set. If the intersection of any $m-1$ of them is non-empty, then the intersection of all the C_j is non-empty.*

[(1)] This result is important in numerous applications (c.f., for example, Sion's theorem, p. 210) and has recently given rise to various generalisations (c.f. Berge, *C.R.Ac.Sc.* Paris, **248**, 1959, p. 2698; Ghouila-Houri, *C.R.Ac.Sc.* Paris, **252**, 1961, p. 494). The result is still true if the C_j, instead of being compact, are merely closed; this follows from the theorem on p. 169.

Proof. We shall first prove the theorem in the case $m = 2$. Let C_1 and C_2 be two non-empty compact convex sets such that $C_1 \cup C_2$ is convex. If C_1 and C_2 are disjoint, there is a plane P which separates them strictly, by the second separation theorem. Since there exist points of $C_1 \cup C_2$ on both sides of P and since $C_1 \cup C_2$ is convex, there exist points of $C_1 \cup C_2$ on P; but this is impossible since P meets neither C_1 nor C_2.

Suppose now that the result holds for $m = r$ convex sets; we shall prove that this implies that it holds for $m = r+1$ convex sets $C_1, C_2, \ldots, C_{r+1}$. Put $C = \bigcap_{j=1}^{r} C_j$; by hypothesis $C \neq \varnothing$, $C_{r+1} \neq \varnothing$. Suppose that these two sets are disjoint; then there exists a plane P which separates them strictly. Writing $C'_j = C_j \cap P$, we have

$$\bigcup_{j=1}^{r} C'_j = \bigcup_{j=1}^{r} (C_j \cap P) \cup (C_{r+1} \cap P) = P \cap \Big(\bigcup_{j=1}^{r+1} C_j\Big).$$

Therefore the union of the sets $C'_1, C'_2, \ldots, C'_r$ is convex. Also the intersection of any $r-1$ of $C_1, C_2, \ldots, C_r$ meets C and C_{r+1} and hence meets P; therefore the intersection of any $r-1$ of $C'_1, C'_2, \ldots, C'_r$ is not empty. But, by hypothesis, this implies that

$$\bigcap_{j=1}^{r} C'_j = C \cap P \neq \varnothing,$$

contradicting the fact that P is a plane which separates C and C_{r+1} strictly. It follows that

$$C \cap C_{r+1} \neq \varnothing$$

and so the result holds for $m = r+1$.

Hence, by induction, the theorem is true for all $m \geqq 2$.

Helly's theorem. *Let $C_1, C_2, \ldots, C_m$ (where $m \geqq n+1$) be compact convex sets in $\mathbf{R}^n$. If the intersection of any $n+1$ of them is non-empty, then the intersection of all the C_j is non-empty.*

Proof. We shall show that if $p \geqq n+1$ and the intersection of any p of $C_1, C_2, \ldots, C_{p+1}$ is non-empty, then the intersection of $C_1, C_2, \ldots, C_{p+1}$ is non-empty. Consider the sets

$$\bigcap_{\substack{j \leq p+1 \\ j \neq i}} C_j \qquad (i = 1, 2, \ldots, p+1).$$

If these are non-empty, there exist points $a_1, a_2, \ldots, a_{p+1}$, one belonging to each set. If

$$A = \{a_1, a_2, \ldots, a_{p+1}\}$$

then, by Theorem 4 of § 4, Chapter VII, each point of the convex closure $[A]$ can be expressed as a linear combination of $n+1$ points of $[A]$, with coefficients $p_1, p_2, \ldots, p_{n+1} \geqq 0$ such that $\Sigma p_i = 1$. Therefore each point of $[A]$ belongs to at least one of the C_j and so

$$[A] \subset \bigcup_{j=1}^{p+1} C_j.$$

Putting $C'_j = C_j \cap [A]$, we obtain a set of compact convex sets C'_j whose union is a convex set $[A]$. The intersection of any p of these sets is non-empty; therefore, by the Intersection Theorem, the intersection of all the C'_j is non-empty. Hence the intersection of all the C_j is non-empty.

Theorem 3. *A closed convex set is equal to the intersection of the half-spaces which contain it.*

Proof. Let C be a closed convex set and let A be the intersection of the half-spaces which contain it. If $x_0 \notin C$, then $\{x_0\}$ is a compact convex set which does not meet C; therefore there exists a plane E_f^α such that

$$f(x_0) < \alpha < \inf_{x \in C} f(x).$$

We then have $H_f^\alpha \supset C$ and $x_0 \notin H_f^\alpha$; consequently x_0 does not belong to the intersection of the half-spaces H_f^α containing C: that is, $x_0 \notin A$. Hence $C \supset A$ and so, since $A \supset C$ by definition, we have $A = C$.

Let C be any set. A plane E_f^α containing at least one point of C and such that all the points of C are on one side of E_f^α, is called a **plane of support** of C (see Chapter VII, § 4). We observe that, if C is compact, then, for any linear function f which is not identically zero, there exists a plane of support having equation $f(x) = \alpha$ (it is sufficient to take $\alpha = \min_{x \in C} f(x)$).

'Plane of support' theorem. *If C is a compact non-empty convex set, it admits an extreme point; in fact every plane of support contains an extreme point of C.*

Proof. (1) The theorem is true in **R**, for a compact convex set in **R** is a closed segment $[\alpha, \beta]$ and contains two extreme points α and β; the planes of support $\{x \,/\, x = \alpha\}$ and $\{x \,/\, x = \beta\}$ contain α and β respectively.

(2) Suppose now that the theorem holds in $\mathbf{R}^r$. We shall prove that this implies that it holds in $\mathbf{R}^{r+1}$. Let C be a compact convex set in $\mathbf{R}^{r+1}$ and let E_f^α be a plane of support. The intersection $E_f^\alpha \cap C$ is a non-empty closed convex set; since $E_f^\alpha \cap C$ is contained in the compact set C, it is also a compact set.

The set $E_f^\alpha \cap C$ can be regarded as a compact convex set in $\mathbf{R}^r$ and so, by hypothesis, it admits an extreme point x_0. Let $[x_1, x_2]$ be a linear segment of centre x_0, with $x_1 \neq x_0$ and $x_2 \neq x_0$. Since x_0 is an extreme point of $E_f^\alpha \cap C$, we have $[x_1, x_2] \not\subset E_f^\alpha \cap C$. Therefore, if x_1 and $x_2 \in C$, we have $x_1, x_2 \notin E_f^\alpha$ and hence x_1, x_2 are separated by E_f^α; but this contradicts the definition of E_f^α as a plane of support of C. It follows that there is no segment $[x_1, x_2]$ of centre x_0 contained in C and so x_0 is an extreme point of C; by definition x_0 is in E_f^α.

Thus, if the theorem holds for $n = r$, it holds for $n = r+1$; but, as we have seen, it holds for $n = 1$. Hence, by induction, the theorem is true for all $n \geqq 1$.

The **closed-convex closure** $\bar{c}\,[A]$ of a set A is the intersection of the closed convex sets containing A.

Theorem of Krein and Milman. *A non-empty compact convex set C is equal to the closed-convex closure of its profile $\ddot{C}$.*

Proof. Since $\ddot{C} \subset C$, we have

$$\bar{c}\,[\ddot{C}] \subset \bar{c}\,[C] = C.$$

To prove the theorem, it is therefore sufficient to show that $C \subset \bar{c}\,[\ddot{C}]$. Suppose that this is false; then there exists a point x_0 in C such that $x_0 \notin \bar{c}\,[\ddot{C}]$. By the second separation theorem, there exists a linear function f such that

$$f(x_0) < \inf \{f(x) \,/\, x \in \bar{c}\,[\ddot{C}]\}$$

and so

$$f(C) \not\subset f(\bar{c}\,[\ddot{C}]).$$

On the other hand, $f(C)$ is a compact convex set in $\mathbf{R}$ and is therefore a segment $[\alpha, \beta]$. By the preceding theorem, the plane E_f^α, (which is a plane of support of C), contains an extreme point. Then

$$\alpha \in f(\bar{c}\,[\ddot{C}]).$$

By symmetry, a similar relation holds with β in place of α, and hence, since $f(\bar{c}\,[\ddot{C}])$ is a convex set (see Theorem 5, § 3, Chapter VII), we also have

$$f(C) = [\alpha, \beta] \subset f(\bar{c}\,[\ddot{C}]).$$

Thus we have reached a contradiction and so the hypothesis $C \not\subset \bar{c}\,[\ddot{C}]$ is false.

If C and D are two compact convex sets, each having at least one interior point, then they are homeomorphic; this can be proved by using the idea of radial projection (cf. page 154), as shown in the following theorem.

Theorem of Sz. Nagy. *If two closed convex sets C and D in $\mathbf{R}^n$ each have* 0 *as an interior point and are bounded, then the radial projection $y = \sigma x$ of C on D is a contracting mapping in C and is also a homeomorphism.*

Proof. Let $B_\lambda(0)$ be a ball containing $C \cup D$ and let $B_\mu(0)$ be a ball contained in $C \cap D$. Let j and k be the gauges of C and D respectively; we have

$$\frac{1}{\lambda} \| x \| \leqq j(x) \leqq \frac{1}{\mu} \| x \|;$$

also

$$| j(x)-j(x') | \leqq \max \{j(x-x'), j(x'-x)\} \leqq \frac{1}{\mu} \| x-x' \|$$

and similar inequalities hold with $k(x)$ in place of $j(x)$.

We must show that σ is a contracting mapping; that is, that there exists a number α such that

$$(\forall_C x)\,(\forall_C x') : \| \sigma x - \sigma x' \| \leqq \alpha \| x-x' \|.$$

Let x and x' be points of C such that $k(x) \geqq k(x')$ and write

$u = \dfrac{x}{k(x)}$, $v = \dfrac{x'}{k(x)}$; we have

$$\begin{aligned}
\frac{1}{k(x)}\left\| \frac{j(x)}{k(x)}\, x - \frac{j(x')}{k(x')}\, x' \right\| &= \left\| j(u)u - j(v)\,\frac{k(x)}{k(x')}\, v \right\| = \left\| j(u)u - \frac{j(v)}{k(v)}\, v \right\| \\
&= \left\| j(u)\,(u-v) + \left[j(u) - \frac{j(v)}{k(v)} \right] v \right\| \\
&= \left\| j(u)\,(u-v) + [j(u)-j(v)]v + \frac{j(v)}{k(v)}\,[k(v)-k(u)]v \right\| \\
&\leqq \left[j(u) + \frac{1}{\mu} \| v \| + \frac{1}{\mu}\frac{j(v)}{k(v)} \| v \| \right] \| u-v \|
\end{aligned}$$

Then

$$\begin{aligned}
\| \sigma x - \sigma x' \| &\leqq \left[\frac{j(x)}{k(x)} + \frac{1}{\mu}\frac{\| x' \|}{k(x)} + \frac{1}{\mu}\frac{j(x')}{k(x')}\frac{\| x' \|}{k(x)} \right] \| x-x' \| \\
&\leqq \left[\frac{1/\mu}{1/\lambda} + \frac{1}{\mu} \cdot \lambda + \frac{1/\mu^2}{1/\lambda} \cdot \lambda \right] \| x-x' \| = \alpha \| x-x' \|
\end{aligned}$$

It follows that σ is continuous in C; by symmetry σ^{-1} is continuous in D and therefore σ is a homeomorphism.

§ 2. Simplexes; Kakutani's Theorem

We have seen that every closed convex set C is the intersection of the closed half-spaces which contain it. If a set T is the intersection of a finite number of closed half-spaces $\bar{H}_f^\alpha$, we say that T is a **truncation**. The planes E_f^α are called the **generating planes** of T. A truncation is a closed convex set, since it is the intersection of sets H_f^α which are closed and convex.

A bounded truncation is called a **convex polyhedron**; a convex polyhedron is a compact convex set (it is compact since it is closed and bounded in $\mathbf{R}^n$). The extreme points of a truncation or a convex polyhedron are called its **vertices.**

EXAMPLE 1. The unit cube

$$K = \{x \,/\, 0 \leqq x^1 \leqq 1, 0 \leqq x^2 \leqq 1, \ldots, 0 \leqq x^n \leqq 1\}$$

is a truncation generated by the half-spaces

$$\{x \,/\, x^i \leqq 1\} \quad \text{and} \quad \{x \,/\, x^i \geqq 0\}.$$

Since K is bounded, it is a convex polyhedron. Its vertices are the points

$$e^I = \{x \,/\, x^i = 0 \text{ if } i \in I, x^i = 1 \text{ if } i \notin I\}$$

where

$$I \subset \{1, 2, \ldots, n\}.$$

EXAMPLE 2. The set $\mathbf{P}_n = \{x/x^i \geqq 0 \text{ for all } i, \sum_{i=1}^{n} x^i = 1\}$ is a truncation generated by the half-spaces $\{x \,/\, x^i \geqq 0\}$ and the half-spaces $\{x/\sum_{i=1}^{n} x^i \leqq 1\}$ and $\{x/\sum_{i=1}^{n} x^i \geqq 1\}$. Since $\mathbf{P}_n$ is bounded, it is a convex polyhedron; its vertices are the points $\delta_1, \delta_2, \ldots, \delta_n$ (where δ_j^i is the Kronecker symbol).

If a is a point of a truncation T, we denote by V_a the linear variety formed by the privileged lines of T passing through a (cf. page 148); if a is a vertex, so that $V_a = \emptyset$, we also say that $\{a\}$ is a **face of order** 0; if V_a has dimension 1, we say that $V_a \cap T$ is an **edge**, or a face of order 1, of T; if V_a has dimension k, we say that $V_a \cap T$ is a **face of order** k of T. Since every face is an intersection of generating planes, they are finite in number.

Theorem 1. *If $\{a_1, a_2, \ldots, a_k\}$ is a finite set in $\mathbf{R}^n$, its convex closure $[\{a_1, a_2, \ldots, a_k\}] = [a_1, a_2, \ldots, a_k]$ is a compact set.*

Proof. Let ρ be the single-valued mapping of $\mathbf{P}_k$ into $\mathbf{R}^n$ defined by

$$\rho(p_1, p_2, \ldots, p_k) = p_1 a_1 + p_2 a_2 + \ldots + p_k a_k.$$

Then ρ is a continuous mapping, for

$$\max_i |p_i - p_i'| \leqq \frac{\varepsilon}{\Sigma \| a_i \|} \Rightarrow \left| \sum_i p_i a_i - \sum_i p_i' a_i \right| \leqq \sum_i |p_i - p_i'| \cdot \| a_i \| \leqq \varepsilon.$$

Since $\mathbf{P}_k$ is a compact set in $\mathbf{R}^k$, so is its image

$$\rho\, \mathbf{P}_k = [a_1, a_2, \ldots, a_k].$$

COROLLARY. *If C is a convex polyhedron, its profile $\ddot{C}$ is a finite set and $C = [\ddot{C}]$.*

Proof. The extreme points of C are finite in number since they are faces of order 0; therefore $\ddot{C}$ is a finite set. Since C is a compact convex set, it is equal to the closed-convex closure of $\ddot{C}$, by the theorem of Krein and Milman; by the above theorem, $[\ddot{C}]$ is a closed set and so

$$C = \bar{c}\,[\ddot{C}] = [\ddot{C}].$$

A convex polyhedron whose vertices $a_1, a_2, \ldots, a_{k+1}$ are $k+1$ linearly independent vectors is called a **k-simplex**; clearly a k-simplex S_k has dimension k. A 0-simplex is a point, a 1-simplex is a linear segment, a 2-simplex is called a triangle, a 3-simplex is called a tetrahedron. For $k > n$, there are no k-simplexes.

Let S_k be the k-simplex $[\{a_1, a_2, \ldots, a_{k+1}\}] = [a_1, a_2, \ldots, a_{k+1}]$: if $q < k$, a face of order q of this simplex is just the convex closure of $q+1$ of the a_i; thus a face of the simplex is also a simplex.

Two faces S and T of S_k are said to be **opposite** if

$$\ddot{S} \cup \ddot{T} = \ddot{S}_k, \quad \ddot{S} \cap \ddot{T} = \emptyset.$$

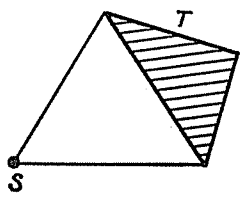


FIG. 37

Put

$$]a_1, a_2, \ldots, a_k[=$$

$$\{p_1a_1+p_2a_2+ \ldots +p_ka_k \,/\, p_1 > 0, p_2 > 0, \ldots, p_k > 0, \sum_{i=1}^{k} p_i = 1\}.$$

Such a set is called an **open simplex**; an open 0-simplex is a point, an open 1-simplex is an interval $]x_1, x_2[$, an open 2-simplex is the interior of a triangle, etc.

If $T_k =]a_1, a_2, \ldots, a_{k+1}[$ is an open simplex, we shall again use the term vertex for the points a_i and we shall denote the set $\{a_1, a_2, \ldots a_{k+1}\}$ by $\ddot{T}_k$. An open simplex such as $]a_{i_1}, a_{i_2}, a_{i_3}[$ (where $i_1, i_2, i_3 \leqq k+1$) will be called a *face* of the open simplex T_k.

Let $S_n = [a_1, a_2, \ldots, a_{n+1}]$ be an n-simplex. A family $\mathscr{T}$ of open simplexes T^i such that

(1) $\bigcup T^i = S_n$
(2) $i \neq j$ implies that $T^i \cap T^j = \emptyset$,
(3) every face of a T^i belongs to $\mathscr{T}$,

is called a **triangulation** of S_n.

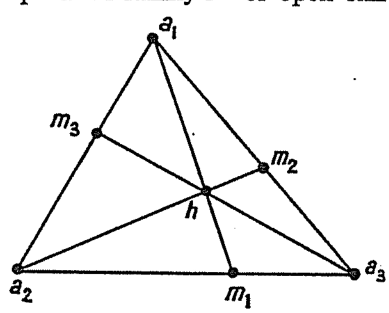


FIG. 38

EXAMPLE 1. Consider the triangle $S_2 = [a_1, a_2, a_3]$ in $\mathbf{R}^2$. The triangulation $\mathscr{T}$ represented in figure 38 consists of

(1) the open 2-simplexes

$$]a_1, h, m_3[,\ ]m_3, h, a_2[,\ ]a_2, h, m_1[,\ ]m_1, h, a_3[,\ ]a_3, h, m_2[,\ ]m_2, h, a_1[$$

(2) the open 1-simplexes

$$]a_1, h[,\ ]a_2, h[,\ ]a_3, h[,\ ]m_1, h[,\ ]m_2, h[,\ ]m_3, h[$$
$$]a_1, m_3[,\ ]m_3, a_2[,\ ]a_2, m_1[,\ ]m_1, a_3[,\ ]a_3, m_2[,\ ]m_2, a_1[$$

(3) the open 0-simplexes

$$\{a_1\},\ \{a_2\},\ \{a_3\},\ \{m_1\},\ \{m_2\},\ \{m_3\},\ \{h\}$$

EXAMPLE 2. If S_n is an n-simplex $[a_1, a_2, \ldots, a_{n+1}]$, the sets $\{a_i\}$, $]a_i, a_j[$, $]a_i, a_j, a_k[$, etc. are open simplexes of a triangulation of S_n. These sets are disjoint, for otherwise we should have (for example)

$$]a_1, a_2[\, \cap\,]a_3, a_4, a_5[\ \neq \varnothing,$$

so that

$$p_1a_1+p_2a_2 = p_3a_3+p_4a_4+p_5a_5,$$

for certain numbers p_1, p_2, p_3, p_4, p_5 not all zero, which would imply that a_1, a_2, a_3, a_4, a_5 were not linearly independent. The sets of this triangulation of S_n are called the **elementary faces** of the simplex S_n.

Sperner's lemma. *Let $\mathscr{T}$ be a triangulation of an n-simplex*

$$S_n = [a_1, a_2, \ldots, a_{n+1}],$$

and let σ be a single-valued mapping which maps each vertex e of an open simplex of $\mathscr{T}$ onto a vertex $a_i = \sigma e$ of the elementary face of S_n containing e. Then there exists an open n-simplex T_n of $\mathscr{T}$ such that $\sigma\ddot{T}_n = \{a_1, a_2, \ldots, a_{n+1}\}$.

Proof. Let $\mathscr{T}_0$ be the set of n-simplexes T_n of $\mathscr{T}$ such that

$$\sigma\ddot{T}_n = \{a_1, a_2, \ldots, a_{n+1}\}.$$

We shall in fact prove that the number q of elements of $\mathscr{T}_0$ is odd.

For $n = 0$, this result is clearly true. We shall show that the assumption that it is true for $n = r$ implies that it is true for $n = r+1$; the theorem then follows by induction.

Let $T^1, T^2, \ldots, T^m$ be the $(r+1)$-simplexes of a triangulation $\mathscr{T}$ of an $(r+1)$-simplex $S_{r+1} = [a_1, a_2, \ldots, a_{r+2}]$. We say that an r-face of T^i is *marked* if its vertices have $a_1, a_2, \ldots, a_{r+1}$ as images. Let r_i be the number of marked faces of the $(r+1)$-simplex T^i. We shall determine the sum $r_1+r_2+\ldots+r_m$ in two different ways.

(1) Consider the following different cases:

$$\begin{array}{ll} \text{(i)} & T^i \in \mathcal{T}_0 \quad\Rightarrow\quad r_i = 1, \\ \text{(ii)} & T^i \notin \mathcal{T}_0,\ \sigma\ddot{T}_i \not\supset \{a_1, a_2, \ldots, a_{r+1}\} \quad\Rightarrow\quad r_i = 0, \\ \text{(iii)} & T^i \notin \mathcal{T}_0,\ \sigma\ddot{T}_i \supset \{a_1, a_2, \ldots, a_{r+1}\} \quad\Rightarrow\quad r_i = 2. \end{array}$$

The first two of these are immediate. To prove (iii), suppose that $T^i \notin \mathcal{T}_0$ and that $\ddot{T}^i = \{e_1, e_2, \ldots, e_{r+2}\}$ is such that $\sigma e_1 = a_1, \sigma e_2 = a_2, \ldots, \sigma e_{r+1} = a_{r+1}$; then we have $\sigma e_{r+2} \neq a_{r+2}$ (for $T^i \notin \mathcal{T}_0$). If, for example, $\sigma e_{r+2} = a_1$, then T^i has two marked faces $[e_1, e_2, \ldots, e_{r+1}]$ and $[e_{r+2}, e_2, \ldots, e_{r+1}]$.

It follows from these considerations that

$$(r_1 + r_2 + \ldots + r_m) = q + 2k; \qquad k \in \mathbf{N}.$$

(2) Let T_r be an r-simplex of the triangulation $\mathcal{T}$. Then the following possibilities can occur:

(i) T_r is not contained in any face of order r of S_{r+1} $\Rightarrow$ (*a*) T_r is not a marked face of any T_{r+1}, or (*b*) T_r is a marked face of T_{r+1} and T'_{r+1} (which have T_r in common).

(ii) T_r is contained in a face of order r of S_{r+1} but this face is not $[a_1, a_2, \ldots, a_{r+1}]$ $\Rightarrow$ T_r is not a marked face.

(iii) $T_r \subset [a_1, a_2, \ldots, a_{r+1}]$ $\Rightarrow$ (*a*) if two vertices of T_r have the same a_i for image, then T_r is not a marked face of any T_{r+1}, (*b*) if the vertices of T_r have distinct images, T_r is the marked face of only one T_{r+1}.

Let q' be the number of T_r contained in $S_r = [a_1, a_2, \ldots, a_{r+1}]$ and whose vertices have distinct images; we then have

$$(r_1 + r_2 + \ldots + r_m) = q' + 2k'; \qquad k' \in \mathbf{N}.$$

It follows that q and q' have the same parity. By hypothesis, q' is odd and therefore q is also odd.

Kuratowski-Knaster-Mazurkiewicz theorem. *Let $S_n = [a_1, a_2, \ldots, a_{n+1}]$ be an n-simplex and let $F_1, F_2, \ldots, F_{n+1}$ be $n+1$ closed sets. If for each set $\{i, j, \ldots, l\}_f \subset \{1, 2, \ldots, n+1\}$ we have*

$$[a_i, a_j, \ldots, a_l] \subset F_i \cup F_j \cup \ldots \cup F_l,$$

then the F_i have a non-empty intersection.

Proof. We observe that the F_i form a closed covering of the compact set S_n, since

$$S_n = [a_1, a_2, \dots, a_{n+1}] \subset F_1 \cup F_2 \cup \dots \cup F_{n+1}.$$

By Lebesgue's theorem (§ 6, Chapter V) there exists a number ε such that, for each set A of diameter less than ε, the intersection of the F_i meeting A is non-empty. Moreover, we can always find a triangulation $\mathscr{T}$ of S_n whose simplexes T have diameter $\delta(T) \leqq \varepsilon$. With each vertex e of a simplex of $\mathscr{T}$, we associate a vertex $a_i = \sigma e$ of the elementary face of S_n containing e, such that $F_i \ni e$. By Sperner's lemma, there exists an n-simplex T_n of $\mathscr{T}$ such that $\sigma\ddot{T}_n = \{a_1, a_2, \dots, a_{n+1}\}$. In other words the sets $F_1, F_2, \dots, F_{n+1}$ each contain an element of $\ddot{T}_n$, so that

$$(\forall i) : F_i \cap \ddot{T}_n \neq \emptyset.$$

Since $\delta(\ddot{T}_n) \leqq \varepsilon$, we have

$$\bigcap_{i=1}^{n} F_i \neq \emptyset.$$

COROLLARY. *Let $F_1, F_2, \dots F_{n+1}$ be closed sets covering an n-simplex $S_n = [a_1, a_2, \dots, a_{n+1}]$, such that F_i contains a_i and does not meet the face opposite to a_i. Then the F_i have a non-empty intersection.*

Proof. Let $S^i_{n-1} = [a_1, a_2, \dots, a_{i-1}, a_{i+1}, a_{i+2}, \dots, a_{n+1}]$ be the face of S_n opposite to a_i. If $S_k = [a_{i_1}, a_{i_2}, \dots, a_{i_{k+1}}]$ is a face of S_n and if $i \neq i_1, i_2, \dots, i_{k+1}$, we have $S_k \subset S^i_{n-1}$. Since F_i does not meet S^i_{n-1}, F_i does not meet S_k. This is true for all $i \neq i_1, i_2, \dots, i_{k+1}$, and so

$$S_k = [a_{i_1}, a_{i_2}, \dots, a_{i_{k+1}}] \subset F_{i_1} \cup F_{i_2} \cup \dots \cup F_{i_{k+1}}.$$

Therefore, by the theorem, we have

$$\bigcap_{i=1}^{n} F_i \neq \emptyset.$$

We now use the above results to prove certain deep properties of mappings of $\mathbf{R}^n$ into $\mathbf{R}^n$.

Theorem 2. *If Γ_1 and Γ_2 are two u.s.c. (resp. l.s.c.) mappings of $X = \mathbf{R}^n$ into itself, the mapping $\Gamma_1 + \Gamma_2$ is u.s.c. (resp. l.s.c.).*

Proof. The mapping defined by $(\Gamma_1 \times \Gamma_2)x = \Gamma_1 x \times \Gamma_2 x$ is a u.s.c. mapping of X into $X \times X$; the single valued mapping σ defined by $\sigma(x, y) = x + y$ is a continuous mapping of $X \times X$ into X and so it is u.s.c. Hence, by Theorem 1′ of § 2, Chapter VI,

$$\Gamma_1 + \Gamma_2 = \sigma(\Gamma_1 \times \Gamma_2)$$

is a u.s.c. mapping of X into X.

Kakutani's theorem.[1] *Let C be a non-empty compact convex set in* $\mathbf{R}^n$*; if* Γ *is a u.s.c. mapping of C into C and if the set* Γx *is convex and non-empty for each x, then there exists a point* x_0 *in C such that*

$$x_0 \in \Gamma x_0.$$

Proof. (1) We first prove the theorem in the case in which C is an n-simplex $[a_1, a_2, \ldots, a_{n+1}]$ and Γ is a single-valued mapping.

Every simplex of dimension n has an internal point. Without loss of generality, we can assume that the origin 0 is an internal point (we can always make a translation if necessary). Consider the convex cone

$$K_i = \{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_{i-1} a_{i-1} + \lambda_{i+1} a_{i+1} + \ldots + \lambda_{n+1} a_{n+1} \,/\, \lambda_1, \ldots, \lambda_{n+1} \geqq 0\}$$

and let $B_\lambda(0)$ be a ball of centre 0 and radius λ chosen sufficiently large for $B_\lambda(x)$ to contain C for all $x \in C$. The cone K_i is closed and so

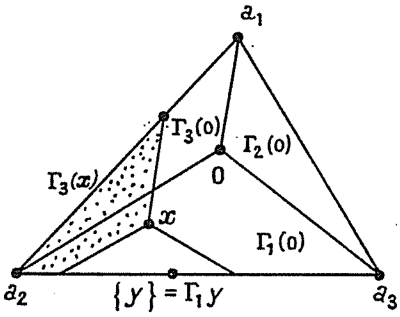


FIG. 39

$$K_i' = K_i \cap B_\lambda(0)$$

is compact. By the preceding theorem, the mapping defined by

$$\Gamma_i(x) = (x + K_i')$$

is u.s.c.; this is also true for the mapping $\Gamma \cap \Gamma_i$, and so the set

$$(\Gamma \cap \Gamma_i)^+ \varnothing = \{x \,/\, \Gamma x \cap \Gamma_i x = \varnothing\}$$

is open.

Let F_i be the complement of this set. Then F_i is the set of points x such that Γx meets $\Gamma_i x$ and F_i is closed. The sets F_i form a covering of C (for $\Gamma_1 x \cup \Gamma_2 x \cup \ldots \cup \Gamma_{n+1} x \supset C$ and so Γx meets at least one of the $\Gamma_i x$). Furthermore, F_i contains a_i, for

$$\Gamma_i a_i \supset C \supset \Gamma a_i.$$

If F_i meets the face S_{n-1}^i opposite to a_i, there exists a point y such that

$$y \in S_{n-1}^i; \; \Gamma y \cap \Gamma_i y \neq \varnothing.$$

This implies that $\Gamma y = y$, which is the desired result.

Suppose therefore that F_i does not meet the face S_{n-1}^i opposite to a_i. Then, by the corollary to the Kuratowski-Knaster-Mazurkiewicz theorem, we have

$$\bigcap_{i=1}^{n} F_i \neq \varnothing.$$

[1] Although this theorem is of a purely set-theoretic character, it has never been proved without the assistance of combinatorial arguments, except in **R**.

If x_0 belongs to this intersection, then Γx_0 meets all the $\Gamma_i x_0$; therefore $x_0 = \Gamma x_0$.

(2) We now prove the theorem in the case in which C is an n-simplex S and Γ is a u.s.c. multi-valued mapping.

Given a simplex $[a_1, a_2, \ldots, a_j]$, the point $\frac{1}{j}(a_1+a_2+\ldots+a_j)$ is called the **barycentre** of this simplex. The barycentres of all the faces of orders 0, 1, 2, ... of a simplex S determine a family $\mathscr{S}^{(1)}$ of n-simplexes, called the **barycentric division** of order 1 of S, with the properties:

(i) each n-simplex of $\mathscr{S}^{(1)}$ contains the barycentre of a face of dimension 0, of a face of dimension 1, etc.,

(ii) if an n-simplex of $\mathscr{S}^{(1)}$ contains the barycentres of $[a_{j_1}, a_{j_2}, \ldots, a_{j_k}]$ and $[a_{i_1}, a_{i_2}, \ldots, a_{i_p}]$, where $k < p$, then

$$\{j_1, j_2, \ldots, j_k\} \subset \{i_1, i_2, \ldots, i_p\}.$$

If we divide each n-simplex of $\mathscr{S}^{(1)}$ in a similar manner, we obtain a new family of n-simplexes, called the barycentric division of order 2, which we denote by $\mathscr{S}^{(2)}$. Continuing this process, we can define the barycentric division $\mathscr{S}^{(k)}$ of order k.

For each vertex a^k of the kth barycentric division $\mathscr{S}^{(k)}$ of the simplex S, let b^k be a point of Γa^k and write $b^k = \phi_k(a^k)$. For each $x \in S$, we consider an n-simplex $[a^k_{i_1}, a^k_{i_2}, \ldots, a^k_{i_{n+1}}] \in \mathscr{S}^{(k)}$ which contains it and we put

$$\phi_k(x) = \phi_k(p_1 a^k_{i_1} + p_2 a^k_{i_2} + \ldots) = p_1 \phi_k(a^k_{i_1}) + p_2 \phi_k(a^k_{i_2}) + \ldots$$

The function ϕ_k so defined is linear in the interior of a simplex of $\mathscr{S}^{(k)}$ and is therefore continuous; moreover it is uniquely determined, even at the points which belong to several n-simplexes of $\mathscr{S}^{(k)}$; consequently ϕ_k is continuous in S. Therefore, by part (1) of the proof, there exists a point $x_k \in S$ such that $x_k = \phi_k(x_k)$. Since S is compact, the sequence (x_k) has a cluster point x_0; we shall prove that $x_0 \in \Gamma x_0$.

If $[a^k_1, a^k_2, \ldots, a^k_{n+1}]$ is an n-simplex of $\mathscr{S}^{(k)}$ which contains the point x_k, we can write

$$x_k = \phi_k(x_k) = \phi_k(p^k_1 a^k_1 + p^k_2 a^k_2 + \ldots) = p^k_1 \phi_k(a^k_1) + p^k_2 \phi_k(a^k_2) + \ldots$$
$$= p^k_1 b^k_1 + p^k_2 b^k_2 + \ldots + p^k_{n+1} b^k_{n+1}$$

where $b^k_1 \in \Gamma a^k_1$ and $b^k_2 \in \Gamma a^k_2$, etc.

Let (x_{k_m}) be a sub-sequence of (x_k) which converges to x_0 and let (l_m) be a sub-sequence of (k_m) such that

$$(b^{l_m}_i) \to b^0_i \qquad (i = 1, 2, \ldots, n+1),$$
$$(p^{l_m}_i) \to p^0_i \qquad (i = 1, 2, \ldots, n+1).$$

We have

$$p_i^0 \geqq 0, \qquad \sum_{i=1}^{n+1} p_i^0 = 1, \qquad \sum p_i^0 b_i^0 = x_0.$$

Since $(a_i^{l_m}) \to x_0$, $(b_i^{l_m}) \to b_i^0$, $b_i^{l_m} \in \Gamma a_i^{l_m}$ and Γ is u.s.c., we have

$$b_i^0 \in \Gamma x_0.$$

But, by hypothesis, Γx_0 is convex, and so

$$x_0 = \sum p_i^0 b_i^0 \in \Gamma x_0.$$

(3) Finally, we prove the theorem in the case in which C is any compact convex set.

Let C have dimension n, choose an interior point of C for origin 0 and consider an n-simplex S which contains C in its interior. Let ϕ be the single-valued mapping defined by

$$\begin{aligned} \phi(x) &= x \qquad \text{if } x \in C, \\ \phi(x) &= \text{the extremity of the segment } [0, x] \cap C \\ &\qquad\qquad \text{if } x \in S - C. \end{aligned}$$

Then ϕ is a continuous mapping of S into C and $\Gamma\phi$ is a u.s.c. mapping defined in S. By part (2) of the proof, there exists a point x_0 such that

$$x_0 \in \Gamma[\phi(x_0)].$$

Since $\Gamma[\phi(x_0)] \subset C$, it follows that $x_0 \in C$ and so $\phi(x_0) = x_0$. Therefore

$$x_0 \in \Gamma[\phi(x_0)] = \Gamma x_0.$$

COROLLARY (Brouwer's theorem). *If C is a non-empty compact convex set in $\mathbf{R}^n$ and σ is a single-valued continuous mapping of C into C, there exists a point x_0 in C such that $x_0 = \sigma x_0$.*

This follows immediately, because σ satisfies the conditions of Kakutani's theorem.

§ 3. Matrices

In this section we recall some well-known results which are required in the sequel.

Let α be a linear mapping which associates with each point $x = (x^1, x^2, \ldots, x^n)$ of $\mathbf{R}^n$ a point $y = (y^1, y^2, \ldots, y^n) = \alpha(x)$ of $\mathbf{R}^n$. We have

$$y = \alpha\left(\sum_{i=1}^{n} x^i \delta_i\right) = \sum_{i=1}^{n} x^i \alpha(\delta_i) = \sum_{i=1}^{n} x^i a_i,$$

where $\alpha(\delta_i) = a_i = (a_i^1, a_i^2, \dots, a_i^n)$; thus

$$\begin{cases} y^1 = x^1\, a_1^1 + x^2\, a_2^1 + \dots + x^n\, a_n^1, \\ y^2 = x^1\, a_1^2 + x^2\, a_2^2 + \dots + x^n\, a_n^2, \\ \dots\dots\dots\dots\dots\dots \\ y^n = x^1\, a_1^n + x^2\, a_2^n + \dots + x^n\, a_n^n. \end{cases}$$

The linear mapping α is completely determined by the array

$$\begin{pmatrix} a_1^1 & a_2^1 & a_3^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & a_3^2 & \dots & a_n^2 \\ \vdots & \vdots & \vdots & & \vdots \\ a_1^n & a_2^n & a_3^n & \dots & a_n^n \end{pmatrix}$$

Conversely, such an array uniquely determines, by means of the above equations, a linear mapping α. Such an array is called a **matrix of order** n (more precisely a square matrix of order n). We denote the matrix with coefficients a_j^i by $\mathbf{A}$.

The vector $a_i = (a_i^1, a_i^2, \dots, a_i^n)$ is called a **column-vector** of $\mathbf{A}$ and $a^j = (a_1^j, a_2^j, \dots, a_n^j)$ is called a **row-vector** of $\mathbf{A}$. We can write

$$\mathbf{A} = \begin{pmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \vdots & \vdots & & \vdots \\ a_1^n & a_2^n & \dots & a_n^n \end{pmatrix} = (a_1, a_2, \dots, a_n) = \begin{pmatrix} a^1 \\ a^2 \\ \cdot \\ \cdot \\ \cdot \\ a^n \end{pmatrix}.$$

We also put $\mathbf{A}x = \alpha(x)$.

In this section and elsewhere in the present chapter we put

$$\mathrm{N} = \{1, 2, \dots, n\}.$$

In certain cases, it will be convenient to denote the matrix $\mathbf{A}$ by $\mathbf{A}_\mathrm{N}^\mathrm{N}$ to stress the row and column indices of $\mathbf{A}$. If $\mathbf{I} \subset \mathrm{N}$, $\mathbf{J} \subset \mathrm{N}$, the (rectangular) matrix obtained from $\mathbf{A}_\mathrm{N}^\mathrm{N}$ by suppressing the column-vectors a_j for which $j \notin \mathbf{J}$ and the row vectors a^i for which $i \notin \mathbf{I}$, is denoted by $\mathbf{A}_\mathbf{J}^\mathbf{I}$. If $\mathbf{I}, \mathbf{J}, \mathbf{K}, \mathbf{L} \subset \mathrm{N}$ and $\mathbf{I} \cap \mathbf{K} = \emptyset$, $\mathbf{J} \cap \mathbf{L} = \emptyset$, we denote by $\mathbf{A}_\mathbf{J}^\mathbf{I} \otimes \mathbf{B}_\mathbf{L}^\mathbf{K}$ the (square) matrix $\mathbf{C}_\mathrm{N}^\mathrm{N}$ defined by

$$c_j^i \begin{cases} = a_j^i \text{ if } i \in \mathbf{I} \text{ and } j \in \mathbf{J}, \\ = b_j^i \text{ if } i \in \mathbf{K} \text{ and } j \in \mathbf{L}, \\ = 0 \text{ in the other cases.} \end{cases}$$

For any matrix $\mathbf{A}$, we have

$$a_j^i = \langle \delta_i, a_j \rangle = \langle \delta_i, \mathbf{A}\delta_j \rangle.$$

If $\mathbf{A}x = \mathbf{B}x$ for all x, then we say that the matrices $\mathbf{A}$ and $\mathbf{B}$ are **equal**; we then have

$$a_j^i = \langle \delta_i, \mathbf{A}\delta_j \rangle = \langle \delta_i, \mathbf{B}\delta_j \rangle = b_j^i.$$

The **sum** $\mathbf{A}+\mathbf{B}$ of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined to be the matrix $\mathbf{S}$ corresponding to the linear mapping σ such that

$$\sigma(x) = \mathbf{A}x+\mathbf{B}x.$$

We have

$$\begin{aligned} s^i_j = \langle \delta_i, \mathbf{A}\delta_j+\mathbf{B}\delta_j\rangle &= \langle \delta_i, \mathbf{A}\delta_j\rangle+\langle \delta_i, \mathbf{B}\delta_j\rangle \\ &= a^i_j+b^i_j. \end{aligned}$$

The **product** $\lambda\mathbf{A}$ of a matrix $\mathbf{A}$ and a scalar λ is defined to be the matrix $\mathbf{T}$ corresponding to the linear mapping τ such that

$$\tau(x) = \lambda(\mathbf{A}x).$$

We have

$$t^i_j = \langle \delta_i, \lambda\mathbf{A}\delta_j\rangle = \lambda\langle \delta_i, \mathbf{A}\delta_j\rangle = \lambda a^i_j.$$

Writing $\mathbf{O}$ for the matrix of order n whose elements are all zero, we see that the set of matrices of order n forms a vector space; the following properties are easily verified:

(1) $\mathbf{A}+\mathbf{B} = \mathbf{B}+\mathbf{A}$,
(2) $\mathbf{A}+(\mathbf{B}+\mathbf{C}) = (\mathbf{A}+\mathbf{B})+\mathbf{C}$,
(3) $\mathbf{A}+\mathbf{O} = \mathbf{A}$,
(4) $\mathbf{A}+(-1)\mathbf{A} = \mathbf{O}$,
(5) $\lambda(\mathbf{A}+\mathbf{B}) = \lambda\mathbf{A}+\lambda\mathbf{B}$,
(6) $(\lambda+\mu)\mathbf{A} = \lambda\mathbf{A}+\mu\mathbf{A}$,
(7) $(\lambda\mu)\mathbf{A} = \lambda(\mu\mathbf{A})$,
(8) $1\mathbf{A} = \mathbf{A}$.

The vector space determined by the matrices of order n has n^2 dimensions, for we can write

$$\mathbf{A} = \sum_{k,l} a^k_l \mathbf{E}^k_l$$

where $\mathbf{E}^k_l$ denotes the matrix with components $(e^k_l)^i_j$ satisfying

$$\begin{aligned} (e^k_l)^i_j &= 0 \qquad \text{if } i \neq k \text{ or } j \neq l, \\ &= 1 \qquad \text{if } i = k \text{ and } j = l. \end{aligned}$$

Thus

$$\mathbf{E}^k_l = \begin{array}{cc} \begin{array}{ccccccc} & & & & l & & \end{array} & \\ \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \vdots & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & & \cdots & & 0 \end{pmatrix} & \begin{array}{c} \\ \\ \\ \cdots k \\ \\ \\ \end{array} \end{array}$$

In what follows, we shall identify the space of matrices of order n with $\mathbf{R}^{n^2}$.

Let $\mathbf{A}$ be the matrix determined by the linear mapping α; then the **inverse** matrix of $\mathbf{A}$ is defined to be the matrix $\mathbf{B}$ determined by the inverse mapping α^{-1}, provided that this exists (that is, if and only if the determinant of the coefficients a_j^i is non-zero).

The **transposed** matrix of $\mathbf{A}$ is a matrix $\mathbf{A}^*$ such that

$$\langle x, \mathbf{A}^* y\rangle = \langle \mathbf{A}x, y\rangle$$

for all x and all y; $\mathbf{A}^*$ exists for each $\mathbf{A}$ and is unique. We have

$$a^{*i}_{\ j} = \langle \delta_i, \mathbf{A}^*\delta_j\rangle = \langle \mathbf{A}\delta_i, \delta_j\rangle = a_i^j.$$

If $\mathbf{A}$ and $\mathbf{B}$ are the matrices determined by the linear mappings α and β, the **product** $\mathbf{AB}$ is the matrix determined by the mapping $\gamma = \alpha\beta$; if $\mathbf{C} = \mathbf{AB}$, we have

$$\begin{aligned} c_j^i = \langle \delta_i, \mathbf{AB}\delta_j\rangle = \langle \mathbf{A}^*\delta_i, \mathbf{B}\delta_j\rangle &= \langle a^i, b_j\rangle \\ &= \sum_k a_k^i b_j^k. \end{aligned}$$

The **unit matrix** is the matrix $\mathbf{E}$ whose coefficients are the Kronecker symbols δ_j^i, so that

$$\mathbf{E} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ & & 1 & \ddots & \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

REMARK. The above matrices represent single-valued linear mappings. However, the ideas can be extended so as to give matrices which represent linear mappings that are not necessarily single-valued. Let Γ be a multi-valued linear mapping of $\mathbf{R}^n$ into itself; if $x = (x^1, x^2, \dots, x^n) \in \mathbf{R}^n$ we have

$$\Gamma x = \Gamma\,(\sum_{i\in N} x^i \delta_i) = \sum_{i \in N} x^i\, \Gamma(\delta_i).$$

If, for all x, Γx is a Cartesian product of n one-dimensional sets of the form $A^1(x) \times A^2(x) \times \dots \times A^n(x)$, and if we put $A^i(\delta_j) = A_j^i$, then we have

$$\begin{cases} A^1(x) = x^1 A_1^1 + x^2 A_2^1 + \dots + x^n A_n^1, \\ A^2(x) = x^1 A_1^2 + x^2 A_2^2 + \dots + x^n A_n^2, \\ \dots \\ A^n(x) = x^1 A_1^n + x^2 A_2^n + \dots + x^n A_n^n. \end{cases}$$

The array formed by the sets A_j^i is called a **multi-valued matrix**; it is written

$$\mathbf{A} = \begin{pmatrix} A_1^1 & A_2^1 & \dots & A_n^1 \\ A_1^2 & A_2^2 & \dots & A_n^2 \\ \dots & \dots & \dots & \dots \\ A_1^n & A_2^n & \dots & A_n^n \end{pmatrix}.$$

§ 4. Bistochastic matrices

We now study an important family of matrices of order n, which plays a significant role in the theory of convexity. A matrix P of order n is said to be **bistochastic** or **doubly stochastic** if its coefficients p_j^i satisfy

$$(\forall i)\,(\forall j) : p_j^i \geqq 0,$$
$$(\forall i) : \sum_{j=1}^{n} p_j^i = 1,$$
$$(\forall j) : \sum_{i=1}^{n} p_j^i = 1.$$

If $e = (1, 1, \ldots, 1)$ denotes the point of $\mathbf{R}^n$ whose coordinates are all equal to unity, we have

$$\langle e, p^i \rangle = \langle e, p_j \rangle = 1.$$

EXAMPLE. Consider the bistochastic matrix

$$\mathsf{P} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

This matrix has only one coefficient in each row equal to 1 and only one coefficient in each column equal to 1; the image of the point $x = (x^1, x^2, x^3, x^4)$ is the point

$$\mathsf{P}x = (x^3, x^1, x^4, x^2)$$

Such a matrix is called a **permutation matrix.**

The product of two permutation matrices is again a permutation matrix; for example

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

A permutation matrix P has determinant equal to either $+1$ or -1. Its inverse is the transposed matrix P^*; for example

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

(Thus permutation matrices form a group).

Theorem 1. *The product of two bistochastic matrices is also a bistochastic matrix.*

Proof. Let $\mathbf{P}$ and $\mathbf{Q}$ be bistochastic matrices; if $\mathbf{C} = \mathbf{PQ}$, we have

$$(1)\quad c_j^i = \langle p^i, q_j\rangle \geqq 0,$$

$$(2)\quad \sum_{j=1}^{n} c_j^i = \sum_{j=1}^{n} \langle p^i, q_j\rangle = \langle p^i, \sum_{j=1}^{n} q_j\rangle = \langle p^i, e\rangle = 1,$$

$$(3)\quad \sum_{i=1}^{n} c_j^i = \sum_{i=1}^{n} \langle p^i, q_j\rangle = \langle \sum_{i=1}^{n} p^i, q_j\rangle = \langle e, q_j\rangle = 1.$$

Therefore $\mathbf{C}$ is a bistochastic matrix.

LEMMA. *A row-vector p^k of a bistochastic matrix $\mathbf{P}$ has norm $\| p^k \| \leqq 1$; equality occurs if and only if p^k is a unit vector δ_j.*

Proof. Since $0 \leqq p_j^i \leqq 1$, we have $(p_j^i)^2 \leqq p_j^i$, whence

$$\| p^k \|^2 = (p_1^k)^2+(p_2^k)^2+ \dots +(p_n^k)^2 \leqq p_1^k+p_2^k+ \dots +p_n^k = 1.$$

If p^k is not a unit vector, then $(p_i^k)^2 < p_i^k$ for at least one index i and so the inequality is satisfied strictly. On the other hand, if p^k is a unit vector, then clearly $\| p^k \| = 1$.

Theorem 2. *If $\mathbf{P}$ is a bistochastic matrix having an inverse $\mathbf{Q}$ which is also bistochastic, then $\mathbf{P}$ is a permutation matrix.*

Proof. Write

$$\mathbf{Q}\delta_1 = (x^1, x^2, \dots, x^n) = x.$$

Since $\delta_1 = \mathbf{P}x$, we have

$$1 = p_1^1 x^1+p_2^1 x^2+ \dots +p_n^1 x^n = \langle p^1, x\rangle.$$

Using the Cauchy-Schwartz inequality, we have

$$1 = | \langle p^1, x\rangle | \leqq \| p^1 \| \cdot \| x \| = \| p^1 \| \cdot \| q_1 \|.$$

Therefore, by the lemma just proved,

$$\| p^1 \| = \| q_1 \| = 1.$$

Similarly we have $\| p^2 \| = 1$, $\| p^3 \| = 1, \dots, \| p^n \| = 1$. Thus each row of $\mathbf{P}$ is a unit vector. Since the determinant of $\mathbf{P}$ is not zero, the matrix $\mathbf{P}$ contains an element equal to unity in each row and also in each column; therefore it is a permutation matrix.

Theorem 3. *The set P of bistochastic matrices of order n is a convex polyhedron in $\mathbf{R}^{n^2}$; its dimension is at most $(n-1)^2$.*

Proof. The set P is the intersection of the half-spaces $H_j^i = \{x \,/\, x_j^i \geqq 0\}$, the planes $E^i = \{x \,/\, \sum_{j=1}^{n} x_j^i = 1\}$ and the planes $E_j = \{x \,/\, \sum_{i=1}^{n} x_j^i = 1\}$; each plane is itself the intersection of two half-spaces. Therefore P is a truncation, and since $\| \mathbf{P} \| = \sqrt{\sum_{i,j} (p_j^i)^2} \leqq \sqrt{n}$, it is a polyhedron.

Since P is contained in $(2n-1)$ linearly independent planes of the form E^i or E_j, the dimension of P is at most $n^2-(2n-1) = (n-1)^2$.

Theorem of Birkhoff and von Neumann. *The profile $\ddot{P}$ is equal to the set P_0 of permutation matrices of order n.*

Proof. We first show that $P_0 \subset \ddot{P}$: that is, we show that if $\mathbf{Q}$ is a permutation matrix, it is an extreme point of P. Suppose therefore that

$$\mathbf{Q} = \lambda \mathbf{P} + \lambda' \mathbf{P}'; \quad \mathbf{P}, \mathbf{P}' \in P; \quad \mathbf{Q} \in P_0, \quad \text{where } \lambda, \lambda' > 0 \text{ and } \lambda + \lambda' = 1.$$

Since $p_j^i \neq 0$ implies that $q_j^i \neq 0$, the only non-zero p_j^i are those whose position corresponds to that occupied by an element of $\mathbf{Q}$ equal to unity; therefore, since $\mathbf{P}$ is bistochastic, we have $\mathbf{P} = \mathbf{Q}$. Similarly, $\mathbf{P}' = \mathbf{Q}$ and therefore the matrix $\mathbf{Q}$ is an extreme point of P.

We now prove that $\ddot{P} \subset P_0$, using an inductive argument. The inclusion clearly holds for matrices of order 1; we shall prove that if it holds for matrices of order $n-1$, then it holds for matrices of order n. In this proof, it is convenient to use the notation introduced in § 3 (page 177) to indicate the number of rows and columns of the matrices involved.

If $\mathbf{P}_{\mathrm{N}}^{\mathrm{N}} \in \ddot{P}$, there exists an index i_0 such that the row-vector p^{i_0} is a unit vector δ_{j_0}. For otherwise there would be at least $2n$ non-zero coefficients p_j^i; but the polyhedron P, in a linear variety of $(n-1)^2$ dimensions, is defined by generating planes of equations $x_j^i = 0$ and so each of its vertices belongs to at least $(n-1)^2$ of these planes, whence there are at least $(n-1)^2$ zero coefficients p_j^i. Thus we have a contradiction, since

$$2\,n + (n-1)^2 > n^2.$$

We therefore conclude that

$$\mathbf{P}_{\mathrm{N}}^{\mathrm{N}} = \begin{array}{c} \begin{array}{ccccccc} & & & j_0 & & & \end{array} \\ \begin{pmatrix} p_1^1 & p_2^1 & \cdots & 0 & \cdots & p_n^1 \\ p_1^2 & p_2^2 & \cdots & 0 & \cdots & p_n^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ p_1^n & p_2^n & \cdots & 0 & \cdots & p_n^n \end{pmatrix} \end{array} \begin{array}{c} \\ \\ \\ \\ i_0 \\ \\ \\ \end{array}$$

Put $\{i_0\} = \mathrm{I}$, $\{j_0\} = \mathrm{J}$. The matrix $\mathbf{P}_{\mathrm{N-J}}^{\mathrm{N-I}}$ (obtained from $\mathbf{P}_{\mathrm{N}}^{\mathrm{N}}$ by suppressing the column of index j_0 and the row of index i_0) is then a bistochastic matrix of order $n-1$. Moreover, if $\lambda, \lambda' > 0$ and $\lambda+\lambda' = 1$ and if $e_j^i = 1$ for all i and all j, we have

$$\begin{aligned}
\mathbf{P}_{\mathrm{N-J}}^{\mathrm{N-I}} = \lambda \mathbf{Q}_{\mathrm{N-J}}^{\mathrm{N-I}} + \lambda' \mathbf{Q}'^{\mathrm{N-I}}_{\mathrm{N-J}} &\Rightarrow \\
\mathbf{P}_{\mathrm{N}}^{\mathrm{N}} = \lambda(\mathbf{Q}_{\mathrm{N-J}}^{\mathrm{N-I}} \otimes \mathbf{E}_{\mathrm{J}}^{\mathrm{I}}) + \lambda'(\mathbf{Q}'^{\mathrm{N-I}}_{\mathrm{N-J}} \otimes \mathbf{E}_{\mathrm{J}}^{\mathrm{I}}) &\Rightarrow \\
\mathbf{Q}_{\mathrm{N-J}}^{\mathrm{N-I}} \otimes \mathbf{E}_{\mathrm{J}}^{\mathrm{I}} = \mathbf{Q}'^{\mathrm{N-I}}_{\mathrm{N-J}} \otimes \mathbf{E}_{\mathrm{J}}^{\mathrm{I}} &\Rightarrow \\
\mathbf{Q}_{\mathrm{N-J}}^{\mathrm{N-I}} = \mathbf{Q}'^{\mathrm{N-I}}_{\mathrm{N-J}}.&
\end{aligned}$$

Therefore $\mathbf{P}_{\mathrm{N-J}}^{\mathrm{N-I}}$ is an extreme point of the set of bistochastic matrices of order $n-1$ and so it is a permutation matrix; therefore $\mathbf{P}_{\mathrm{N}}^{\mathrm{N}}$ is a permutation matrix.

COROLLARY. *If* $\mathbf{P}$ *is a bistochastic matrix, we can write*

$$\mathbf{P} = \lambda_1 \mathbf{Q}_1 + \lambda_2 \mathbf{Q}_2 + \ldots + \lambda_m \mathbf{Q}_m,$$

where

$$(\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathbf{P}_m; \quad m \leqq (n-1)^2+1;$$
$$\mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_m \in P_0.$$

Proof. By the corollary to Theorem 1, § 2, a convex polyhedron is equal to the convex closure of its profile $\ddot{C}$. Choose C to be the set P of bistochastic matrices of order n; then, since P has dimension at most $(n-1)^2$, the result follows from Theorem 4 of § 4, Chapter VII.

LEMMA. *Let* $x = (x^1, x^2, \ldots, x^n) \in \mathbf{R}^n$ *and suppose that*

$$x^1 \leqq x^2 \leqq \ldots \leqq x^i \leqq \ldots \leqq x^j \leqq \ldots \leqq x^n;$$

suppose also that $\varepsilon > 0$. *If* $y = (y^1, y^2, \ldots, y^n)$ *is a vector of* $\mathbf{R}^n$ *such that*

$$\begin{aligned}
y^i &= x^i + \varepsilon \leqq x^j, \\
y^j &= x^j - \varepsilon \geqq x^i, \\
y^k &= x^k \text{ if } k \neq i, j,
\end{aligned}$$

there exists a bistochastic matrix $\mathbf{P}$ *such that* $y = \mathbf{P}x$.

Proof. Since $x^j - x^i \geqq \varepsilon > 0$, we can define a number λ by

$$\lambda(x^j - x^i) = +\varepsilon,$$

so that

$$\begin{aligned}
y^i &= x^i + \varepsilon = \lambda x^j + (1-\lambda)x^i, \\
y^j &= x^j - \varepsilon = \lambda x^i + (1-\lambda)x^j.
\end{aligned}$$

Also

$$\lambda = \frac{\varepsilon}{x^j - x^i} > 0; \quad \lambda = \frac{\varepsilon}{x^j - x^i} \leqq \frac{x^j - x^i}{x^j - x^i} = 1.$$

Then the matrix $\mathbf{P}$ such that

$$p_i^i = p_j^j = 1-\lambda, \qquad p_j^i = p_i^j = \lambda,$$
$$p_l^k = \delta_l^k \qquad (k, l \neq i, j)$$

is bistochastic, and $y = \mathbf{P}x$.

Theorem of Hardy, Littlewood and Polya. *If $x = (x^1, x^2, \ldots, x^n)$ and $y = (y^1, y^2, \ldots, y^n)$ are two vectors of $\mathbf{R}^n$ such that*

$$x^1 \leqq x^2 \leqq \ldots \leqq x^n,$$
$$y^1 \leqq y^2 \leqq \ldots \leqq y^n,$$

then the following three conditions are equivalent:

(1) *there exists a bistochastic matrix $\mathbf{P}$ such that $y = \mathbf{P}x$,*
(2) *for any convex function ϕ defined in $\mathbf{R}$, we have*

$$\phi(x^1)+\phi(x^2)+\ldots+\phi(x^n) \geqq \phi(y^1)+\phi(y^2)+\ldots+\phi(y^n),$$

(3) *we have*

$$x^1+x^2+\ldots+x^k \leqq y^1+y^2+\ldots+y^k \quad (\textit{if } k \leqq n),$$
$$x^1+x^2+\ldots+x^n = y^1+y^2+\ldots+y^n.$$

Proof. (1) implies (2). We say that a numerical function ϕ defined in $\mathbf{R}$ is convex (see § 5) if, for all $t_1, t_2, \ldots, t_m \in \mathbf{R}$, $(p_1, p_2, \ldots, p_m) \in \mathbf{P}_m$, and all m, we have

$$\phi(\sum_{i=1}^{m} p_i t_i) \leqq \sum_{i=1}^{m} p_i\, \phi(t_i).$$

If $y = \mathbf{P}x$, then

$$\begin{cases} y^1 = p_1^1x^1+p_2^1x^2+\ldots+p_n^1x^n, \\ y^2 = p_1^2x^1+p_2^2x^2+\ldots+p_n^2x^n, \\ \ldots\ldots\ldots\ldots \end{cases}$$

and so

$$\begin{cases} \phi(y^1) \leqq p_1^1\, \phi(x^1)+p_2^1\, \phi(x^2)+\ldots+p_n^1\, \phi(x^n), \\ \phi(y^2) \leqq p_1^2\, \phi(x^1)+p_2^2\, \phi(x^2)+\ldots+p_n^2\, \phi(x^n), \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{cases}$$

Adding these together we get

$$\phi(y^1)+\phi(y^2)+\ldots+\phi(y^n) \leqq \phi(x^1)+\phi(x^2)+\ldots+\phi(x^n).$$

(2) implies (3). For each $k \leqq n$, let ϕ_k be the numerical function defined by

$$\phi_k(t) = x^k-t \quad \text{if} \quad t \leqq x^k,$$
$$\phi_k(t) = 0 \quad \text{if} \quad t > x^k.$$

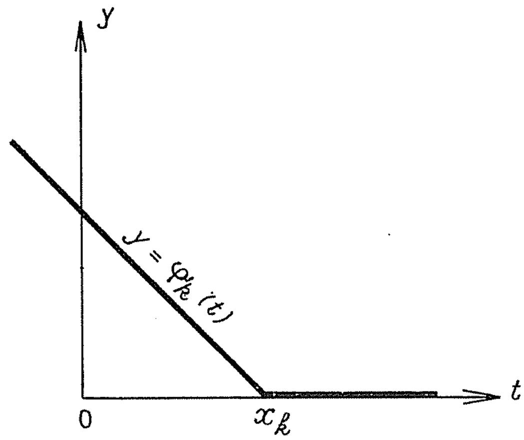


FIG. 40

As we can see from figure 40, ϕ_k is a positive convex function and so, by (2),

$$\sum_{i=1}^{k} (x^k - x^i) = \sum_{i=1}^{n} \phi_k(x^i) \geqq \sum_{i=1}^{n} \phi_k(y^i) \geqq \sum_{i=1}^{k} (x^k - y^i) + 0.$$

Hence

$$\sum_{i=1}^{k} x^i \leqq \sum_{i=1}^{k} y^i$$

and so we have the inequalities stated in (3). If $k = n$, we have

$$x^1 + x^2 + \ldots + x^n \leqq y^1 + y^2 + \ldots + y^n.$$

On the other hand, putting $\phi(t) = t$ in (2), we get

$$x^1 + x^2 + \ldots + x^n \geqq y^1 + y^2 + \ldots + y^n$$

and therefore

$$x^1 + x^2 + \ldots + x^n = y^1 + y^2 + \ldots + y^n.$$

(3) implies (1). To prove this, we use induction. If $n = 1$, the result is clearly true; we shall suppose that it holds for all integers strictly less than n and shall show that this implies that it holds for n.

Put

$$a = \frac{x^1 + x^2 + \ldots + x^n}{n} = \frac{y^1 + y^2 + \ldots + y^n}{n}.$$

We can assume that $x^1 < a$ and $x^n > a$, for otherwise we have

$$x^1 = x^2 = \ldots = x^n = a,$$

so that

$$y^n \geqq y^{n-1} \geqq \ldots \geqq y^1 \geqq x^1 \geqq a,$$

and therefore $y^1 = y^2 = \dots = y^n = a$; condition (1) then clearly holds, for the unit matrix $\mathbf{E}$ is bistochastic and $y = \mathbf{E}x$.

We consider two cases:

Case 1. There exists an integer k, strictly less than n, such that

$$\sum_{h=1}^{k} x^h = \sum_{h=1}^{k} y^h.$$

Then we have

$$\begin{aligned}
&x^1 \leqq y^1,\\
&x^1+x^2 \leqq y^1+y^2,\\
&\dots\dots\dots\\
&x^1+x^2+\dots+x^k = y^1+y^2+\dots+y^k,\\
&x^{k+1} \leqq y^{k+1},\\
&x^{k+1}+x^{k+2} \leqq y^{k+1}+y^{k+2},\\
&\dots\dots\dots\dots\\
&x^{k+1}+x^{k+2}+\dots+x^n = y^{k+1}+y^{k+2}+\dots+y^n.
\end{aligned}$$

Put $\mathbf{K} = \{1, 2, \dots, k\}$; by hypothesis there exist two bistochastic matrices $\mathbf{P}_{\mathrm{K}}^{\mathrm{K}}$ and $\mathbf{Q}_{\mathrm{N-K}}^{\mathrm{N-K}}$ such that

$$\mathbf{Y}^{\mathrm{K}} = \mathbf{P}_{\mathrm{K}}^{\mathrm{K}}\mathbf{X}^{\mathrm{K}}, \qquad \mathbf{Y}^{\mathrm{N-K}} = \mathbf{Q}_{\mathrm{N-K}}^{\mathrm{N-K}}\mathbf{X}^{\mathrm{N-K}}.$$

Then the matrix $\mathbf{P}_{\mathrm{K}}^{\mathrm{K}} \otimes \mathbf{Q}_{\mathrm{N-K}}^{\mathrm{N-K}}$ is bistochastic and we have

$$y = (\mathbf{P}_{\mathrm{K}}^{\mathrm{K}} \otimes \mathbf{Q}_{\mathrm{N-K}}^{\mathrm{N-K}})x.$$

Thus (1) is established.

Case (2). For each integer k strictly less than n, we have

$$\sum_{h=1}^{k} x^h < \sum_{h=1}^{k} y^h.$$

Let i be the greatest index such that $x^i < a$ and j the smallest index such that $x^j > a$. By the lemma, there exists a bistochastic transformation

$$x_0 = \mathbf{P}_\varepsilon x$$

such that

$$\begin{cases} x_0^i = x^i+\varepsilon \leqq a,\\ x_0^j = x^j-\varepsilon \geqq a,\\ x_0^k = x^k \quad (\text{if } k \neq i, j). \end{cases}$$

If this transformation leads us to a situation in which the condition of case (1) is satisfied, the proposition is proved, for

$$y = \mathbf{P}x_0 = \mathbf{P}\mathbf{P}_\varepsilon x.$$

We therefore suppose that, for all ε, we have

$$\begin{cases} x_0^1 < y^1, \\ x_0^1+x_0^2 < y^1+y^2, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \vdots \\ x_0^1+x_0^2+\ldots+x_0^{n-1} < y^1+y^2+\ldots+y^{n-1}, \\ x_0^1+x_0^2+\ldots+x_0^n = y^1+y^2+\ldots+y^n. \end{cases}$$

Putting $\varepsilon = \min\{(a-x^i), (x^j-a)\}$, the transformation $\mathbf{P}_\varepsilon$ enables us to increase the number of components of x equal to a. We can then repeat the process: by a new bistochastic transformation we can again increase the number of components of x equal to a, or else verify (1) by reverting to case (1). If sooner or later we do not revert to case (1), we finally obtain a vector

$$x_1 = (x_1^1, x_1^2, \ldots, x_1^n)$$

where

$$\begin{cases} x_1^1 = x_1^2 = \ldots = x_1^n = a, \\ x_1^1 < y^1, \\ x_1^1+x_1^2 < y^1+y^2, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ x_1^1+x_1^2+\ldots+x_1^n = y^1+y^2+\ldots+y^n. \end{cases}$$

But this implies that

$$a = \frac{y^1+y^2+\ldots+y^n}{n} \geqq y^1 > x_1^1 = a$$

and so this situation cannot arise.

COROLLARY. *Let* $x = (x^1, x^2, \ldots, x^n)$ *and* $y = (y^1, y^2, \ldots, y^n)$ *be such that* $x^1 \leqq x^2 \leqq \ldots \leqq x^n$, $y^1 \leqq y^2 \leqq \ldots \leqq y^n$, *and put* $z \leqq y$ *whenever* $z^i \leqq y^i$ *for all i. There exists a bistochastic matrix* $\mathbf{P}$ *such that* $y \geqq \mathbf{P}x$ *if and only if, for each integer* $k \leqq n$, *we have*

$$\sum_{i=1}^{k} x^i \leqq \sum_{i=1}^{k} y^i.$$

Proof. If $z = \mathbf{P}x$ and $z \leqq y$, we have

$$x^1+x^2+\ldots+x^k \leqq z^1+z^2+\ldots+z^k \leqq y^1+y^2+\ldots+y^k$$

for all k.

Conversely, suppose that these inequalities are satisfied for each integer $k \leqq n$; we must prove that there exists a vector z such that $z = \mathbf{P}x$ and $z \leqq y$.

By means of a transformation $y^1 \to z^1$, we can reduce the value of y^1 to obtain

$$\begin{cases} x^1+x^2+\ldots+x^k \leqq z^1+y^2+\ldots+y^k & (\text{if } k < m), \\ x^1+x^2+\ldots+x^m = z^1+y^2+\ldots+y^m, & \\ x^1+x^2+\ldots+x^k < z^1+y^2+\ldots+y^k & (\text{if } k > m). \end{cases}$$

If $m = n$, the result follows at once from the theorem. If $m < n$, we have $x^{m+1} < y^{m+1}$; by a transformation $y^{m+1} \to z^{m+1} = x^{m+1}$, which reduces the value of y^{m+1}, we can replace the $(m+1)$th inequality by an equality. Moreover, we always have

$$z^1 \leqq y^2 \leqq \ldots \leqq y^m \leqq z^{m+1} \leqq y^{m+2} \leqq \ldots \leqq y^n,$$

since

$$z^{m+1} < y^m \;\Rightarrow\; x^m \leqq x^{m+1} = z^{m+1} < y^m \;\Rightarrow\; x^1+x^2+\ldots+x^{m-1} > y^1+y^2+\ldots+y^{m-1}.$$

If necessary we repeat this operation until we obtain

$$x^1+x^2+\ldots+x^n = z^1+y^2+\ldots+y^m+z^{m+1}+\ldots+z^n.$$

Then $z = (z^1, \ldots, z^n)$ is the required vector.

Remark. In the statement of the theorem of Hardy, Littlewood and Polya, conditions (1) and (2) do not involve the order in which we have arranged the components of x or of y. Condition (3) can also be expressed in a form which does not depend on this order. If we denote the family of subsets $\mathbf{I}$ of $\mathbf{N} = \{1, 2, \ldots, n\}$ having k elements by $\mathscr{T}_k$, then (3) can be replaced by

$$(3') \quad \begin{cases} \max\limits_{\mathbf{I}\in\mathscr{T}_k} \sum\limits_{i\in\mathbf{I}} x^i \leqq \max\limits_{\mathbf{J}\in\mathscr{T}_k} \sum\limits_{j\in\mathbf{J}} y^j, \\ \sum\limits_{i\in\mathbf{N}} x^i = \sum\limits_{j\in\mathbf{N}} y^i. \end{cases}$$

The corollary can then be stated as follows:

If $x, y \in \mathbf{R}^n$, a necessary and sufficient condition for there to exist a bistochastic matrix $\mathbf{P}$ such that $y \geqq \mathbf{P}x$ is that, for each integer $k \geqq n$, we have

$$\max_{\mathbf{I}\in\mathscr{T}_k} \sum_{i\in\mathbf{I}} x^i \leqq \max_{\mathbf{J}\in\mathscr{T}_k} \sum_{j\in\mathbf{J}} y^j.$$

§ 5. Convex functions[1]

Let C be a convex set in $\mathbf{R}^n$. We say that a numerical function f is **convex** in C if

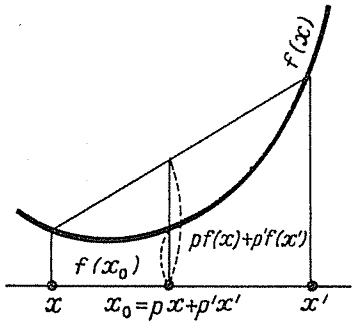
$$f(px+p'x') \leqq pf(x)+p'f(x')$$

[1] For the remainder of this chapter, we are concerned mainly with convex functions and their generalisations.

for all $x, x' \in C$ and $(p, p') \in \mathbf{P}_2$. We say that f is **concave** if

$$f(px+p'x') \geqq pf(x)+p'f(x').$$

Thus f is concave if and only if $-f$ is convex. If the above inequalities are strict for $x \neq x'$ and $p, p' \neq 0$, we say that f is **strictly convex**, or **strictly concave**, as the case may be.



Example of a convex function in R
FIG. 41

EXAMPLE 1. A numerical function g such that $g(x) = \langle a, x\rangle + \alpha$, where $a \in \mathbf{R}^n$, $\alpha \in \mathbf{R}$, is called a **linear affine function**; it is a convex function in $\mathbf{R}^n$

EXAMPLE 2. The gauge j of a convex set $C \subset \mathbf{R}^n$, admitting 0 as an internal point, is a convex function in $\mathbf{R}^n$; in fact, if $x, x' \in \mathbf{R}^n$,

$$j(px+p'x') \leqq j(px)+j(p'x') = pj(x)+p'j(x').$$

If $C \ni 0$ and if 0 is not an internal point of C, the function j can take the value $+\infty$; in this case, j is an example of a *generalised convex function.*

EXAMPLE 3. If $r(x) = d(x, C)$ is the distance from a point x to a convex set $C \subset \mathbf{R}^n$, then r is a convex function in $\mathbf{R}^n$. For suppose that $x, x' \in \mathbf{R}^n$ and that $\varepsilon > 0$; then there exist points c and c' of C such that

$$\| x-c \| < r(x)+\varepsilon, \qquad \| x'-c' \| < r(x')+\varepsilon.$$

Therefore

$$\begin{aligned} r(px+p'x') &\leqq \| px+p'x'-(pc+p'c') \| \leqq \| p(x-c) \| + \| p'(x'-c') \| \\ &\leqq p[r(x)+\varepsilon]+p'[r(x')+\varepsilon]. \end{aligned}$$

Since this holds for all ε, we have

$$r(px+p'x') \leqq pr(x)+p'r(x').$$

Theorem 1. *In order that a function f should be convex in C it is necessary and sufficient that*

$$\left\{ \begin{array}{l} (p_1, p_2, \ldots, p_m) \in \mathbf{P}_m \\ x_1, x_2, \ldots, x_m \in C \end{array} \right. \Rightarrow f\Big(\sum_{i=1}^m p_i x_i\Big) \leqq \sum_{i=1}^m p_i f(x_i)$$

for all integers m.

Proof. If the condition is satisfied, then clearly f is convex: we simply consider the case $m = 2$.

Suppose conversely that f is a convex function. Clearly the condition is satisfied for $m = 1$ and $m = 2$. Suppose that it is satisfied for $m = r$; then, if $p_{r+1} \neq 1$,

$$\sum_{i=1}^{r+1} p_i x_i = \sum_{i=1}^{r} p_i x_i + p_{r+1} x_{r+1} = \left(\sum_{i=1}^{r} p_i\right) a + p_{r+1} x_{r+1},$$

where

$$a = \sum_{i=1}^{r} \frac{p_i}{\sum_{i=1}^{r} p_i} x_i \in C.$$

Therefore

$$f\left(\sum_1^{r+1} p_i x_i\right) \leqq \left(\sum_1^{r} p_i\right) f(a) + p_{r+1} f(x_{r+1}) \leqq 1\left[\sum_1^{r} p_i f(x_i)\right] + p_{r+1} f(x_{r+1}).$$

It follows that the condition is satisfied for $m = r+1$; hence, by induction, it holds for all m.

Theorem 2. *In order that a function f should be convex in C it is necessary and sufficient that, for each pair of points c and c' in C, the function ϕ defined by*

$$\phi(\lambda) = f(\lambda c + (1-\lambda)c')$$

is convex in the segment $[0, 1]$.

Proof. If ϕ is a convex function in $[0, 1]$ and if $(p, p') \in \mathbf{P}_2$, we have

$$\begin{aligned} f(pc+p'c') = \phi(p) &= \phi(p \times 1 + p' \times 0) \\ &\leqq p\phi(1) + p'\phi(0) = pf(c) + p'f(c'). \end{aligned}$$

This holds for all c and c' and therefore the function f is convex in C.

Conversely, suppose that f is a convex function in C. Given $\lambda_1, \lambda_2 \in [0, 1]$, we write

$$\begin{aligned} c_1 &= \lambda_1 c + (1-\lambda_1)\, c' \in [c, c'], \\ c_2 &= \lambda_2 c + (1-\lambda_2)\, c' \in [c, c']. \end{aligned}$$

Then, if $(p_1, p_2) \in \mathbf{P}_2$, we have

$$\begin{aligned} p_1 c_1 + p_2 c_2 &= p_1\lambda_1 c + p_1 c' - p_1\lambda_1 c' + p_2\lambda_2 c + p_2 c' - p_2\lambda_2 c' \\ &= (p_1\lambda_1 + p_2\lambda_2)\, c + [1 - (p_1\lambda_1 + p_2\lambda_2)]\, c' \end{aligned}$$

and so

$$\phi(p_1\lambda_1 + p_2\lambda_2) = f(p_1 c_1 + p_2 c_2) \leqq p_1 f(c_1) + p_2 f(c_2) = p_1\phi(\lambda_1) + p_2\phi(\lambda_2).$$

Therefore ϕ is convex in $[0, 1]$.

Theorem 3. *A function f is convex in $C \subset \mathbf{R}^n$ if and only if the set*

$$\overline{T}_f = \{\bar{x} = (x^1, x^2, \ldots, x^{n+1}) \,/\, x = (x^1, x^2, \ldots, x^n) \in C, f(x) \leqq x^{n+1}\}$$

in $\mathbf{R}^{n+1}$ is convex.

Proof. Suppose that the set $\overline{T}_f$ is convex. If $x \in C$, $x' \in C$, $y = f(x)$ and $y' = f(x')$, then

$$p(x^1, x^2, \ldots, x^n, y) + p'(x'^1, x'^2, \ldots, x'^n, y') = (px^1 + p'x'^1, \ldots, py + p'y') \in \overline{T}_f$$

and therefore

$$f(px + p'x') \leqq py + p'y' = pf(x) + p'f(x').$$

Hence the function f is convex.

Conversely, suppose that the function f is convex. If $(x^1, x^2, \ldots, x^n, y)$ and $(x'^1, x'^2, \ldots, x'^n, y')$ are in $\overline{T}_f$, we have

$$f(px + p'x') \leqq pf(x) + p'f(x') \leqq py + p'y'.$$

Then

$$p(x^1, x^2, \ldots, x^n, y) + p'(x'^1, x'^2, \ldots, x'^n, y') \in \overline{T}_f$$

and therefore $\overline{T}_f$ is a convex set.

A function ϕ in $\mathbf{R}^m$ is said to be **increasing** if

$$(\forall i) : z_i \leqq z'_i \Rightarrow \phi(z_1, z_2, \ldots, z_m) \leqq \phi(z'_1, z'_2, \ldots, z'_m).$$

Theorem 4. *If $f_1, f_2, \ldots, f_m$ are convex (resp. concave) functions in $C \subset \mathbf{R}^n$ and if ϕ is an increasing convex (resp. concave) function in $\mathbf{R}^m$, then the function g defined by*

$$g(x) = \phi[f_1(x), f_2(x), \ldots, f_m(x)]$$

is convex (resp. concave) in C.

Proof. If $x, x' \in C$, then, for all i,

$$f_i(px + p'x') \leqq pf_i(x) + p'f_i(x').$$

Therefore

$$\begin{aligned} g(px+p'x') &= \phi[f_1(px+p'x'), \ldots] \leqq \phi[pf_1(x) + p'f_1(x'), \ldots] \\ &\leqq p\phi[f_1(x), \ldots] + p'\phi[f_1(x'), \ldots] = pg(x) + p'g(x') \end{aligned}$$

and so g is convex. A similar proof holds for concave functions.

Theorem 5. *If $(f_i \mid i \in I)$ is a family of convex functions in C, then the function f defined by $f(x) = \sup_{i \in I} f_i(x)$ is convex in C.*

Proof. If f_i is convex, the set

$$\overline{T}_{f_i} = \{\bar{x} = (x^1, x^2, \ldots, x^{n+1}) \mid x \in C, x^{n+1} \geqq f_i(x^1, x^2, \ldots, x^n)\}$$

is convex and therefore so is the set

$$\overline{T}_f = \bigcap_{i \in I} \overline{T}_{f_i}.$$

Hence f is a convex function.

In particular, if A is a set in $\mathbf{R}^n$, the function f defined by

$$\sup_{a \in A} \langle a, x \rangle = f(x)$$

is convex.

LEMMA. *If $C \subset \mathbf{R}^n$ is a closed set and f is a lower semi-continuous function in $\mathbf{R}^n$, the set*

$$\overline{T}_f = \{\bar{x} = (x^1, x^2, \ldots, x^{n+1}) \,/\, x = (x^1, x^2, \ldots, x^n) \in C, f(x) \leqq x^{n+1}\}$$

is closed in $\mathbf{R}^{n+1}$.

Proof. Suppose that $\bar{a} = (a^1, a^2, \ldots, a^{n+1}) \notin \overline{T}_f$. We shall prove that there exists a ball of centre $\bar{a}$ not meeting $\overline{T}_f$; there are two cases which can occur.

Case 1. If $a = (a^1, a^2, \ldots, a^n) \notin C$ then, since C is closed, there exists a number ε such that

$$\| x - a \| \leqq \varepsilon \Rightarrow x \notin C.$$

Therefore

$$\| \bar{x} - \bar{a} \| \leqq \varepsilon \Rightarrow \| x - a \| \leqq \varepsilon \Rightarrow x \notin C \Rightarrow \bar{x} \notin \overline{T}_f$$

and so there exists a ball of centre $\bar{a}$ not meeting $\overline{T}_f$.

Case 2. If $f(a) > a^{n+1}$, there exists a number ε such that

$$f(a) - \varepsilon > a^{n+1} + \varepsilon.$$

Since f is lower semi-continuous, there exists a number η such that

$$\| a - x \| \leqq \eta \Rightarrow f(x) \geqq f(a) - \varepsilon.$$

Therefore, if $\| \bar{a} - \bar{x} \| \leqq \min(\eta, \varepsilon)$, then

$$f(x) \geqq f(a) - \varepsilon > a^{n+1} + \varepsilon \geqq x^{n+1}$$

and so there exists a ball of centre $\bar{a}$ not meeting $\overline{T}_f$. Thus the lemma is proved.

Theorem 6. *Let $C \subset \mathbf{R}^n$ be a closed convex set and let f be a convex lower semi-continuous function in $\mathbf{R}^n$. If $\mathscr{G} = (g_i \,/\, i \in I)$ is the family of linear affine functions such that $g_i(x) \leqq f(x)$ for all $x \in C$, then we have*

$$f(x) = \sup_{i \in I} g_i(x)$$

for all $x \in C$.

Proof. Let a be the point $(a^1, a^2, \ldots, a^n) \in C$. Suppose that there exists a positive number ε such that

$$\sup_{i \in I} g_i(a) < f(a) - \varepsilon.$$

If $a^{n+1} = f(a) - \varepsilon$, the point $\bar{a} = (a^1, a^2, \ldots, a^n, a^{n+1})$ in $\mathbf{R}^{n+1}$ does not belong to $\overline{T}_f$, for

$$f(a) > f(a) - \varepsilon = a^{n+1}.$$

By the lemma, $\overline{T}_f$ is closed and convex. Therefore, by the second separation theorem (§ 1) there exists a plane in $\mathbf{R}^{n+1}$ separating $\bar{a}$ and $\overline{T}_f$, say $x^{n+1} - g(x^1, x^2, \ldots, x^n) = 0$, where g is a linear affine function in $\mathbf{R}^n$. Then we have

$$(\forall \bar{x};\ \bar{x} \in \overline{T}_f) : a^{n+1} - g(a) \leqq 0 \leqq x^{n+1} - g(x).$$

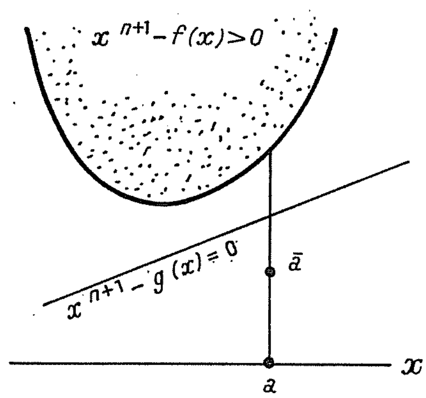


FIG. 42

In particular, if we take $x^{n+1} = f(x)$, we have $g(x) \leqq f(x)$ (for all x in C). Therefore $g \in \mathscr{G}$ and so

$$g(a) \leqq \sup_{i \in I} g_i(a) < f(a) - \varepsilon = a^{n+1}.$$

But

$$a^{n+1} - g(a) \leqq 0$$

and so we have a contradiction. Therefore there is no strictly positive number ε such that

$$\sup_{i \in I} g_i(a) < f(a) - \varepsilon$$

and the theorem is proved.

Theorem 7. *If G is a convex open set in $\mathbf{R}^n$ and f is a convex function in G, then f is continuous in G.*

Proof. We first show that if $a \in G$, there exist numbers δ and α such that

$$\| x - a \| \leqq \delta \ \Rightarrow\ f(x) \leqq \alpha.$$

In fact, there exists an n-simplex $[a_1, a_2, \ldots, a_{n+1}]$ whose vertices all belong to G and for which a is an interior point. If $B_\delta(a)$ is a ball contained in this simplex and if $\alpha = \max f(a_i)$, then, for all x in $B_\delta(a)$,

$$f(x) = f\Big(\sum_1^{n+1} p_i\, a_i\Big) \leqq \sum_1^{n+1} p_i f(a_i) \leqq \alpha.$$

Suppose now that x_0 is a point of $B_\delta(a)$; we shall prove that f is continuous at x_0. By means of a translation we can replace x_0 by 0 and, by replacing $f(x)$ by $f(x)-f(x_0)$, we can make $f(0)$ equal to 0. If $B_\lambda(0) \subset B_\delta(a)$ and ε is a strictly positive number, then for all x such that $\left\|\dfrac{x}{\varepsilon}\right\| \leqq \lambda$, we have

$$f(x) = f\left[(1-\varepsilon)\,0+\varepsilon\,\frac{x}{\varepsilon}\right] \leqq \varepsilon f\left(\frac{x}{\varepsilon}\right) \leqq \varepsilon\alpha.$$

Also

$$0 = f(0) = f\left[\frac{1}{1+\varepsilon}\,x+\frac{\varepsilon}{1+\varepsilon}\left(\frac{-x}{\varepsilon}\right)\right]$$

$$\leqq \frac{1}{1+\varepsilon} f(x)+\frac{\varepsilon}{1+\varepsilon} f\left(\frac{-x}{\varepsilon}\right) \leqq \frac{1}{1+\varepsilon} f(x)+\frac{\varepsilon\alpha}{1+\varepsilon}.$$

Therefore $f(x) \geqq -\varepsilon\alpha$ and so

$$\|\, x \,\| \leqq \lambda\varepsilon \quad \Rightarrow \quad |f(x)| \leqq \alpha\varepsilon.$$

It follows that f is continuous at x_0; since x_0 is arbitrary, f is continuous in G.

Theorem 8. *Let C be a convex set and let f be a convex function in C. If f attains its maximum on C at an interior point of C, then f is constant in C.*

Proof. Suppose that f is not constant in C. Write $\alpha = \max\limits_{y \in C} f(y)$. Then there exists a point x in C such that $f(x) < \alpha$. If $a \in \mathring{C}$, there exists a point y such that

$$y \in C; \quad a = px+qy; \quad p, q > 0; \quad p+q = 1.$$

Therefore

$$f(a) = f(px+qy) \leqq pf(x)+qf(y) < p\alpha+q\alpha = \alpha$$

and so f does not attain its maximum at a.

§ 6. Differentiable convex functions

The property of 'admitting derivatives' in the space $\mathbf{R}^n$ has been advantageously replaced, by Stolz and Fréchet, by the now classical notion of 'differentiability'.

Let f be a numerical function, with values $f(x) = f(x^1, x^2, \ldots, x^n)$, defined in an open set G of $\mathbf{R}$. We say that f is **differentiable** in G if for all $x = (x^1, x^2, \ldots, x^n)$ in G and all $\Delta x = (\Delta x^1, \Delta x^2, \ldots, \Delta x^n)$ such that $x + \Delta x \in G$, we have

$$f(x+\Delta x)-f(x) = \alpha_1(x)\,\Delta x^1+\alpha_2(x)\,\Delta x^2+\ldots+\alpha_n(x)\,\Delta x^n+\beta(x, \Delta x)\,\|\Delta x\|,$$

where the numerical functions $\alpha_1, \alpha_2, \ldots, \alpha_n$ are finite and the numerical function $\beta(x, \Delta x)$ tends to 0 whenever Δx tends to 0 (the point x remaining fixed).

We say that f admits a partial derivative with respect to x^1 if

$$\frac{f(x+h\,\delta_1)-f(x)}{h} = \frac{f(x^1+h, x^2, \ldots, x^n)-f(x^1, x^2, \ldots, x^n)}{h}$$

tends to a limit when h tends to 0; this limit, which we denote by $f_1' = \dfrac{\partial f}{\partial x^1}$, is called the **partial derivative** of f with respect to x^1.

Theorem 1. *In an open set G, a differentiable function is continuous and admits partial derivatives with respect to all the variables.*

Proof. We have

$$|f(x+\Delta x)-f(x)| \leqq \sum_{i=1}^{n} |\alpha_i| \,.\, |\Delta x^i| + |\beta(x, \Delta x)| \,.\, \|\Delta x\|.$$

Let ε be a strictly positive number and choose a number η such that

$$(1)\quad \|\Delta x\| \leqq \eta \;\Rightarrow\; |\beta(x, \Delta x)| \leqq 1,$$

$$(2)\quad \eta \leqq \frac{\varepsilon}{2},$$

$$(3)\quad \eta \leqq \frac{\varepsilon}{2\,\Sigma\,|\alpha_i|}.$$

Then

$$\|\Delta x\| \leqq \eta \;\Rightarrow\; |f(x+\Delta x)-f(x)| \leqq \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon$$

and so f is continuous at x. Since x is an arbitrary point of G, f is continuous in G.

Furthermore, f admits partial derivatives, because

$$\frac{f(x+h\delta_1)-f(x)}{h} = \alpha_1(x)+\beta(x, h\delta_1)\,\frac{|h|}{h}$$

tends to $\alpha_1(x)$ when h tends to 0; we then have

$$\alpha_1(x) = \frac{\partial f}{\partial x^1}.$$

REMARK. The converse of this proposition is not true. For example, consider the function f in $\mathbf{R}^2$ such that

$$f(x, y) = \frac{xy}{\sqrt{x^2+y^2}} \qquad \text{if } x \neq 0 \text{ or } y \neq 0,$$

$$f(x, y) = 0 \qquad \text{if } x = 0 \text{ and } y = 0.$$

This is continuous at $0 = (0, 0)$, but it is easily seen that the principal part of $f(h, k)-0$ is not linear in h and k, so that this function is not differentiable at 0.

Theorem 2. *Suppose that f is a function which admits, in an open set $G \subset \mathbf{R}^n$, continuous partial derivatives with respect to all the variables. Then f is differentiable in G.*

Proof. For simplicity, we prove the theorem for $\mathbf{R}^2$, but the same argument can be applied to the general case. We write

$$\begin{aligned}\Delta f &= f(x+h, y+k)-f(x, y)\\ &= [f(x+h, y+k)-f(x+h, y)]+[f(x+h, y)-f(x, y)].\end{aligned}$$

If $(x, y) \in G$ and h, k are sufficiently small, then, by the Mean Value Theorem, we have

$$\begin{aligned}\Delta f &= kf'_y(x+h, y+\theta_1 k)+hf'_x(x+\theta_2 h, y) = k[f'_y+\varepsilon_1]+h[f'_x+\varepsilon_2]\\ &= h\frac{\partial f}{\partial x}+k\frac{\partial f}{\partial y}+\left[\frac{k\varepsilon_1+h\varepsilon_2}{\| \Delta x \|}\right] \| \Delta x \|,\end{aligned}$$

where ε_1 and ε_2 are functions which tend to 0 with $\| \Delta x \|$. The expression between the brackets in the last term tends to 0, for

$$\left|\frac{k\varepsilon_1+h\varepsilon_2}{\| \Delta x \|}\right| \leqq \frac{| k |}{\| \Delta x \|} | \varepsilon_1 | + \frac{| h |}{\| \Delta x \|} | \varepsilon_2 | \leqq | \varepsilon_1 | + | \varepsilon_2 |$$

and therefore f is differentiable.

Theorem 3. *Let $f_1, f_2, \ldots, f_m$ be differentiable functions in $G \subset \mathbf{R}$ and let ϕ be a differentiable function in $\mathbf{R}^m$. Then the function g determined by*

$$g(x) = \phi[f_1(x), f_2(x), \ldots, f_m(x)]$$

is differentiable in G.

Proof. Corresponding to the change Δx in x we have

$$\Delta z_k = f_k(x+\Delta x)-f_k(x) = \sum_{i=1}^{n} \frac{\partial f_k}{\partial x^i} \Delta x^i+\beta_k(x, \Delta x) \| \Delta x \|.$$

The corresponding change in g is

$$\Delta g = \sum_{k=1}^{m} \frac{\partial \phi}{\partial z_k} \Delta z_k + \gamma(z, \Delta z) \| \Delta z \| = \sum_{i=1}^{n} \Delta x^i \sum_{k=1}^{m} \frac{\partial \phi}{\partial z_k} \frac{\partial f_k}{\partial x^i} +$$
$$+ \| \Delta x \| \left[\sum_{k=1}^{m} \frac{\partial \phi}{\partial z_k} \beta_k(x, \Delta x) + \gamma(z, \Delta z) \frac{\| \Delta z \|}{\| \Delta x \|} \right].$$

Clearly the expression in the last term between the brackets tends to 0 with $\| \Delta x \|$ and so the function g is differentiable.

We note that we have the following well-known formula:

$$\boxed{\frac{\partial g}{\partial x^k} = \sum_{i=1}^{m} \frac{\partial \phi}{\partial z_i} \frac{\partial f_i}{\partial x^k}}.$$

REMARK. Let f be a function defined in an open set G of $\mathbf{R}^n$. Then f is said to be **bi-differentiable** in G if, whenever x and $x+\Delta x$ belong to G, we have

$$f(x+\Delta x) - f(x) = \sum_{k=1}^{n} \alpha_k \Delta x^k + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_j^i \Delta x^i \Delta x^j + \beta(x, \Delta x) \| \Delta x \|^2,$$

where the α_k and the α_j^i are finite functions of x and the function $\beta(x, \Delta x)$ tends to 0 with $\| \Delta x \|$, the point x remaining fixed. We then write

$$\alpha_k = \frac{\partial f}{\partial x^k} = f_k'; \qquad \alpha_j^i = \frac{\partial^2 f}{\partial x^i \partial x^j} = f_{ij}''.$$

Bi-differentiable functions possess the following properties:

(1) *If the partial derivatives* $\dfrac{\partial f}{\partial x^i}$ *are differentiable in an open set G, then f is bi-differentiable; in particular, this occurs if, for all i,* $\dfrac{\partial f}{\partial x^i}$ *admits continuous partial derivatives in G.*

(2) *If f is bi-differentiable in an open set G, we have*

$$\frac{\partial}{\partial x^i}\left(\frac{\partial f}{\partial x^j}\right) = \frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial^2 f}{\partial x^j \partial x^i} = \frac{\partial}{\partial x^j}\left(\frac{\partial f}{\partial x^i}\right).$$

(3) *If* $f_1, f_2, \ldots, f_m$ *are bi-differentiable functions in an open set G and if* ϕ *is bi-differentiable in* $\mathbf{R}^m$, *the function g defined by*

$$g(x) = \phi[f_1(x), f_2(x), \ldots, f_m(x)]$$

is bi-differentiable in G.

(These results are no longer true if we remove the restriction that G is open.)

Fundamental Theorem. *If a function ϕ in an interval $D \subset \mathbf{R}$ possesses a derivative ϕ', it is convex in D if and only if ϕ' is increasing in D.*

Proof. Suppose first that ϕ' is increasing in D. If t_1 and t_2 are real numbers such that $t_1 < t_2$ and if $(p_1, p_2) \in \mathbf{P}_2$, then, by the Mean Value Theorem,

$$\begin{cases} \phi(p_1t_1+p_2t_2)-\phi(t_1) = (p_1t_1+p_2t_2-t_1)\mu_1 = p_2(t_2-t_1)\mu_1 \\ \phi'(t_1) \leqq \mu_1 \leqq \phi'(p_1t_1+p_2t_2). \end{cases}$$

Similarly, we have

$$\begin{cases} \phi(t_2)-\phi(p_1t_1+p_2t_2) = (t_2-p_1t_1-p_2t_2)\mu_2 = p_1(t_2-t_1)\mu_2 \\ \phi'(p_1t_1+p_2t_2) \leqq \mu_2 \leqq \phi'(t_2). \end{cases}$$

Since ϕ' is increasing and $t_1 \leqq p_1t_1+p_2t_2 \leqq t_2$, we have $\mu_1 \leqq \mu_2$. Therefore

$$p_1p_2(t_2-t_1)\mu_1 \leqq p_1p_2(t_2-t_1)\mu_2,$$

or

$$p_1[\phi(p_1t_1+p_2t_2)-\phi(t_1)] \leqq p_2[\phi(t_2)-\phi(p_1t_1+p_2t_2)].$$

Therefore

$$\phi(p_1t_1+p_2t_2) \leqq p_1\phi(t_1)+p_2\phi(t_2)$$

and so ϕ is convex.

Conversely, suppose that ϕ is a convex function admitting a derivative in D. Let $t_1, t_2, x_1, x_2 \in D$ be such that $t_1 < x_1 \leqq x_2 < t_2$ with $x_2 = x_1$ initially. Since the point $m(x_1) = (x_1, \phi(x_1))$ is below the segment $[m(t_1), m(t_2)]$, we have

$$\frac{\phi(x_1)-\phi(t_1)}{x_1-t_1} = \text{slope}\,[m(x_1), m(t_1)] \leqq \text{slope}\,[m(x_2), m(t_2)] = \frac{\phi(x_2)-\phi(t_2)}{x_2-t_2}.$$

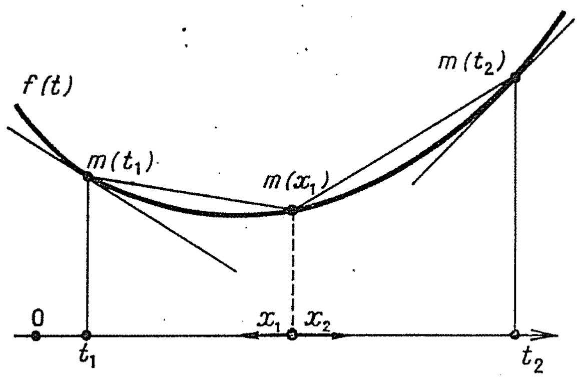


FIG. 43

Letting x_1 tend to t_1, we get

$$\phi'(t_1) \leqq \frac{\phi(x_2)-\phi(t_2)}{x_2-t_2}$$

and then, letting x_2 tend to t_2, we get

$$\phi'(t_1) \leqq \phi'(t_2)$$

Therefore ϕ' is increasing.

Theorem 4. *Let D be an open interval in the space* $\mathbf{R}$ *and let ϕ be a function admitting a second derivative ϕ'' in D. Then*

(1) *ϕ is convex if and only if $\phi''(t) \geqq 0$ for all t in D,*

(2) *ϕ is strictly convex if and only if $\phi''(t) \geqq 0$ for all t in D but is not identically zero in any non-trivial sub-interval of D.*

Proof. The first part of the theorem follows immediately from the fundamental theorem.

If ϕ is strictly convex, then ϕ'' cannot be identically zero in an interval $D' \subset D$, because then ϕ would be linear in D'.

If ϕ is not strictly convex, there exist points t_1 and t_2 (where $t_1 \neq t_2$) such that the negative function g given by

$$g(\lambda) = \phi[(1-\lambda)t_1+\lambda t_2]-(1-\lambda)\,\phi(t_1)-\lambda\,\phi(t_2)$$

is zero at a point λ_0 of $]0, 1[$; since g is a convex function which attains its maximum at an interior point λ_0, then, by Theorem 8 of § 5, $g(\lambda)$ is constant in $]0, 1[$; in fact $g(\lambda) = 0$ in $]0, 1[$. Therefore ϕ is linear in $[t_1, t_2]$ and so $\phi''(t) = 0$ for all t in $[t_1, t_2]$.

COROLLARY 1. *If a function f is differentiable in an open convex set $C \subset \mathbf{R}^n$, then f is convex if and only if for all c and c' in C, the function g given by*

$$g(t) = \sum_{i=1}^{n} (c^i - c'^i) f_i'(tc+(1-t)\,c')$$

increases as t increases from 0 *to* 1.

Proof. The function f is convex in C if and only if the function ϕ defined by

$$\phi(t) = f(tc+(1-t)c')$$

is convex in $[0, 1]$, by Theorem 2, § 5. By Theorem 3, ϕ admits a derivative, which is precisely the function g.

COROLLARY 2. *If a function f is bi-differentiable in an open convex set $C \subset \mathbf{R}^n$, it is convex in C if and only if the quadratic form*

$$\sum_{j=1}^{n} \sum_{i=1}^{n} h^i h^j f_{ij}''(c_0)$$

is positive for all $c_0 \in C$.

Proof. The function f is convex if (and only if) for all c and c' in C and t in $[0, 1]$, we have

$$\phi''(t) = \sum_{j=1}^{n} \sum_{i=1}^{n} (c^i - c'^i)(c^j - c'^j) f''_{ij}\,[tc+(1-t)\,c'] \geqq 0.$$

Since $tc+(1-t)c' = c_0$ is a point of C and C is open, this condition can be written

$$(\forall h) : \sum_{j=1}^{n} \sum_{i=1}^{n} h^i h^j f''_{ij}(c_0) \geqq 0.$$

APPLICATION. *If* $(p_1, p_2, \ldots, p_n) \in \mathbf{P}_n$ *and* $x_1, x_2, \ldots, x_n$ *are positive numbers, then*

$$p_1 x_1 + p_2 x_2 + \ldots + p_n x_n \geqq (x_1)^{p_1} (x_2)^{p_2} \ldots (x_n)^{p_n}.$$

Proof. The function defined by $f(x) = \log x$ is concave, for

$$f''(x) = -\frac{1}{x^2} \leqq 0.$$

Therefore, by Theorem 1 of § 5,

$$\log\Big(\sum_{i=1}^{n} p_i\, x_i\Big) \geqq \sum_{i=1}^{n} p_i \log x_i = \log\Big[\prod_{i=1}^{n} (x_i)^{p_i}\Big]$$

and the required inequality follows (in particular, we deduce that 'the arithmetic mean is greater than the geometric mean').

§7. The fundamental properties of convex functions

Nearly all the known properties of convex functions which are difficult to prove can be deduced from the two fundamental theorems which we are now going to prove.

First fundamental theorem. *Let C be a convex set $\subset \mathbf{R}^n$ and let $f_1, f_2, \ldots, f_m$ be convex functions. If the system*

$$\begin{cases} f_i(x) \leqq 0 & (i = 1, 2, \ldots, k) \\ f_i(x) < 0 & (i = k+1, k+2, \ldots, m) \end{cases}$$

admits no solution $x \in C$, there exists a function f, given by

$$f(x) = \sum_{i=1}^{m} p_i f_i(x), \quad \textit{where } (p_1, p_2, \ldots, p_m) \in \mathbf{P}_m,$$

such that

$$f(x) \geqq 0, \quad (x \in C).$$

Proof. Write

$$G(x) = \{(\xi_1, \xi_2, \ldots, \xi_m) \,/\, \xi_i \geqq f_i(x) \text{ for } i = 1, 2, \ldots, k; \\ \xi_i > f_i(x) \text{ for } i = k+1, k+2, \ldots, m\}.$$

By hypothesis, the set $G = \bigcup_{x \in C} G(x)$ does not contain the point 0. Also G is convex, since if $\xi \in G(x)$, $\eta \in G(y)$, p and q are positive and $p+q = 1$, then

$$\begin{aligned} p\xi_i+q\eta_i &\geqq pf_i(x)+qf_i(y) \geqq f_i(px+qy) \quad (i \leqq k), \\ p\xi_i+q\eta_i &> pf_i(x)+qf_i(y) \geqq f_i(px+qy) \quad (i > k) \end{aligned}$$

and so $p\xi+q\eta \in G(px+qy)$.

By the first separation theorem, two disjoint convex sets in $\mathbf{R}^m$ can be separated by a plane. Therefore there exist coefficients $p_1, p_2, \ldots, p_m$ (not all zero) such that

$$\sum_{i=1}^{m} p_i \xi_i \geqq 0 \qquad ((\xi_1, \xi_2, \ldots, \xi_m) \in G).$$

Since each ξ_i can be chosen as large as we please, $p_i \geqq 0$. We can also suppose that $\sum_{i=1}^{m} p_i = 1$, because we can always multiply the p_i by a constant factor without affecting the situation.

Let ε be a strictly positive number; writing $\xi_i = f_i(x)+\varepsilon$, we have

$$\sum_{i=1}^{m} p_i f_i(x)+\varepsilon \geqq 0 \qquad (x \in C).$$

Hence, if

$$f = \sum p_i f_i,$$

we have

$$\inf_{x \in C} f(x) \geqq -\varepsilon.$$

This holds for all strictly positive ε and therefore we have

$$f(x) \geqq 0$$

as required.

REMARK. We can show that, in the above theorem, we can take all but $n+1$ of the p_i to be zero.

COROLLARY. *Let C be a convex set in $\mathbf{R}^n$ and let $f_1, f_2, \ldots, f_m$ be concave functions. If the system*

$$\begin{aligned} f_i(x) &\geqq 0 \qquad (i = 1, 2, \ldots, k), \\ f_i(x) &> 0 \qquad (i = k+1, k+2, \ldots, m) \end{aligned}$$

admits no solution $x \in C$, then there exists a function f given by

$$f(x) = \sum_{i=1}^{m} p_i f_i(x), \text{ where } (p_1, p_2, \ldots, p_m) \in \mathbf{P}_m,$$

such that

$$f(x) \leqq 0, \qquad (x \in C).$$

This can be deduced at once by changing the signs of the functions occurring in the theorem.

Second fundamental theorem. *Let C be a compact convex set in $\mathbf{R}^n$ and let $(f_k \mid k \in K)$ be continuous convex functions (not necessarily finite in number). If the system*

$$f_k(x) \leqq 0 \qquad (k \in K)$$

does not admit a solution $x \in C$, there exists a function f given by

$$f(x) = \sum_{i=1}^{m} p_i f_{k_i}(x), \text{ where } (p_1, p_2, \ldots, p_m) \in \mathbf{P}_m,$$

such that

$$f(x) > 0 \qquad (x \in C).$$

Proof. With each $x_i \in C$ we associate an index $k_i \in K$ and a number $\varepsilon_i > 0$ such that

$$f_{k_i}(x_i) > \varepsilon_i.$$

The sets $G_i = \{x \mid f_{k_i}(x) > \varepsilon_i\}$ form an open covering of C; since C is compact, this contains a finite open covering, say $G_1, G_2, \ldots, G_m$. Consider the single-valued mapping of C into R^m defined by

$$x \to \bar{x} = (\xi_1, \xi_2, \ldots, \xi_m) = (f_{k_1}(x), f_{k_2}(x), \ldots, f_{k_m}(x)).$$

The image $\bar{C}$ of C does not meet the set $\bar{A}$ in $\mathbf{R}^m$ given by

$$\xi_i < \varepsilon_i \qquad (i = 1, 2, \ldots, m)$$

and the same is true of the convex closure $[\bar{C}]$ of $\bar{C}$, for otherwise there would exist $x_1, x_2, \ldots, x_m \in C$ such that

$$\bar{y} = p_1\bar{x}_1 + p_2\bar{x}_2 + \ldots + p_m\bar{x}_m \in \bar{A},$$
$$(p_1, p_2, \ldots, p_m) \in \mathbf{P}_m;$$

and this is impossible, because for at least one index $i \leqq m$ we have

$$\bar{y}_i = p_1 f_{k_i}(x_1) + p_2 f_{k_i}(x_2) + \ldots + p_m f_{k_i}(x_m)$$
$$\geqq f_{k_i}(p_1x_1 + p_2x_2 + \ldots + p_mx_m) > \varepsilon_i.$$

Thus we have a non-empty open convex set $\bar{A} \subset \mathbf{R}^n$ and a convex set $[\bar{C}]$ which does not meet $\bar{A}$; therefore these sets can be separated by a plane with equation

$$\sum_{i=1}^{m} p_i \xi_i = \sum_{i=1}^{m} p_i \varepsilon_i,$$

where

$$(p_1, p_2, \ldots, p_m) \in \mathbf{P}_m.$$

Hence

$$\sum_{i=1}^{m} p_k f_{k_i}(x) \geqq \sum_{i=1}^{m} p_i \varepsilon_i > 0 \qquad (x \in C)$$

and so the theorem is proved.

REMARK 1. It is important to note that neither of the two fundamental theorems is implied by the other. Suppose that the system

$$f_i(x) \leqq 0 \qquad (i = 1, 2, \ldots, m)$$

is inconsistent over a convex set C; if C is not compact, we can apply the first fundamental theorem and if C is compact we can apply the second fundamental theorem to obtain a stronger result. But the hypothesis that C is compact is essential, as the following example shows: consider the convex set

$$C = \{x = (x_1, x_2) \,/\, x_2 > 0\} \cup \{x \,/\, x_1 > 0,\, x_2 = 0\}$$

in $\mathbf{R}^2$. As indicated in figure 44, this set is convex and not compact. The system

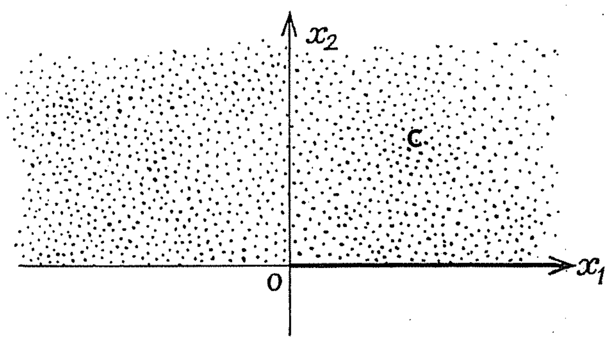


FIG. 44

$$f_1(x) = x_1 \leqq 0,$$
$$f_2(x) = x_2 \leqq 0$$

does not admit a solution $x \in C$, but this does not imply the existence of coefficients p_1, p_2 such that

$$x \in C \;\Rightarrow\; p_1 f_1(x) + p_2 f_2(x) = p_1 x_1 + p_2 x_2 > 0,$$
$$(p_1, p_2) \in \mathbf{P}_2.$$

For clearly if $p_1 > 0$ then $p_1 x_1 + p_2 x_2$ cannot be positive for all $(x_1, x_2) \in C$ (we can choose any value we please for x_1) and thus the condition $p_1 x_1 + p_2 x_2 > 0$ becomes $x_2 > 0$; but, by definition, C contains points for which $x_2 = 0$.

REMARK 2. The second fundamental theorem is equivalent to the following result: *if the system $f_k(x) \leqq 0$ does not admit a solution in a compact convex set $C \subset \mathbf{R}^n$, then there exist coefficients $p_1, p_2, \ldots, p_{n+1}$ such that $(p_1, p_2, \ldots, p_{n+1}) \in \mathbf{P}_{n+1}$ and functions $f_{k_1}, f_{k_2}, \ldots, f_{k_{n+1}}$ such that*

$$\sum_{i=1}^{n+1} p_i f_{k_i} > 0 \qquad (x \in C).$$

This theorem is due to Bohnenblust, Karlin and Shapley. We shall show that it follows from the second fundamental theorem; that the converse is true is trivial.

Consider the compact convex sets

$$C_k = \{x \mid x \in C, f_k(x) \leqq 0\}.$$

By hypothesis, these have an empty intersection: suppose that the intersection of any $n+1$ of them is non-empty. Then every finite intersection of the C_k is non-empty, by Helly's Theorem (§ 1). Since C is a compact space, it follows from the finite intersection axiom (page 69) that the intersection of all the C_k is non-empty, which is contrary to hypothesis.

Suppose then that $C_{k_1}, C_{k_2}, \ldots, C_{k_{n+1}}$ are $n+1$ convex sets having an empty intersection. Applying the second fundamental theorem to the functions $f_{k_1}, f_{k_2}, \ldots, f_{k_{n+1}}$, we obtain the desired result.

Minimax theorem (von Neumann).[1] *Let $C \subset \mathbf{R}^m$ and $D \subset \mathbf{R}^n$ be two non-empty compact convex sets and let f be a numerical function defined on $\mathbf{R}^m \times \mathbf{R}^n = \mathbf{R}^{m+n}$, such that $x \to f(x, y)$ is upper semi-continuous and concave in C (for each $y \in D$) and $y \to f(x, y)$ is lower semi-continuous and convex in D (for each $x \in C$). Then there exist points $x_0 \in C$ and $y_0 \in D$ such that*

$$f(x_0, y_0) = \max_{x \in C} f(x, y_0) = \min_{y \in D} f(x_0, y).$$

Proof. Write

$$\alpha = \inf_{y \in D} \sup_{x \in C} f(x, y) = \min_{y \in D} \max_{x \in C} f(x, y),$$

$$\beta = \sup_{x \in C} \inf_{y \in D} f(x, y) = \max_{x \in C} \min_{y \in D} f(x, y)$$

(see Theorems 2 and 3, § 8, Chapter IV).

We first prove that $\alpha = \beta$. We have

$$(\forall x_0)\,(\forall y) : \max_{x \in C} f(x, y) \geqq f(x_0, y)$$

[1] The 'minimax' theorems, which are given here and in § 8 in a very general form, play an important part in the theory of games.

and so

$$(\forall x_0) : \min_{y \in D} \max_{x \in C} f(x, y) \geqq \min_{y \in D} f(x_0, y),$$

hence

$$\alpha = \min_{y \in D} \max_{x \in C} f(x, y) \geqq \max_{x \in C} \min_{y \in D} f(x, y) = \beta.$$

Therefore $\alpha \geqq \beta$.

Also

$$(\forall y) : \max_{x \in C} f(x, y) \geqq \alpha.$$

Given any strictly positive number ε, the function g_x defined by

$$g_x(y) = f(x, y) - \alpha + \varepsilon,$$

is convex and lower semi-continuous. For all y in D, there exists a function g_x such that $g_x(y) > 0$; therefore, by the second fundamental theorem, there exists a function ϕ_0 such that

$$\phi_0(y) = \sum_{i=1}^{n+1} p_i \, g_{x_i}(y),$$

$$(\forall y) : \phi_0(y) > 0.$$

Therefore

$$0 < \phi_0(y) = \sum p_i [f(x_i, y) - \alpha + \varepsilon] = \sum p_i f(x_i, y) - \alpha + \varepsilon$$
$$\leqq f(\sum p_i x_i, y) - \alpha + \varepsilon,$$

whence

$$\min_{y \in D} f(\sum p_i x_i, y) > \alpha - \varepsilon$$

and so

$$\beta = \max_{x \in C} \min_{y \in D} f(x, y) > \alpha - \varepsilon.$$

Since this is true for all ε, we have $\beta \geqq \alpha$; it therefore follows that $\alpha = \beta$.

Suppose now that x_0 and y_0 are points of C and D such that

$$\max_{x \in C} f(x, y_0) = \min_{y \in D} \max_{x \in C} f(x, y) = \max_{x \in C} \min_{y \in D} f(x, y) = \min_{y \in D} f(x_0, y) = \alpha.$$

We have

$$(\forall x) : f(x, y_0) \leqq \alpha,$$
$$(\forall y) : f(x_0, y) \geqq \alpha.$$

Then

$$\alpha \leqq f(x_0, y_0) \leqq \alpha$$

and so

$$\alpha = f(x_0, y_0) = \max_{x \in C} f(x, y_0) = \min_{y \in D} f(x_0, y).$$

APPLICATION. *Games of Strategies.* Let (A) and (B) be two players and let $\mathbf{A}_N^M$ be a matrix of m rows and n columns. Simultaneously the player (A) chooses a row i and the player (B) chooses a column j. This determines a coefficient a_j^i of the matrix, called the *result* of the game. We examine the following question: how should the players proceed if (A) is attempting to obtain a result as large as possible and (B) is attempting to obtain a result as small as possible? We usually regard a_j^i as a sum of money which (A) receives from (B); if a_j^i is negative, (B) receives the money from (A).

We shall consider two methods of playing.

Method (1). Before the game, (A) decides on a row i, which guarantees a result $\min_j a_j^i$; thus the greatest result that (A) can guarantee is

$$\alpha' = \max_i \min_j a_j^i.$$

Similarly the smallest result that (B) can guarantee is

$$\beta' = \min_j \max_i a_j^i.$$

For example, consider the matrix

$$\begin{array}{cc} j=1 & j=2 \\ \left(\begin{array}{cc} 0 & +1 \\ -2 & +2 \\ +2 & -2 \end{array}\right) \end{array} \begin{array}{l} \\ i=1 \\ i=2 \\ i=3. \end{array}$$

In this case (A) can guarantee $\alpha' = 0$ (choosing the 'pure' strategy $i = 1$) and (B) can guarantee $\beta' = 2$ (choosing either of the two pure strategies). But between 0 and 2, there is an interval of uncertainty, which seems to defy analysis.

Method (2). Instead of deciding *a priori* on a particular row, the player (A) restricts himself to choosing a probability distribution and leaves the final decision to a machine which 'draws lots' for one of the events with the stated probabilities.

Thus, in the above example, the players (A) and (B) restrict their respective choices to points of the sets

$$C = \{x = (x^1, x^2, x^3) \,/\, x^1, x^2, x^3 \geqq 0;\quad x^1+x^2+x^3 = 1\},$$
$$D = \{y = (y^1, y^2) \,/\, y^1, y^2 \geqq 0;\quad y^1+y^2 = 1\}.$$

Having chosen $x \in C$ and $y \in D$, the players must expect a 'mean result' of

$$f(x, y) = \sum_{i,j} a_j^i x^i y^j.$$

The greatest mean result which (A) can guarantee is

$$\alpha = \max_{x \in C} \min_{y \in D} f(x, y)$$

The smallest mean result which (B) can guarantee is

$$\beta = \min_{y \in D} \max_{x \in C} f(x, y).$$

By the minimax theorem, we have $\alpha = \beta$ and so the interval of uncertainty is now made good (which shows the superiority of this method of playing).

In the above example, with the 'combined strategy' $x = (4/5, 0, 1/5)$ the player (A) is guaranteed a mean result of 2/5 (which is preferable to a certain result of 0, if only that the game should continue). With the combined strategy $y = (3/5, 2/5)$, the player (B) is guaranteed a mean result of 2/5 (instead of a certain result of 2).

§ 8. Quasi convex functions

Let f be a function defined in a convex set $C \subset \mathbf{R}^n$, with values $f(x) = f(x^1, x^2, \dots, x^n)$. We say that f is **quasi convex** in C if the set

$$\{x \mid x \in C;\ f(x) \leqq \alpha\}$$

is convex for all α. Similarly we say that a function f is **quasi concave** if the set $\{x \mid x \in C;\ f(x) \geqq \alpha\}$ is convex for all α. A function f is quasi concave if and only if $-f$ is quasi convex.

EXAMPLE 1. If f is a convex function, then

$$\begin{cases} f(x) \leqq \alpha \\ f(x') \leqq \alpha \end{cases} \Rightarrow f(px+p'x') \leqq pf(x)+p'f(x') \leqq p\alpha+p'\alpha = \alpha$$

and so f is also quasi convex.

EXAMPLE 2. Let f be a function in $\mathbf{R}$, decreasing in the range $-\infty$ to t_0 and increasing in the range t_0 to $+\infty (t_0 \in \hat{\mathbf{R}})$. Then f is quasi convex, for the set $\{x \mid f(x) \leqq \alpha\}$ is either $\varnothing$ or an interval containing t_0.

(This example shows that a quasi convex function is not necessarily convex. Thus, by the first example, the family of quasi convex functions is larger than the family of convex functions.)

Theorem 1. *If $f_1, f_2, \dots, f_m$ are convex functions in a convex set $C \subset \mathbf{R}^n$ and if ϕ is an increasing quasi convex function in $\mathbf{R}^m$, then the function g given by*

$$g(x) = \phi[f_1(x), f_2(x), \dots, f_m(x)]$$

is quasi convex in C.

Proof. If $g(x) \leqq \alpha$, $g(x') \leqq \alpha$, we have
$g(px+p'x') = \phi[f_1(px+p'x'), \dots] \leqq \phi[pf_1(x)+p'f_1(x'), \dots] \leqq \alpha.$

Theorem 2. *A function f is quasi convex in C if and only if, for all c and c' in C, the function ϕ given by $\phi(\lambda) = f(\lambda c+(1-\lambda)c')$ is quasi convex in $[0, 1]$.*

Proof. As in the proof of Theorem 2, § 5, we observe that

$$\begin{cases} c_1 = \lambda_1 c+(1-\lambda_1)c' \\ c_2 = \lambda_2 c+(1-\lambda_2)c' \end{cases} \Rightarrow p_1c_1+p_2c_2 = (p_1\lambda_1+p_2\lambda_2)c+[1-(p_1\lambda_1+p_2\lambda_2)]c'$$

where $(p_1, p_2) \in \mathbf{P}_2$ and $\lambda_1, \lambda_2 \in [0, 1]$. If f is quasi convex, then

$$\begin{cases} \phi(\lambda_1) \leqq \alpha \\ \phi(\lambda_2) \leqq \alpha \end{cases} \Rightarrow \begin{cases} f(c_1) \leqq \alpha \\ f(c_2) \leqq \alpha \end{cases} \Rightarrow f(p_1c_1+p_2c_2) \leqq \alpha \Rightarrow \phi(p_1\lambda_1+p_2\lambda_2) \leqq \alpha$$

and so ϕ is quasi convex.

Conversely, if ϕ is quasi convex in $[0, 1]$, then

$$\begin{cases} f(c_1) \leqq \alpha \\ f(c_2) \leqq \alpha \end{cases} \Rightarrow \begin{cases} \phi(\lambda_1) \leqq \alpha \\ \phi(\lambda_2) \leqq \alpha \end{cases} \Rightarrow \phi(p_1\lambda_1+p_2\lambda_2) \leqq \alpha \Rightarrow f(p_1c_1+p_2c_2) \leqq \alpha$$

and so f is quasi convex.

Corollary. *Let f be a differentiable function; then f is quasi convex in C if and only if, for all c and c' in C, the function defined by*

$$\phi'(\lambda) = \sum_{i=1}^{n} (c^i-c'^i) f'_i\,[\lambda c+(1-\lambda)c']$$

has constant sign in $[0, 1]$, or changes sign only once as λ varies from 0 to 1 (to pass from negative to positive).

Theorem 3. *Let K be a convex cone in $\mathbf{R}^n$ and let f be a function such that, for $x \in K-\{0\}$ and $\lambda \geqq 0$, we have*

$$f(x) > 0, \qquad f(\lambda x) = \lambda f(x).$$

Then the function f is convex (resp. concave) in K if and only if it is quasi convex (resp. quasi concave).

Proof. If f is convex in K, it is also quasi convex in K, as shown in Example 1 above.

Conversely, suppose that f is quasi convex in K. By Theorem 3, § 5, it is sufficient to prove that the set

$$\overline{T} = \{\bar{x} \,/\, \bar{x} = (x^1, x^2, \dots, x^{n+1}) \in \mathbf{R}^{n+1}, x = (x^1, x^2, \dots, x^n) \in K, f(x) \leqq x^{n+1}\}$$

is convex. We observe that $\overline{T}$ is a cone, for if $\lambda > 0$, then

$$\bar{a} \in \overline{T} \;\Rightarrow\; \lambda\bar{a} \in \overline{T}.$$

Every generator D_a of the cone $\bar{T}$ meets the plane $\bar{P}$ of equation $x^{n+1} = 1$; moreover, the intersection $\bar{A} = \bar{P} \cap \bar{T}$ projects into the set

$$A = \{x \,/\, x \in K,\ f(x) \leqq 1\}$$

in $\mathbf{R}^n$. By hypothesis, A is convex and therefore so is the set $\bar{A}$, whence $\bar{T}$, being the conical closure of the set $\bar{A}$, is a convex set (see Theorem 2, § 4, Chapter VII).

APPLICATION. *If $\lambda_1 \geqq 0$, $\lambda_2 \geqq 0, \ldots \lambda_n \geqq 0$, the function f given by*

$$f(x) = (x^1)^{\lambda_1}\,(x^2)^{\lambda_2} \ldots, (x^n)^{\lambda_n}$$

is quasi concave in the 'unpointed' cone

$$K_+ = \{x = (x^1, x^2, \ldots x^n) \,/\, x^1 > 0, x^2 > 0, \ldots x^n > 0\};$$

if the λ_i also satisfy $\lambda_1 + \lambda_2 + \ldots + \lambda_n \leqq 1$ then the function f is concave in K_+. (The set K_+ is not a cone according to our definition; however $K_+ \cup \{0\}$ is a cone and so we refer to K_+ as an unpointed cone.)

Proof. (1) We first prove that f is quasi concave in K_+. Since the logarithmic function is increasing, it is sufficient to prove that the function given by

$$\log f(x) = \lambda_1 \log x^1 + \lambda_2 \log x^2 + \ldots + \lambda_n \log x^n$$

is quasi concave. Since the function ϕ given by

$$\phi(z_1, z_2, \ldots, z_n) = \lambda_1 z_1 + \lambda_2 z_2 + \ldots + \lambda_n z_n$$

is increasing and concave and $h_i(x) = \log x^i$ is concave, then, by Theorem 4, § 5, the function given by

$$\log f(x) = \phi(h_1(x), h_2(x), \ldots, h_n(x))$$

is concave and so is quasi concave.

(2) We now show that, if $\lambda_1 + \lambda_2 + \ldots + \lambda_n \leqq 1$, then f is a concave function in K_+. If $\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1$, this follows at once from Theorem 3 and (1) above. Suppose therefore that $\lambda_1 + \lambda_2 + \ldots + \lambda_n < 1$ and put $\lambda_0 = 1 - (\lambda_1 + \lambda_2 + \ldots + \lambda_n)$. The function g defined by

$$g(x) = (x^1)^{\frac{\lambda_1}{1-\lambda_0}}\,(x^2)^{\frac{\lambda_2}{1-\lambda_0}} \ldots (x^n)^{\frac{\lambda_n}{1-\lambda_0}}$$

is concave in K_+, as has already been shown. Also the function defined by $\phi(t) = t^{1-\lambda_0}$ is concave and increasing in $[0, +\infty[$: therefore, by Theorem 4, § 5, $\phi(g(x)) = (x^1)^{\lambda_1}\,(x^2)^{\lambda_2} \ldots (x^n)^{\lambda_n}$ determines a concave function in K_+.

The theory of quasi convex functions is important in certain problems occurring in Applied Mathematics, for it enables us to generalise the minimax theorem of von Neumann (§ 7); in fact we shall prove the following generalisation of this theorem.

General minimax theorem (Sion).[1] *Let* $A \subset \mathbf{R}^m$ *and* $B \subset \mathbf{R}^n$ *be two compact convex sets and let* f *be a numerical function defined on* $A \times B$ *and such that* $x \to f(x, y)$ *is upper semi-continuous and quasi concave in* A *(for each* $y \in B$*) and* $y \to f(x, y)$ *is lower semi-continuous and quasi convex in* B *(for each* $x \in A$*). Then there exist points* $x_0 \in A$ *and* $y_0 \in B$ *such that*

$$f(x, y_0) \leqq f(x_0, y_0) \leqq f(x_0, y) \qquad (x \in A,\ y \in B).$$

Proof. (1) As in § 7, we have

$$\min_{y \in B} \max_{x \in A} f(x, y) \geqq \max_{x \in A} \min_{y \in B} f(x, y).$$

(2) Let γ_1 and γ_2 be numbers such that

$$\gamma_1 < \min_{y \in B} \max_{x \in A} f(x, y),$$
$$\gamma_2 > \max_{x \in A} \min_{y \in B} f(x, y).$$

With each $x \in A$ we associate the compact convex set

$$B_x = \{y \mid y \in B \text{ and } f(x, y) \leqq \gamma_1\}$$

and with each $y \in B$ we associate the compact convex set

$$A_y = \{x \mid x \in A \text{ and } f(x, y) \geqq \gamma_2\}.$$

We have

$$\bigcap_{x \in A} B_x = \bigcap_{y \in B} A_y = \varnothing.$$

Therefore there exist points $x_1, x_2, \ldots, x_p \in A$ such that $\bigcap_{i=1}^{p} B_{x_i} = \varnothing$ and also there exist points $y_1, y_2, \ldots, y_q \in B$ such that $\bigcap_{j=1}^{q} A_{y_j} = \varnothing$.

Let $\mathscr{F}$ be the family of pairs (I, J) satisfying

$$I \subset (1, 2, \ldots, p), \quad I \neq \varnothing, \qquad J \subset (1, 2, \ldots, q), \quad J \neq \varnothing$$

such that the intersection of the $(B_{x_i})_{i \in I}$ does not meet the convex polyhedron generated by the $(y_j)_{j \in J}$ and the intersection of the $(A_{y_j})_{j \in J}$ does not meet the convex polyhedron generated by the $(x_i)_{i \in I}$. This family is finite and non-empty; let (I_0, J_0) be a minimal element: that is, an element such that

$$(I, J) \in \mathscr{F},\ I \subset I_0 \text{ and } J \subset J_0 \;\Rightarrow\; I = I_0 \text{ and } J = J_0.$$

[1] Cf. M. Sion, *C.R. Acad. Sciences*, vol. 244, p. 2120, 15th April 1957. There exist other minimax theorems using the idea of quasi convex functions, but with more restrictive topological hypotheses. One of the most elegant proofs is contained in the theorem due to Nash, which supposes $f(x, y)$ to be continuous in (x, y) and which depends on Kakutani's Theorem (cf. Berge, *Théorie générale des jeux*, Gauthier-Villars, 1957, p. 72). The result due to Nikaido, which supposes only that $f(x, y)$ is continuous separately in x and y, depends on Brouwer's Theorem (cf. H. Nikaido, *Pac. J. of Math.*, 1954, p. 65). We observe that whenever we use the idea of quasi convexity, the proof appeals to Sperner's lemma or its consequences.

Let r and s be the numbers of elements of I_0 and J_0 respectively. Let A' be the convex polyhedron generated by the $(x_i)_{i \in I_0}$ and let B' be the convex polyhedron generated by the $(y_j)_{j \in J_0}$.

The intersection of the sets $(A_{y_j})_{j \in J_0}$ does not meet A', but the intersection of any $s-1$ of them meets A', since (I_0, J_0) is minimal. It follows from the intersection theorem (§ 1) that A' is not contained in the union of the $(A_{y_j})_{j \in J_0}$. Let x_0 be a point of A' not belonging to this union. We have

$$f(x_0, y_j) < \gamma_2 \text{ for } j \in J_0,$$

and, since $f(x, y)$ is quasi convex in y,

$$f(x_0, y) < \gamma_2 \text{ for } y \in B'.$$

In a similar manner, we can show that there exists a point y_0 in B' such that

$$f(x, y_0) > \gamma_1 \text{ for } x \in A'.$$

We therefore have

$$\gamma_1 < f(x_0, y_0) < \gamma_2.$$

(3) In part (1), we proved that

$$\min_{y \in B} \max_{x \in A} f(x, y) \geqq \max_{x \in A} \min_{y \in B} f(x, y)$$

and in (2), we showed that

$$\gamma_1 < \min_{y \in B} \max_{x \in A} f(x, y) \text{ and } \gamma_2 > \max_{x \in A} \min_{y \in B} f(x, y) \Rightarrow \gamma_1 < \gamma_2.$$

Combining these results, we obtain

$$\min_{y \in B} \max_{x \in A} f(x, y) = \max_{x \in A} \min_{y \in B} f(x, y).$$

(4) Let $x_0 \in A$ and $y_0 \in B$ be such that

$$\max_{\substack{x \in A \\ y \in B}} [f(x, y_0) - f(x_0, y)] = \min_{\substack{x' \in A \\ y' \in B}} \max_{\substack{x \in A \\ y \in B}} [f(x, y') - f(x', y)] = 0.$$

We have

$$f(x, y_0) \leqq f(x_0, y) \text{ for } x \in A, y \in B,$$

whence

$$f(x, y_0) \leqq f(x_0, y_0) \leqq f(x_0, y) \text{ for } x \in A, y \in B.$$

Thus x_0 and y_0 are the required points.

§ 9. The fundamental inequality of convexity

In this section we consider

(1) the convex unpointed cone $K_+ = \{x \,/\, x^1 > 0, x^2 > 0, \dots x^n > 0\}$ in $\mathbf{R}^n$ (see § 8) and the convex cone $K_0 = \{x \,/\, x^1 \geqq 0, x^2 \geqq 0, \dots, x^n \geqq 0\}$;

(2) the vector space Φ of numerical functions defined on an arbitrary set T and the cone Φ_0 of positive (strictly or otherwise) functions in the set Φ.

A gauge j, defined on Φ_0, is said to be **increasing** if

$$\left\{\begin{array}{l} \phi, \phi' \in \Phi_0 \\ (\forall t) : \phi(t) \geqq \phi'(t) \end{array}\right. \Rightarrow \quad j(\phi) \geqq j(\phi').$$

Fundamental theorem (Bourbaki). *Let f be a numerical function continuous and concave in K_0 and such that, for $x \in K_+$, $\lambda \geqq 0$ we have*

$$f(x) > 0; \qquad f(\lambda x) = \lambda f(x).$$

Let $\phi_1, \phi_2, \ldots, \phi_n$ be functions in Φ_0 and let j be an increasing gauge in Φ_0 such that $j(\phi_1), j(\phi_2), \ldots, j(\phi_n) < +\infty$. Then the function given by $g(t) = f[\phi_1(t), \phi_2(t), \ldots, \phi_n(t)]$ satisfies

$$j(g) \leqq f[j(\phi_1), j(\phi_2), \ldots, j(\phi_n)].$$

Proof. In $\mathbf{R}^n$, the set

$$C = \{x \,/\, x \in K_0, f(x) \geqq 1\}$$

is closed (for f is continuous in K_0 and K_0 is closed) and convex (for f is concave). Then C is the intersection of the closed half-spaces which contain it (by Theorem 3, § 1); we suppose that these half-spaces are given by

$$f^h(x) = \sum_{i=1}^{n} a_i^h x^i \geqq b^h.$$

(1) If $x \in K_+$, we have $\lambda x \in C$ whenever λ is greater than a certain number λ_0, so that

$$\lambda \geqq \lambda_0 \quad \Rightarrow \quad \lambda f^h(x) = f^h(\lambda x) \geqq b^h.$$

This shows that $f^h(x) \geqq 0$ (for all x in K_+) so that all the a_i^h are positive. Furthermore, the half-space of equation $f^h(x) \geqq 0$ contains C, for otherwise there exists an x in C such that $f^h(x) < 0$ (which contradicts $a_i^h \geqq 0$); thus we can assume that all the b^h are positive. Because of the continuity of f in K_0 we have

$$f(x) \geqq 0; \quad f(\lambda x) = \lambda f(x)$$

for $\lambda \geqq 0$ and $x \in K_0$.

(2) We now show that

$$\sum_{i=1}^{n} a_i^h \phi_i(t) \geqq b^h g(t).$$

If $g(t) = 0$, this inequality is trivial. We therefore suppose that $g(t) \neq 0$. Consider the point $y(t)$ whose coordinates are

$$y^i = \frac{\phi_i(t)}{g(t)}.$$

Then $y(t)$ belongs to C, for

$$f[y(t)] = f\left(\frac{\phi_1(t)}{g(t)}, \frac{\phi_2(t)}{g(t)}, \ldots, \frac{\phi_n(t)}{g(t)}\right) = \frac{1}{g(t)} f[\phi_1(t), \phi_2(t), \ldots, \phi_n(t)] = 1.$$

Therefore we have

$$\sum_{i=1}^{n} a_i^h \frac{\phi_i(t)}{g(t)} \geqq b^h,$$

so that, since $g(t) > 0$,

$$\sum_{i=1}^{n} a_i^h \phi_i(t) \geqq b^h g(t).$$

(3) We now show that

$$f[j(\phi_1), j(\phi_2), \ldots, j(\phi_n)] \geqq j(g).$$

The case $j(g) = 0$ is trivial; we therefore suppose that $j(g) \neq 0$. Since j is increasing in Φ_0, we have

$$\sum a_i^h j(\phi_i) \geqq j(\sum a_i^h \phi_i) \geqq j(b^h g) = b^h j(g).$$

Then

$$z = \left(\frac{j(\phi_1)}{j(g)}, \frac{j(\phi_2)}{j(g)}, \ldots, \frac{j(\phi_n)}{j(g)}\right) \in C$$

and so

$$f(z) = f\left(\frac{j(\phi_1)}{j(g)}, \frac{j(\phi_2)}{j(g)}, \ldots, \frac{j(\phi_n)}{j(g)}\right) \geqq 1.$$

Since $j(g) > 0$, we have

$$f[j(\phi_1), j(\phi_2), \ldots, j(\phi_n)] \geqq j(g).$$

REMARK. When f satisfies the hypotheses of the theorem apart from concavity, the following conditions are equivalent:

(1) f is concave in K_0,
(2) f is quasi concave in K_0,
(3) f is quasi concave in K_+,
(4) f is concave in K_+.

It is clear that (1) $\Rightarrow$ (2) $\Rightarrow$ (3). By Theorem 3, § 8, (3) $\Rightarrow$ (4); and (4) $\Rightarrow$ (1) because of the continuity of f on K_0.

APPLICATION 1. *The Hölder inequalities.*

Let f be the function defined by $f(x^1, x^2) = (x^1)^{\lambda_1}(x^2)^{\lambda_2}$, where $\lambda_1, \lambda_2 > 0$, $\lambda_1+\lambda_2 = 1$. Then f is a continuous and concave function in K_+ (see page 209). Therefore

$$j[(\phi_1)^{\lambda_1}(\phi_2)^{\lambda_2}] \leqq [j(\phi_1)]^{\lambda_1}[j(\phi_2)]^{\lambda_2}.$$

Putting $\lambda_i = \dfrac{q}{p_i}$, $(\phi_i)^{\frac{1}{p_i}} = g_i$, we obtain

$$j(g_1^q g_2^q) \leqq [j(g_1^{p_1})]^{\frac{q}{p_1}}[j(g_2^{p_2})]^{\frac{q}{p_2}},$$

whence

$$\boxed{[j(g_1^q g_2^q)]^{\frac{1}{q}} \leqq [j(g_1^{p_1})]^{\frac{1}{p_1}}[j(g_2^{p_2})]^{\frac{1}{p_2}}; \qquad \frac{1}{p_1}+\frac{1}{p_2} = \frac{1}{q}}$$

(1) If $T = \{1, 2, \ldots\}$, $g_1(n) = |a_n|$, $g_2(n) = |b_n|$, $j(g) = \sum_{n=1}^{\infty}|g(n)|$,

we have

$$\sqrt[q]{\sum |a_n|^q |b_n|^q} \leqq \sqrt[p_1]{\sum |a_n|^{p_1}}\,\sqrt[p_2]{\sum |b_n|^{p_2}},$$

(first Hölder inequality).

If we put $q = 1$, $p_1 = p_2 = 2$, we recover the Cauchy-Schwartz inequality (§ 1).

(2) If $T = [0, 1]$, $j(g) = \int_0^1 |g(t)|\,\mathrm{d}t$, we have

$$\sqrt[q]{\int_0^1 |g_1(t)\,g_2(t)|^q\,\mathrm{d}t} \leqq \sqrt[p_1]{\int_0^1 |g_1(t)|^{p_1}\,\mathrm{d}t}\;\sqrt[p_2]{\int_0^1 |g_2(t)|^{p_2}\,\mathrm{d}t}$$

(second Hölder inequality).

APPLICATION 2. *The Minkowski inequalities.*

From Corollary 2 to Theorem 4 of § 6, we deduce that the function given by

$$f(x^1, x^2) = [(x^1)^\lambda+(x^2)^\lambda]^{\frac{1}{\lambda}}; \qquad \lambda \leqq 1$$

is concave in K_+. We therefore have

$$j[(\phi_1^\lambda+\phi_2^\lambda)^{\frac{1}{\lambda}}] \leqq [[j(\phi_1)]^\lambda+[j(\phi_2)]^\lambda]^{\frac{1}{\lambda}}.$$

Putting $p = \frac{1}{\lambda}$, $(\phi_i)^{\frac{1}{p}} = g_i$, we obtain

$$\boxed{(j[(g_1+g_2)^p])^{\frac{1}{p}} \leqq (j(g_1^p))^{\frac{1}{p}}+(j(g_2^p))^{\frac{1}{p}}; \quad p \geqq 1}$$

(1) If $T = \{1, 2, \ldots\}$, $g_1(n) = | a_n |$, $g_2(n) = | b_n |$, $j(g) = \sum_{i=1}^{\infty} | g(n) |$, we have

$$\sqrt[p]{\sum | a_n+b_n |^p} \leqq \sqrt[p]{\sum | a_n |^p}+\sqrt[p]{\sum | b_n |^p},$$

(first Minkowski inequality).

If we put $p = 2$, we recover the triangular inequality for the Euclidean norm.

(2) If $T = [0, 1]$, $j(g) = \int_0^1 | g(t) | \,\mathrm{d}t$, we have

$$\sqrt[p]{\int_0^1 | g_1(t)+g_2(t) |^p \,\mathrm{d}t} \leqq \sqrt[p]{\int_0^1 | g_1(t) |^p \,\mathrm{d}t}+\sqrt[p]{\int_0^1 | g_2(t) |^p \,\mathrm{d}t},$$

(second Minkowski inequality).

§ 10.* Sub-Φ functions[1]

Let $[x_1, x_2] = \{p_1x_1+p_2x_2 \,/\, (p_1, p_2) \in \mathbf{P}_2\}$ be an interval in $\mathbf{R}^n$ and let $D(x_1, x_2) = \{\lambda_1x_1+\lambda_2x_2 \,/\, \lambda_1+\lambda_2 = 1\}$ be the straight line passing through x_1 and x_2. We denote by Φ a family of numerical functions defined in a convex set $C \subset \mathbf{R}^n$ such that, if $\phi, \phi' \in \Phi$, then

$$(1) \quad \left\{\begin{array}{l} \phi(x_1) < \phi'(x_1) \\ \phi(x_2) < \phi'(x_2) \end{array}\right. \Rightarrow \phi(x) < \phi'(x) \quad \text{for all } x \text{ in } [x_1, x_2],$$

$$(2) \quad \left\{\begin{array}{l} \phi(x_1) = \phi'(x_1) \\ \phi(x_2) = \phi'(x_2) \end{array}\right. \Rightarrow \phi(x) = \phi'(x) \quad \text{for all } x \text{ in } D(x_1, x_2) \cap C.$$

We say that a numerical function f is **sub-Φ** in C if

$$\left\{\begin{array}{l} \phi \in \Phi \\ x_1, x_2 \in C \\ f(x_1) \leqq \phi(x_1) \\ f(x_2) \leqq \phi(x_2) \end{array}\right. \Rightarrow f(x) \leqq \phi(x) \quad \text{for all } x \text{ in } [x_1, x_2].$$

[1] The idea of a 'sub-Φ function' was introduced by E. F. Beckenbach (*Bull. A.M.S.*, vol. 43, 1937, p. 363).

EXAMPLE 1. Let Φ be the family of linear affine functions; then conditions (1) and (2) are clearly satisfied. Let ϕ be a function in Φ such that

$$f(x_1) = \phi(x_1), \qquad f(x_2) = \phi(x_2).$$

Then f is sub-Φ if and only if

$$f(p_1x_1+p_2x_2) \leqq \phi(p_1x_1+p_2x_2) = p_1\,\phi(x_1)+p_2\,\phi(x_2) = p_1 f(x_1)+p_2 f(x_2).$$

In other words, the sub-Φ functions in this case are the convex functions.

EXAMPLE 2. Let Φ be the family of numerical functions with constant values. Clearly Φ satisfies conditions (1) and (2). A function f is sub-Φ if and only if

$$\begin{matrix} f(x_1) \leqq \alpha \\ f(x_2) \leqq \alpha \end{matrix} \quad \Rightarrow \quad f(p_1x_1+p_2x_2) \leqq \alpha.$$

In other words the sub-Φ functions in this case are the quasi convex functions.

We say that the family Φ is **total** if for all x_0 in C and α_0 in $\mathbf{R}$, there exists a function ϕ in Φ such that $\phi(x_0) = \alpha_0$. We say that Φ is **bi-total** if for all x_1 and x_2 in C, α_1 and α_2 in $\mathbf{R}$, there exists a function such that $\phi(x_1) = \alpha_1$ and $\phi(x_2) = \alpha_2$. In what follows, we suppose that Φ is a bi-total family. For a given function f and given points x_1 and x_2, we denote by ϕ_{12} any particular function in Φ such that

$$\begin{aligned} \phi_{12}(x_1) &= f(x_1), \\ \phi_{12}(x_2) &= f(x_2). \end{aligned}$$

PROPOSITION 1. *Let D be a straight line in $\mathbf{R}^n$ and let ϕ and ϕ' be functions in Φ. Then ϕ and ϕ' either*

(*i*) *do not have the same value at any point of D, or*
(*ii*) *have the same value at a unique point x_1 of D, or*
(*iii*) *have the same value at all points of D.*

In case (ii), the sign of $\phi(x)-\phi'(x)$ is constant on each side of x_1 and changes when we pass through x_1.

Proof. Let D be the straight line containing the points x_0, x_1, x_2 (in this order) and suppose that

$$\begin{aligned} \phi(x_1) &= \phi'(x_1), \\ \phi(x_2) &> \phi'(x_2). \end{aligned}$$

Then $\phi(x_0) \neq \phi'(x_0)$, for otherwise ϕ and ϕ' would coincide on D; and $\phi(x_0) \not> \phi'(x_0)$ for, since $x_1 \in [x_0, x_2]$,

$$\phi(x_0) > \phi'(x_0) \quad \Rightarrow \quad \phi(x_1) > \phi'(x_1).$$

Therefore we have $\phi(x_0) < \phi'(x_0)$.

Suppose now that x_3 is a point on D on the same side of x_1 as x_2. Then $\phi(x_3) \neq \phi'(x_3)$, for otherwise ϕ and ϕ' would coincide on D; and $\phi(x_3) \not< \phi'(x_3)$ for, since $x_1 \in [x_0, x_3]$,

$$\phi(x_3) < \phi'(x_3) \quad \Rightarrow \quad \phi(x_1) < \phi'(x_1).$$

PROPOSITION 2. *If f is a given sub-Φ function, we have*

$$f(x_3) \geqq \phi_{12}(x_3),$$

for all x_3 in $D(x_1, x_2)-[x_1, x_2]$.

Proof. Let x_3 be a point of $D(x_1, x_2)-[x_1, x_2]$ situated (for example) on the x_2 side. We have

$$\begin{aligned} \phi_{13}(x_1) &= f(x_1) = \phi_{12}(x_1), \\ \phi_{13}(x_2) &\geqq f(x_2) = \phi_{12}(x_2). \end{aligned}$$

If $\phi_{13}(x_2) = \phi_{12}(x_2)$, these functions coincide everywhere on the line, whence

$$f(x_3) = \phi_{13}(x_3) = \phi_{12}(x_3).$$

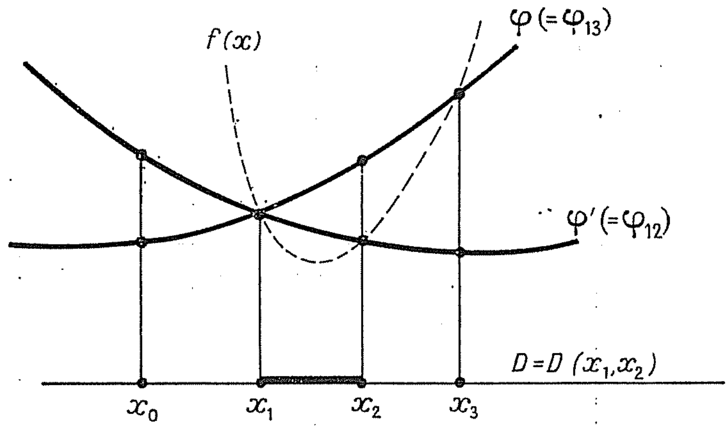


FIG. 45

If $\phi_{13}(x_2) > \phi_{12}(x_2)$, then, by Proposition 1,

$$f(x_3) = \phi_{13}(x_3) > \phi_{12}(x_3).$$

Therefore, taking both cases into account, we have

$$f(x_3) \geqq \phi_{12}(x_3).$$

Theorem 1. *If the functions in Φ are continuous and if $[x_1, x_2] \subset C$, a sub-Φ function f is continuous at $x_0 \in]x_1, x_2[$ and upper semi-continuous at x_1 and x_2, with respect to the usual topology of the segment $[x_1, x_2]$.*

Proof. Orientate the line $D(x_1, x_2)$ in such a way that $x_1 < x_2$; if $x_0 \in]x_1, x_2[$ there exists a positive number h such that

$$x_1 < x_0 - h < x_0 < x_0 + h < x_2.$$

Using Proposition 2, we have

$$\phi_{10}(x_0 - h) \geqq f(x_0 - h) \geqq \phi_{02}(x_0 - h).$$

As h tends to 0, $\phi_{10}(x_0-h)$ and $\phi_{02}(x_0-h)$ both tend to $f(x_0) = \phi_{10}(x_0) = \phi_{02}(x_0)$, so that $f(x_0-h)$ also tends to $f(x_0)$. Similarly $f(x_0+h)$ tends to $f(x_0)$. Therefore f is continuous at the point $x_0 \in]x_1, x_2[$.

Let ε be any strictly positive number. Then there exists a number η such that

$$h \leqq \eta \quad \Rightarrow \quad f(x_1+h) \leqq \phi_{12}(x_1+h) \leqq \phi_{12}(x_1)+\varepsilon = f(x_1)+\varepsilon.$$

Thus the function f is upper semi-continuous at the point x_1; a similar argument applies to the point x_2.

Theorem 2. *If Φ is a convex set in the vector space of numerical functions defined in $\mathbf{R}^n$, then so is the set of sub-Φ functions.*

Proof. Given two sub-Φ functions f' and f'', we must show that the function f such that $f(x) = p'f'(x)+p''f''(x)$, where $(p', p'') \in \mathbf{P}_2$, is also a sub-$\Phi$ function. Suppose that ϕ is a function in Φ such that

$$f(x_1) \leqq \phi(x_1); \quad f(x_2) \leqq \phi(x_2).$$

Let ϕ'_{12} and ϕ''_{12} be functions in Φ corresponding to f' and f'' respectively (for the two points x_1 and x_2) and put

$$\phi_0(x) = p'\phi'_{12}(x)+p''\phi''_{12}(x).$$

Then

$$\phi_0(x_1) = f(x_1) \leqq \phi(x_1),$$
$$\phi_0(x_2) = f(x_2) \leqq \phi(x_2).$$

Since $\phi_0 \in \Phi$, it is also a sub-Φ function (by Proposition 1) and so, for any point x in $[x_1, x_2]$, we have

$$f(x) = p'f'(x)+p''f''(x) \leqq p'\phi'_{12}(x)+p''\phi''_{12}(x) = \phi_0(x) \leqq \phi(x).$$

Therefore f is a sub-Φ function.

REMARK. We observe that Theorems 1 and 2 apply in particular to convex functions, but not to quasi convex functions, for which the family Φ is not bi-total.

§ 11. S-convex functions[1]

In this section, P denotes the set of bistochastic matrices $\mathbf{P} = \mathbf{P}_{\mathrm{N}}^{\mathrm{N}}$ of order n, where, as in §§ 3 and 4, $\mathbf{N}$ is the set $\{1, 2, \ldots, n\}$; D is an interval in $\mathbf{R}$ and D^n is the Cartesian product $D \times D \times \ldots \times D$.

If f is a numerical function defined in D^n, we say that f is *S*-**convex** in D^n if, for all x in D^n and all bistochastic matrices $\mathbf{P}$, we have

$$f(\mathbf{P}x) \leqq f(x).$$

We say that f is **strictly** *S*-**convex** in D^n if, further, we have

$$f(\mathbf{P}x) < f(x)$$

whenever $\mathbf{P}x = (y^1, y^2, \ldots, y^n)$ is not a permutation of $(x^1, x^2, \ldots, x^n)$.

Similarly we say that a function f is *S*-**concave** in D^n if for all x in D^n and all bistochastic matrices $\mathbf{P}$, we have

$$f(\mathbf{P}x) \geqq f(x)$$

(and we can also define strictly *S*-concave functions in the obvious way). A function f is *S*-concave if and only if $-f$ is *S*-convex.

EXAMPLE. The function f such that $f(x) = x^1+x^2+ \ldots +x^n$ is both *S*-convex and *S*-concave in $\mathbf{R}^n$, for, if $y = \mathbf{P}x$, we have

$$f(y) = y^1+y^2+ \ldots +y^n = x^1+x^2+ \ldots +x^n = f(x).$$

Similarly, if ϕ is a convex function, the function defined by $f(x) = \phi(x^1)+\phi(x^2)+ \ldots +\phi(x^n)$ is an *S*-convex function (cf. the Theorem of Hardy, Littlewood and Polya). We now generalise this result.

Theorem 1. *Let f be an S-convex (resp. S-concave) function which is increasing in D^n and let ϕ be a convex (resp. concave) function in $D_0 \subset \mathbf{R}$, with values in D; then the function g given by*

$$g(x) = f[\phi(x^1), \phi(x^2), \ldots, \phi(x^n)]$$

is S-convex (resp. S-concave) in $(D_0)^n$.

Proof. If $y = \mathbf{P}x$, where $x \in D^n$, we have

$$\phi(y^i) \leqq p_1^i\, \phi(x^1)+p_2^i\, \phi(x^2)+ \ldots +p_n^i\, \phi(x^n),$$

whence

$$(\phi(y^1), \phi(y^2), \ldots, \phi(y^n)) \leqq \mathbf{P}(\phi(x^1), \phi(x^2), \ldots, \phi(x^n)),$$

whence

$$g(y) = f(\phi(y^1), \ldots) \leqq f[\mathbf{P}(\phi(x^1), \ldots)] \leqq f[\phi(x^1), \ldots] = g(x)$$

and so g is *S*-convex in $(D_0)^n$.

[1] The idea of '*S*-convex function' is due to I. Schur (*Math. Z.*, vol. 12, 1922, p. 287).

Theorem 2. *If $f_1, f_2, \ldots, f_m$ are S-convex functions and if ϕ is an increasing function in $\mathbf{R}^m$, the function given by*

$$g(x) = \phi[f_1(x), f_2(x), \ldots, f_m(x)]$$

is S-convex.

Proof. If $y = \mathbf{P}x$, we have $f_i(y) \leqq f_i(x)$ for all i, so that

$$g(y) = \phi[f_1(y), \ldots] \leqq \phi[f_1(x), \ldots] = g(x)$$

and therefore g is S-convex.

Theorem 3. *An S-convex function f in $\mathbf{R}^n$ is symmetric in $x^1, x^2, \ldots, x^n$; that is, the value $f(x^1, x^2, \ldots, x^n)$ remains unaltered when we permute the x^i.*

Proof. Let $\mathbf{P}$ be a permutation matrix and let $\mathbf{Q}$ be the transposed permutation matrix (so that $q_j^i = p_i^j$). We have

$$f(x) = f(\mathbf{Q} \,.\, \mathbf{P}x) \leqq f(\mathbf{P}x) \leqq f(x).$$

Therefore $f(\mathbf{P}x) = f(x)$ for each permutation matrix $\mathbf{P}$ and so f is symmetric.

Theorem 4. *If a numerical function f in $\mathbf{R}^n$ is convex and symmetric with respect to $x^1, x^2, \ldots, x^n$, then f is S-convex.*

Proof. Let $\mathbf{P}$ be a bistochastic matrix; by the corollary to the theorem of Birkhoff and von Neumann, there exist positive numbers λ_i, with sum equal to 1, and permutation matrices $\mathbf{Q}_i$ such that

$$\mathbf{P} = \sum_{i=1}^{m} \lambda_i \mathbf{Q}_i.$$

Then

$$f(\mathbf{P}x) = f\Big(\sum_{i=1}^{m} \lambda_i \mathbf{Q}_i x\Big) \leqq \sum_{i=1}^{m} \lambda_i f(\mathbf{Q}_i x) = \sum_{i=1}^{m} \lambda_i f(x) = f(x),$$

and so f is S-convex.

It follows from this theorem that, for symmetric functions, the idea of S-convexity is a generalisation of the idea of convexity.

Lemma. *Let $k < n$ and let $\mathbf{P}_\mathrm{N}^\mathrm{N}$ be a bistochastic matrix such that*

$$(\forall i \leqq k)\,(\forall j > k)\,(\forall l)\,(\forall m > l) : p_j^l\, p_i^m = 0.$$

Then, if $\mathrm{K} = \{1, 2, \ldots, k\}$, we have $\mathbf{P}_\mathrm{N}^\mathrm{N} = \mathbf{P}_\mathrm{K}^\mathrm{K} \otimes \mathbf{P}_{\mathrm{N}-\mathrm{K}}^{\mathrm{N}-\mathrm{K}}$.

Proof. We can write

$$\mathbf{P}_\mathrm{N}^\mathrm{N} = \begin{pmatrix} \mathbf{P}_\mathrm{K}^\mathrm{K} & \mathbf{R}_{\mathrm{N}-\mathrm{K}}^\mathrm{K} \\ \mathbf{S}_\mathrm{N}^{\mathrm{N}-\mathrm{K}} & \mathbf{Q}_{\mathrm{N}-\mathrm{K}}^{\mathrm{N}-\mathrm{K}} \end{pmatrix}$$

If $r^l_j \neq 0$ ($l \in \mathrm{K}, j \in \mathrm{N}-\mathrm{K}$), then $s^m_i = 0$ for all $m > k$ and all $i \leqq k$; therefore $\mathbf{S}^{\mathrm{N}-\mathrm{K}}_{\mathrm{K}} = \mathbf{O}^{\mathrm{N}-\mathrm{K}}_{\mathrm{K}}$ and

$$\sum_{i \in \mathrm{N}-\mathrm{K}} \sum_{j \in \mathrm{N}-\mathrm{K}} q^i_j = n-k,$$

$$\sum_{i \in \mathrm{K}} \sum_{j \in \mathrm{K}} p^i_j = k,$$

whence

$$\sum_{i \in \mathrm{N}-\mathrm{K}} \sum_{j \in \mathrm{K}} r^i_j = 0.$$

But the r^i_j cannot be negative and so they are zero; hence, the supposition $r^l_j \neq 0$ leads to a contradiction. Therefore $\mathbf{R}^{\mathrm{K}}_{\mathrm{N}-\mathrm{K}} = \mathbf{O}^{\mathrm{K}}_{\mathrm{N}-\mathrm{K}}$ and similarly $\mathbf{S}^{\mathrm{N}-\mathrm{K}}_{\mathrm{K}} = \mathbf{O}^{\mathrm{N}-\mathrm{K}}_{\mathrm{K}}$.

Theorem 5 (Ostrowski). *Let D be an open interval in $\mathbf{R}$ and let f be a symmetric differentiable function in D^n. If, for all $x = (x^1, x^2, \ldots, x^n)$ in D^n such that $x^1 \neq x^2$, we have*

$$(x^2-x^1)\left(\frac{\partial f}{\partial x^2}-\frac{\partial f}{\partial x^1}\right) > 0,$$

then the function f is strictly S-convex in D^n.

Proof. (1) Let $x = (x^1, x^2, \ldots, x^n)$ be a point in D^n such that

$$x^1 \leqq x^2 \leqq \ldots \leqq x^n.$$

Since the set P of bistochastic matrices in $\mathbf{R}^{n^2}$ is compact and $f(\mathbf{P}x)$ is a continuous function in the coefficients p^i_j and therefore in $\mathbf{P}$, it attains its maximum on P for some bistochastic matrix $\mathbf{Q}$. Permuting the rows of $\mathbf{Q}$ if necessary, we can suppose that $(y^1, y^2, \ldots, y^n) = \mathbf{Q}x$ satisfies

$$y^1 \leqq y^2 \leqq \ldots \leqq y^n.$$

In order to show that f is strictly S-convex, it is sufficient to show that the matrix $\mathbf{Q}$ satisfies

$$y = \mathbf{Q}x = x.$$

(2) We can suppose that the x^i are not all equal to

$$\alpha = \frac{x^1+x^2+\ldots+x^n}{n} = \frac{y^1+y^2+\ldots+y^n}{n}$$

for otherwise we should have $y^n \geqq y^{n-1} \geqq \ldots \geqq y^1 \geqq x^1 = \alpha$ (theorem of Hardy, Littlewood and Polya), whence $y^1 = y^2 = \ldots = y^n = \alpha$ and the proposition would then be proved.

Suppose therefore that

$$x^1 = x^2 = \ldots = x^k < x^{k+1} \leqq x^{k+2} \leqq \ldots \leqq x^n.$$

If $i \leqq k$ and $j > k$, then $x^j - x^i > 0$. Given $\varepsilon > 0$ and $l < m$, we define z to be the point of coordinates

$$\begin{cases} z^l = y^l - \varepsilon(x^j - x^i), \\ z^m = y^m + \varepsilon(x^j - x^i), \\ z^i = y^i \quad \text{if} \quad i \neq l, m. \end{cases}$$

The function defined by

$$\phi(\varepsilon) = f(y^1, y^2, \ldots, y^l - \varepsilon(x^j - x^i), \ldots, y^m + \varepsilon(x^j - x^i), \ldots, y^n)$$

is differentiable and its derivative is

$$\phi'(\varepsilon) = (x^j - x^i)[f'_m(z) - f'_l(z)].$$

Since $z^m - z^l = y^m - y^l + 2\varepsilon(x^j - x^i) > 0$, we have

$$\phi'(\varepsilon) = \frac{x^j - x^i}{z^m - z^l}(z^m - z^l)\,[f'_m(z) - f'_l(z)] > 0,$$

whenever ε is chosen to be sufficiently small for z to belong to D^n. Hence $\phi(\varepsilon) > \phi(0)$ for ε sufficiently small and so

$$f(z^1, z^2, \ldots, z^n) > f(y^1, y^2, \ldots, y^n).$$

(3) Put

$$\begin{aligned} p_i^l &= q_i^l + \varepsilon, & p_j^l &= q_j^l - \varepsilon, \\ p_i^m &= q_i^m - \varepsilon, & p_j^m &= q_j^m + \varepsilon, \\ p_s^r &= q_s^r \quad \text{if} \quad r \neq l, m; & s &\neq i, j. \end{aligned}$$

We have $f(z) = f(\mathbf{P}x) > f(\mathbf{Q}x)$; since $\mathbf{Q}$ is a bistochastic matrix such that $f(\mathbf{Q}x)$ has the maximum value, the matrix $\mathbf{P}$ is not bistochastic, however small ε may be; therefore we have

$$q_j^l\, q_i^m = 0.$$

This is valid for all $i \leqq k$, all $j > k$, all l and all $m > l$. Therefore, from the lemma,

$$\mathbf{Q}_\mathrm{N}^\mathrm{N} = \mathbf{Q}_\mathrm{K}^\mathrm{K} \otimes \mathbf{Q}_{\mathrm{N-K}}^{\mathrm{N-K}}.$$

(4) The result just proved shows that, in particular, if a function f in $\mathbf{R}^2$ satisfies the conditions of the theorem, then $f(\mathbf{Q}x)$ attains the maximum value when

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We then have $y = \mathbf{Q}x = x$ and therefore the result is true in $\mathbf{R}^2$.

Suppose then that the result is true for integers smaller than n. We shall prove that this implies that it is true for the integer n and so the theorem will be established by induction. Put

$$\alpha = x^1 = x^2 = \dots = x^k < x^{k+1} \leqq x^{k+2} \leqq \dots \leqq x^n.$$

Since $y = \mathbf{Q}_N^N x = (\mathbf{Q}_K^K \otimes \mathbf{Q}_{N-K}^{N-K})x$, we also have

$$y^1 = y^2 = \dots = y^k = \alpha.$$

The function given by $g(\mathbf{X}^{N-K}) = f(\alpha, \alpha, \dots, \alpha, x^{k+1}, x^{k+2}, \dots, x^n)$ satisfies the conditions of the theorem and $g(\mathbf{P}_{N-K}^{N-K}\mathbf{X}^{N-K})$ attains its maximum value with the bistochastic matrix $\mathbf{Q}_{N-K}^{N-K}$; then, by hypothesis,

$$\mathbf{Q}_{N-K}^{N-K}\mathbf{X}^{N-K} = \mathbf{X}^{N-K}.$$

Hence we have

$$y = (\mathbf{X}^K, \mathbf{X}^{N-K}) = x.$$

APPLICATION 1. *If* $\lambda > 1$, *the function* f *such that*

$$f(x) = (x^1)^\lambda + (x^2)^\lambda + \dots + (x^n)^\lambda$$

is strictly S-convex in $]0, +\infty[^n$.

In fact, if $0 < x^1 < x^2$, we have $0 < (x^1)^{\lambda-1} < (x^2)^{\lambda-1}$ and so

$$(x^2 - x^1)\left(\frac{\partial f}{\partial x^2} - \frac{\partial f}{\partial x^1}\right) = \lambda(x^2 - x^1)\,[(x^2)^{\lambda-1} - (x^1)^{\lambda-1}] > 0.$$

APPLICATION 2. *Consider the symmetric functions:*

$$c_1(x) = \sum_i x^i$$
$$c_2(x) = \sum_{i<j} x^i x^j$$
$$c_3(x) = \sum_{i<j<k} x^i x^j x^k$$
$$\dots\dots\dots\dots$$
$$c_n(x) = x^1 x^2 \dots x^n$$

The functions c_i *and* $\dfrac{c_i}{c_{i-1}}$ *are strictly S-concave and increasing in* $]0, +\infty[^n$.

For example, we can prove that c_4 is strictly S-concave as follows. Put

$$c_k'' = c_k(x^3, x^4, \dots, x^n).$$

Then

$$c_4 = x^1 x^2 c_2'' + (x^1 + x^2)c_3'' + c_4''$$

and so, if $x^1 \neq x^2$, we have

$$(x^2-x^1)\left(\frac{\partial c_4}{\partial x^2}-\frac{\partial c_4}{\partial x^1}\right) = (x^2-x^1)\,[x^1c_2''-x^2c_2''+c_3''-c_3'']$$

$$= -(x^2-x^1)^2\, c_2''(x) < 0.$$

Therefore c_4 is strictly S-concave.

Theorem 6. *If D is an open interval of* $\mathbf{R}$, *a necessary and sufficient condition for a differentiable and symmetric function f to be S-convex in D^n is that, for all x in D^n,*

$$(x^2-x^1)\left(\frac{\partial f}{\partial x^2}-\frac{\partial f}{\partial x^1}\right) \geqq 0.$$

Proof. Let f be a differentiable S-convex function in D^n. Let

$$x = (x^1, x^2, \ldots, x^n) \in D^n,$$

and let ε be such that $0 < \varepsilon < 1$. Let $y(\varepsilon) = (y^1, y^2, \ldots, y^n)$ be the point such that

$$y^1 = (1-\varepsilon)x^1+\varepsilon x^2,$$
$$y^2 = \varepsilon x^1+(1-\varepsilon)x^2,$$
$$y^i = x^i \text{ if } i \neq 1, 2.$$

Since y is the result of transforming x by means of a bistochastic matrix, we have

$$\Delta f = f[y(\varepsilon)]-f(x) \leqq 0.$$

Also, since f is differentiable, we can write

$$\Delta f = \frac{\partial f}{\partial x^1}\,(y^1-x^1)+\frac{\partial f}{\partial x^2}\,(y^2-x^2)+\beta(x, \varepsilon)\,\|\, y(\varepsilon)-x \,\|.$$

Then

$$\frac{\Delta f}{\varepsilon} = \frac{\partial f}{\partial x^1}\,(x^2-x^1)+\frac{\partial f}{\partial x^2}\,(x^1-x^2)+\beta(x, \varepsilon)\,\frac{\|\, y(\varepsilon)-x \,\|}{\varepsilon} \leqq 0.$$

Let ε tend to 0; we get

$$(x^2-x^1)\left(\frac{\partial f}{\partial x^2}-\frac{\partial f}{\partial x^1}\right) \geqq 0,$$

as required.

Suppose conversely that f is a symmetric differentiable function satisfying

$$(x^2-x^1)\left(\frac{\partial f}{\partial x^2}-\frac{\partial f}{\partial x^1}\right) \geqq 0.$$

If $\varepsilon > 0$, the function given by

$$g(x) = f(x) + \varepsilon[(x^1)^2 + (x^2)^2 + \ldots + (x^n)^2]$$

is differentiable, symmetric and, for $x^1 \neq x^2$, satisfies

$$(x^2 - x^1)\left(\frac{\partial g}{\partial x^2} - \frac{\partial g}{\partial x^1}\right) = (x^2 - x^1)\left(\frac{\partial f}{\partial x^2} - \frac{\partial f}{\partial x^1}\right) + (x^2 - x^1)\, 2\varepsilon(x^2 - x^1) > 0.$$

Then if $y = \mathbf{P}x$, we have $g(y) \leqq g(x)$ by Theorem 5; hence

$$f(y) + \varepsilon[(y^1)^2 + (y^2)^2 + \ldots + (y^n)^2] \leqq f(x) + \varepsilon[(x^1)^2 + (x^2)^2 + \ldots + (x^n)^2].$$

Let ε tend to 0; we get

$$f(y) \leqq f(x)$$

and so f is S-convex.

Theorem 7. *Let $x = (x^1, x^2, \ldots, x^n)$ and $y = (y^1, y^2, \ldots, y^n)$ be two points of D^n, such that $x^1 \leqq x^2 \leqq \ldots \leqq x^n$, $y^1 \leqq y^2 \leqq \ldots \leqq y^n$ and $y \geqq \mathbf{P}x$ for some bistochastic matrix $\mathbf{P}$. Then, if f is an increasing S-concave function in D^n, if $\mathrm{K} = \{1, 2, \ldots, k\} \subset \mathrm{N}$ and if*

$$a = (a^1, a^2, \ldots, a^n) \in D^n,$$

we have

$$f(\mathrm{Y}^{\mathrm{K}}, \mathbf{A}^{\mathrm{N-K}}) \geqq f(\mathbf{X}^{\mathrm{K}}, \mathbf{A}^{\mathrm{N-K}}).$$

Proof. If $y \geqq \mathbf{P}x$, then, by the corollary to the theorem of Hardy, Littlewood and Polya,

$$\begin{cases} x^1 \leqq y^1, \\ x^1 + x^2 \leqq y^1 + y^2, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ x^1 + x^2 + \ldots + x^n \leqq y^1 + y^2 + \ldots + y^n. \end{cases}$$

Therefore there exists a bistochastic matrix $\mathbf{Q}_{\mathrm{K}}^{\mathrm{K}}$ of order k such that

$$\mathrm{Y}^{\mathrm{K}} \geqq \mathbf{Q}_{\mathrm{K}}^{\mathrm{K}} \mathbf{X}^{\mathrm{K}},$$

whence, since f is increasing,

$$\begin{aligned} f(\mathrm{Y}^{\mathrm{K}}, \mathbf{A}^{\mathrm{N-K}}) \geqq f(\mathbf{Q}_{\mathrm{K}}^{\mathrm{K}} \mathbf{X}^{\mathrm{K}}, \mathbf{A}^{\mathrm{N-K}}) &= f[(\mathbf{Q}_{\mathrm{K}}^{\mathrm{K}} \otimes \mathbf{E}_{\mathrm{N-K}}^{\mathrm{N-K}})(\mathbf{X}^{\mathrm{K}}, \mathbf{A}^{\mathrm{N-K}})] \\ &\geqq f(\mathbf{X}^{\mathrm{K}}, \mathbf{A}^{\mathrm{N-K}}). \end{aligned}$$

Thus the theorem is proved.

APPLICATION. *If $x \in \mathbf{R}^n$ is such that $x^1 > 0, x^2 > 0, \ldots x^n > 0$ and if $y \geqq \mathbf{P}x$ for some bistochastic matrix $\mathbf{P}$, there exists a bistochastic matrix $\mathbf{Q}$ such that*

$$(\log y^1, \log y^2, \ldots, \log y^n) \geqq \mathbf{Q}\,(\log x^1, \log x^2, \ldots, \log x^n).$$

Proof. We can always suppose that

$$x^1 \leqq x^2 \leqq \ldots \leqq x^n,$$
$$y^1 \leqq y^2 \leqq \ldots \leqq y^n.$$

Applying Theorem 7 to the fundamental symmetric function c_n given by $c_n(x) = x^1 x^2 \ldots x^n$, which is S-concave and increasing, and taking $a = (1, 1, \ldots, 1)$, we get

$$\begin{cases} y^1 \geqq x^1, \\ y^1\, y^2 \geqq x^1\, x^2, \\ \ldots\ldots\ldots\ldots \\ y^1\, y^2 \ldots y^n \geqq x^1\, x^2 \ldots x^n. \end{cases}$$

Hence

$$\log y^1 \geqq \log x^1,$$
$$\log y^1 + \log y^2 \geqq \log x^1 + \log x^2,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\log y^1 + \log y^2 + \ldots + \log y^n \geqq \log x^1 + \log x^2 + \ldots + \log x^n.$$

Therefore, by the corollary to the theorem of Hardy, Littlewood and Polya, there exists a bistochastic matrix $\mathbf{Q}$ such that

$$(\log y^1, \log y^2, \ldots, \log y^n) \geqq \mathbf{Q}\,(\log x^1, \log x^2, \ldots, \log x^n)$$

Let $(x^1, x^2, \ldots, x^n)$ and $(y^1, y^2, \ldots, y^n)$ be two n-tuples whose components are arranged in increasing order and which are such that $y \geqq \mathbf{P}x$ for some bistochastic matrix $\mathbf{P}$. Then Theorem 7 enables us to obtain a large number of inequalities. In particular, this situation occurs in the spectral theory of matrices (for example, with n-tuples of latent roots of an Hermitian matrix and the coefficients of the leading diagonal).

§ 12. Extremal problems with convex and concave functions

A problem which often occurs in questions of economics is the following:

MAXIMUM PROBLEM. Given concave functions $f, g_1, g_2, \ldots, g_n$ defined in $\mathbf{R}^m$, find a point $x \in \mathbf{R}^m$ such that

(1) $g_j(x) \geqq 0 \qquad (j = 1, 2, \ldots, n)$,
(2) $f(x)$ is maximal with respect to these constraints.

We shall show that this problem can be reduced to another which is much easier to solve; the method used is a generalisation of the well-known method of 'Lagrange multipliers' and is due to Kuhn and Tucker.

Let $y = (y_0, y_1, y_2, \ldots, y_n)$ be a variable point of $\mathbf{R}^{n+1}$; we associate with the above maximum problem a function, called the **Lagrange function**, given by

$$F(x, y) = y_0 f(x) + \sum_{j=1}^{n} y_j g_j(x).$$

If we write $\bar{y} = (y_1, y_2, \ldots, y_n)$ and $\bar{g}(x) = (g_1(x), g_2(x), \ldots, g_n(x))$, we can also write the Lagrange function in the form

$$F(x, y) = y_0 f(x) + \langle \bar{y}, \bar{g}(x) \rangle.$$

LAGRANGE PROBLEM. Find an $x \in \mathbf{R}^m$ and a $y \in \mathbf{R}^{n+1}$ such that

(1) $F(\xi, y)$ is maximal in $\mathbf{R}^m$ for $\xi = x$,

(2) $\bar{g}(x) \geqq 0,\ y \geqq 0,\ \sum_{j=0}^{n} y_j = 1$,

(3) $\langle \bar{y}, \bar{g}(x) \rangle = 0$.

Whenever the functions under consideration are differentiable, the Lagrange problem reduces to a system of inequalities which we know we can solve, since we can replace (1) by

$$(1') \qquad y_0 \frac{\partial f}{\partial x_i} + \left\langle \frac{\partial \bar{g}}{\partial x_i}, \bar{y} \right\rangle = 0 \qquad (i = 1, 2, \ldots, m).$$

The fundamental result is given in the following theorem:

Theorem of Kuhn and Tucker. *Let $f, g_1, g_2, \ldots, g_n$ be concave functions in $\mathbf{R}^m$ (differentiable or otherwise); for each solution $x \in \mathbf{R}^m$ of the maximum problem, there exists a $y \in \mathbf{R}^{n+1}$ such that (x, y) is a solution of the Lagrange problem; for each solution (x, y) of the Lagrange problem with $y_0 \neq 0$, the point x is a solution of the maximum problem.*

Proof. (1) Let x be a solution of the maximum problem. By hypothesis, the system

$$\begin{aligned} &g_j(\xi) \geqq 0 \qquad (j = 1, 2, \ldots, n) \\ &f(\xi) > f(x) \end{aligned}$$

does not admit a solution $\xi \in \mathbf{R}^m$. Therefore, by the first fundamental theorem (§ 7) there exist coefficients $y_0, y_1, y_2 \ldots y_n \geqq 0$ with sum 1, such that

$$y_0[f(\xi) - f(x)] + \sum_{i=1}^{n} y_i g_i(\xi) \leqq 0 \qquad (\xi \in \mathbf{R}^m),$$

or

$$y_0 f(\xi) + \langle \bar{y}, \bar{g}(\xi) \rangle \leqq y_0 f(x) \qquad (\xi \in \mathbf{R}^m).$$

Putting $\xi = x$, we get $\langle \bar{y}, \bar{g}(x) \rangle \leqq 0$. Hence, since the opposite inequality is also satisfied,

$$\langle \bar{y}, \bar{g}(x) \rangle = 0.$$

Therefore

$$y_0 f(\xi) + \langle \bar{y}, \bar{g}(\xi) \rangle \leqq y_0 f(x) + \langle \bar{y}, \bar{g}(x) \rangle \qquad (\xi \in \mathbf{R}^m).$$

Thus we have shown that the Lagrange function $y_0f(\xi)+\langle\bar{y}, \bar{g}(\xi)\rangle$ is maximal in $\mathbf{R}^m$ for $\xi = x$ and the first part of the theorem is established.

We now suppose that (x, y) is a solution of the Lagrange problem and show that x is a solution of the maximum problem. To do this, we simply reverse the argument given above to prove the first part of the theorem. By hypothesis, we have

$$y_0 f(\xi) + \langle \bar{y}, \bar{g}(\xi)\rangle \leqq y_0 f(x) \quad (\xi \in \mathbf{R}^m)$$

or

$$y_0[f(\xi)-f(x)] + \sum y_j g_j(\xi) \leqq 0 \quad (\xi \in \mathbf{R}^m).$$

Since $y_0 > 0$, the system

$$\begin{cases} g_j(\xi) \geqq 0, \\ f(\xi) > f(x) \end{cases}$$

does not admit a solution $\xi \in \mathbf{R}^m$ and $f(x)$ is a solution of the maximum problem.

We now consider differentiable, but not necessarily concave, functions $f, g_1, g_2, \ldots, g_n$ defined in $\mathbf{R}^m$. Put

$$\bar{g}(x) = (g_1(x), g_2(x), \ldots, g_n(x))$$

$$\frac{\partial \bar{g}}{\partial x_i} = \left(\frac{\partial g_1}{\partial x_i}, \frac{\partial g_2}{\partial x_i}, \ldots, \frac{\partial g_n}{\partial x_i}\right).$$

The gradient of f is denoted by

$$\delta f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right).$$

The global maximum problem consists of finding a point x in the set $G = \{x \,/\, g_j(x) \geqq 0 \text{ for } j = 1, 2, \ldots, n\}$ such that $f(x)$ is maximal. Such a point x will be called a **global maximum** of f.

A point x will be called a **relative maximum** of the function f if the value $f(x)$ decreases when we replace x by $x' = x+\lambda u$, where λ is sufficiently small; this must be verified for each vector u, unless there exists a j such that $g_j(x) = 0$ and $\langle \delta g_j, u\rangle < 0$ (for then the vector u is directed to the exterior of the set G).

In general, the relative maxima are easier to determine than the global maxima and nearly always include them. However, it can happen that a global maximum is not a relative maximum, but is a singular point with respect to the set G. For example, consider the global maximum of $f(x_1) = x_1$, in $\mathbf{R}^2$, with the constraints

$$g_1(x) = (1-x_1)^3 - x_2 \geqq 0,$$
$$g_2(x) = x_1 \geqq 0,$$
$$g_3(x) = x_2 \geqq 0.$$

The set G is indicated in figure 46. The point (1, 0) is the global maximum.

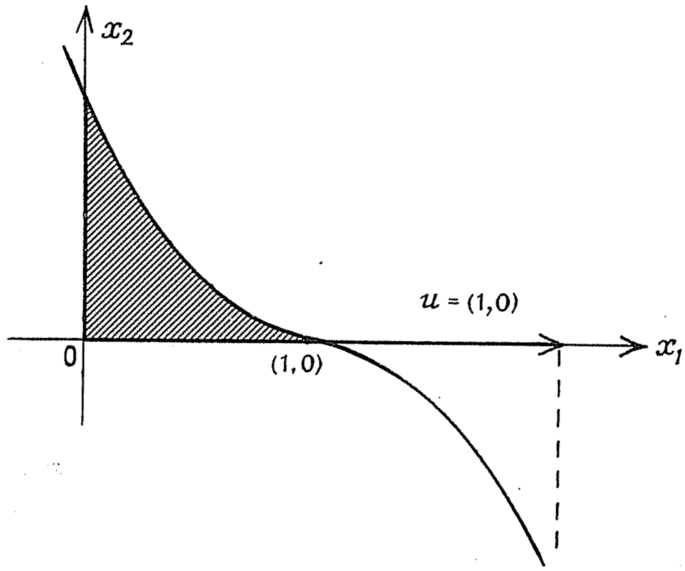


FIG. 46

The vector $u = (1, 0)$ satisfies

$$\langle u, \delta g_1 \rangle = 0, \quad \langle u, \delta g_2 \rangle = 1,$$
$$\langle u, \delta g_3 \rangle = 0.$$

However, $f(x)$ does not increase in the direction of u; the point (1, 0) is not a relative minimum.

We now compare the following two problems.

RELATIVE MAXIMUM PROBLEM. Let $f, g_1, g_2, \ldots, g_n$ be differentiable functions. To find an $x = (x_1, x_2, \ldots, x_m)$ such that

(1) $g(x) = (g_1(x), g_2(x), \ldots, g_n(x)) \geqq 0,$
(2) f has a relative maximum at x.

LAGRANGE PROBLEM. Let $f, g_1, g_2, \ldots, g_n$ be differentiable functions. To find an $x = (x_1, x_2, \ldots, x_m)$ and a $y = (y_1, y_2, \ldots, y_n)$ such that

$$\begin{cases} g(x) \geqq 0, \quad y \geqq 0, \\ \langle g, y \rangle = 0 \\ \dfrac{\partial f}{\partial x_i} + \left\langle \dfrac{\partial g}{\partial x_i}, y \right\rangle = 0 \qquad (i = 1, 2, \ldots, m). \end{cases}$$

Theorem. *If x is a relative maximum, there exists a $\bar{y} = (y_1, y_2, \ldots, y_n)$ such that $(x, \bar{y})$ is a solution of the Lagrange problem.*

Proof. Let $\mathbf{J} \subset \{1, 2, \ldots, n\}$ be such that

$$\begin{cases} g_j(x) = 0 & (j \in \mathbf{J}), \\ g_j(x) > 0 & (j \notin \mathbf{J}). \end{cases}$$

If the point x is a relative maximum we have $\langle \delta f, \Delta x \rangle \leqq 0$ for each increase Δx such that

$$\langle \delta g_j(x), \Delta x \rangle \geqq 0 \quad (j \in \mathbf{J}).$$

By Farkas' corollary (§ 1), there exist numbers $p_j \geqq 0$ such that

$$-\delta f = \sum_{j \in J} p_j \delta g_j(x).$$

Consider the vector $y = (y_1, y_2, \ldots, y_n)$ such that

$$y_j = \begin{cases} p_j \text{ if } j \in \mathbf{J}, \\ 0 \text{ if } j \notin \mathbf{J}. \end{cases}$$

We then have

$$-f'_{x_i} = \left\langle \bar{y}, \frac{\partial \bar{g}}{\partial x_i} \right\rangle \qquad (i = 1, 2, \ldots, m).$$

On the other hand, we have

$$y_j > 0 \;\Rightarrow\; j \in \mathbf{J} \;\Rightarrow\; g_j(x) = 0,$$

whence

$$\langle \bar{y}, \bar{g}(x) \rangle = 0.$$

Thus we have recovered the conditions of Lagrange's problem.

CHAPTER IX

TOPOLOGICAL VECTOR SPACES

§ 1. Normed spaces

Let X be a vector space, on which is defined a numerical function $x \to \| x \|$ such that

$$\begin{aligned} &(1) \quad \| x \| \geqq 0, \\ &(2) \quad \| x \| = 0 \quad \Leftrightarrow \quad x = 0, \\ &(3) \quad \| \lambda x \| = | \lambda | \cdot \| x \| \ (\text{if } \lambda \in \mathbf{R}), \\ &(4) \quad \| x_1 + x_2 \| \leqq \| x_1 \| + \| x_2 \|. \end{aligned}$$

The function $x \to \| x \|$ is called a **norm** (cf. § 5, Chapter VII) and the pair consisting of X and this norm is called a **normed space.** In a normed space X, the distance from a point x to a point y is $d(x, y) = \| x - y \|$; the function d is a metric, because

$$\begin{aligned} &d(x, y) \geqq 0, \\ &d(x, y) = 0 \quad \Leftrightarrow \quad x = y, \\ &d(x, y) = d(y, x), \\ &d(x, z) \leqq d(x, y) + d(y, z). \end{aligned}$$

A normed space is thus a metric space and is therefore a topological space.

We observe that condition (1) for a norm follows from conditions (3) and (4), because

$$0 = | 0 | \cdot \| a \| = \| 0 \cdot a \| = \| x - x \| \leqq \| x \| + \| -x \| = 2 \| x \|.$$

EXAMPLE 1. The space $\mathbf{R}^n$ is normed, with

$$\| x \| = (| x^1 |^2 + | x^2 |^2 + \ldots + | x^n |^2)^{\frac{1}{2}}.$$

EXAMPLE 2. The space L_p consisting of the sequences (x^n) such that $\sum\limits_{n=1}^{\infty} | x^n |^p < +\infty$ is normed, with

$$\| x \| = \Big(\sum_{n=1}^{\infty} | x^n |^p \Big)^{\frac{1}{p}}.$$

To prove this, we use Minkowski's inequality (§ 9, Chapter VIII).

EXAMPLE 3. The space $\mathscr{L}_p$ consisting of functions defined on [0, 1] such that the Lebesgue integral

$$\int_0^1 |\,\phi(t)\,|^p \,\mathrm{d}t$$

exists, is normed, with

$$\|\,\phi\,\| = \left(\int_0^1 |\,\phi(t)\,|^p \,\mathrm{d}t\right)^{\frac{1}{p}}$$

provided that we follow the convention of regarding two functions as being equal if they take the same values on [0, 1] except for a set of measure zero (for if $\|\,\phi\,\| = 0$, then $\phi(t)$ is not necessarily zero for all t in [0, 1]; however, the values of t for which $\phi(t) \neq 0$ form a set of measure zero).

Theorem 1. *In a normed space X, the single-valued mapping σ of $X \times X$ into X defined by $\sigma(x, y) = x+y$ is continuous; the single-valued mapping τ of $\mathbf{R} \times X$ into X defined by $\tau(\lambda, x) = \lambda x$ is continuous.*

Proof. (1) If $(x_0, y_0) \in X \times X$ and if $\varepsilon > 0$, then

$$\max \{\,\|\,x-x_0\,\|,\ \|\,y-y_0\,\|\,\} \leqq \varepsilon$$

implies that

$$\begin{aligned}\|\,(x+y)-(x_0+y_0)\,\| &= \|\,(x-x_0)+(y-y_0)\,\| \\ &\leqq \|\,x-x_0\,\| + \|\,y-y_0\,\| \leqq 2\,\varepsilon.\end{aligned}$$

Therefore σ is continuous at (x_0, y_0) and so, since this point is arbitrary, σ is continuous in $X \times X$.

(2) If $(\lambda_0, x_0) \in \mathbf{R} \times X$ and if $\varepsilon > 0$, then

$$\max \{\,|\,\lambda-\lambda_0\,|,\ \|\,x-x_0\,\|\,\} \leqq \varepsilon$$

implies that

$$\begin{aligned}\|\,\lambda x-\lambda_0 x_0\,\| = \|\,\lambda x-\lambda x_0+\lambda x_0-\lambda_0 x_0\,\| &\leqq \|\,\lambda(x-x_0)\,\| + |\,\lambda-\lambda_0\,| \cdot \|\,x_0\,\| \\ &\leqq |\,\lambda\,|\,\varepsilon+\varepsilon\,\|\,x_0\,\| \leqq (\,|\,\lambda_0\,| + \varepsilon + \|\,x_0\,\|\,)\,\varepsilon.\end{aligned}$$

Therefore τ is continuous at (λ_0, x_0) and so, since this point is arbitrary, τ is continuous in $\mathbf{R} \times X$.

Theorem 2. *If f is a single-valued linear mapping of a normed space X into a normed space Y, then f is continuous if and only if $\|\,f(x)\,\|$ is bounded in the unit ball B of X: that is,*

$$\sup_{x \in B} \|\,f(x)\,\| = \alpha < +\infty.$$

Proof. If $\|\,f(x)\,\|$ is bounded in B by α, then $f(x)$ is continuous at any point x_0, for

$$\| x-x_0 \| \leqq \varepsilon \quad \Rightarrow \quad \frac{x-x_0}{\varepsilon} \in B \quad \Rightarrow \quad \left\| \frac{f(x)-f(x_0)}{\varepsilon} \right\| \leqq \alpha$$
$$\Rightarrow \quad \| f(x)-f(x_0) \| \leqq \alpha\varepsilon.$$

If $\| f(x) \|$ is not bounded in B, there exist points $x_1, x_2, \ldots, x_n, \ldots$ such that $\| f(x_n) \| > n$ for each n; since $\left(\left\| \frac{x_n}{n} \right\| \right) \to 0$, the sequence $\left(\frac{1}{n} x_n \right)$ converges to 0; but we have

$$\left\| f\left(\frac{1}{n} x_n \right) \right\| = \frac{1}{n} \| f(x_n) \| > 1$$

and therefore the sequence $\left(f\left(\frac{1}{n} x_n \right) \right)$ does not converge to $f(0) = 0$, so that f is not continuous at the point 0.

COROLLARY. *A linear mapping f of X into Y is continuous if and only if there exists a number α such that*

$$(\forall x) : \| f(x) \| \leqq \alpha \| x \|.$$

Proof. If f is continuous, then $\sup_{x \in B} \| f(x) \| = \alpha < +\infty$, by the theorem; since $\frac{x}{\| x \|} \in B$, we have

$$\| f(x) \| \leqq \alpha \| x \|.$$

Conversely, if the condition is satisfied, we have

$$\| x \| \leqq 1 \quad \Rightarrow \quad \| f(x) \| \leqq \alpha \| x \| \leqq \alpha.$$

Theorem 3. *Let G be an open convex set and let f be a convex function in G. Then f is continuous in G if and only if it is bounded above in a ball $B(a) \subset G$.*

Proof. If f is continuous in G and $a \in G$, then, to each $\varepsilon > 0$ there corresponds a number η such that

$$x \in B_\eta(a) \quad \Rightarrow \quad f(x) \leqq f(a)+\varepsilon.$$

Therefore $f(x)$ is bounded above in $B_\eta(a)$.

Conversely, if $f(x)$ is bounded above in a ball $B_\lambda(a)$ then we can see, exactly as in the proof of Theorem 7, § 5, Chapter VIII, that $f(x)$ is continuous in $B_\lambda(a)$; we shall prove that f is continuous at an arbitrary point x_0 of G. Let $x_1 \in G$ be such $x_0 \in]a, x_1[$. There is no loss in generality in assuming that $x_1 = 0$ and that $f(0) = 0$, for we can always make a translation and also replace $f(x)$ by $f(x)-f(0)$.

Let y_0 be a point of G such that

$$\| x_0-y_0 \| \leqq \lambda \frac{\| x_0 \|}{\| a \|}.$$

Put

$$y = \frac{\| a \|}{\| x_0 \|} y_0.$$

Then

$$\| y-a \| = \left\| y_0 \frac{\| a \|}{\| x_0 \|} - x_0 \frac{\| a \|}{\| x_0 \|} \right\| = \frac{\| a \|}{\| x_0 \|} \| y_0 - x_0 \| \leqq \lambda,$$

and therefore $y \in B_\lambda(a)$. Consequently

$$f(y_0) = f\left[\frac{\| x_0 \|}{\| a \|} y + \left(1 - \frac{\| x_0 \|}{\| a \|}\right) 0\right] \leqq \frac{\| x_0 \|}{\| a \|} f(y).$$

Therefore $f(y_0)$ is bounded above in a ball of centre x_0 and so f is continuous at x_0.

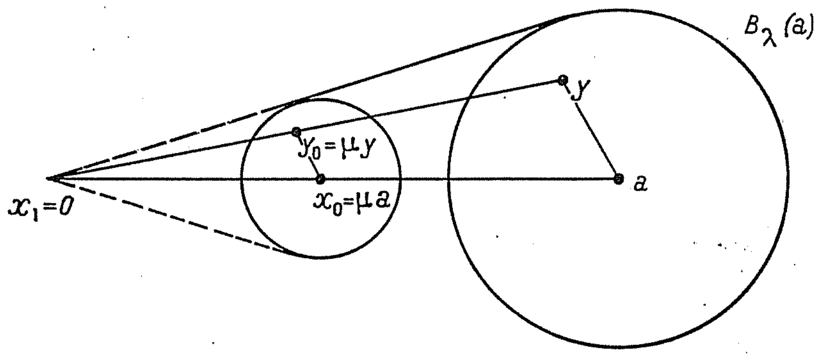


FIG. 47

COROLLARY. *If f is a convex function in an open convex set G and if f is upper semi-continuous at a point of G, then it is continuous in G.*

Proof. This is immediate, since, given $\varepsilon > 0$, there exists η such that

$$\| x - x_0 \| \leqq \eta \quad \Rightarrow \quad f(x) \leqq f(x_0) + \varepsilon.$$

The set of continuous numerical linear functions on X is called the **dual space** of X and is denoted by X'. If we write $[f_1+f_2](x) = f_1(x)+f_2(x)$ and $[\lambda f](x) = \lambda \cdot f(x)$, then X' becomes a vector space; the function which takes the value 0 at each point of X is the neutral element of X'. If X is normed and we write

$$\| f \| = \sup_{x \in B} | f(x) |,$$

where B is the unit ball in X, then the space X' is also normed. To prove this, we verify the following properties:

(1) $\|f\| \geqq 0$,

(2) $\|f\| = 0 \quad \Leftrightarrow \quad f = 0$,

(3) $\|\lambda f\| = \sup\limits_{x\in B} |\lambda f(x)| = |\lambda| \cdot \|f\|$,

(4) $\|f+g\| = \sup\limits_{x\in B} |f(x)+g(x)| \leqq \sup\limits_{x\in B} |f(x)| + \sup\limits_{x\in B} |g(x)| = \|f\|+\|g\|$.

REMARK. It is sometimes convenient to denote the number $f(x)$ by $\langle f, x\rangle$; we then have all the properties of linearity of the scalar product already proved in the space $\mathbf{R}^n$:

$$\text{(1)}\quad \langle \lambda f+\mu g, x\rangle = \lambda\langle f, x\rangle+\mu\langle g, x\rangle,$$
$$\text{(2)}\quad \langle f, \lambda x+\mu y\rangle = \lambda\langle f, x\rangle+\mu\langle f, y\rangle$$

and we also have the following inequality, analogous to the Cauchy-Schwartz inequality (§ 1, Chapter VIII):

$$\text{(3)}\quad |\langle f, x\rangle| \leqq \|f\|\cdot\|x\|.$$

The proof of this formula is immediate, since

$$\frac{|f(x)|}{\|x\|} = \left|f\left(\frac{x}{\|x\|}\right)\right| \leqq \|f\|.$$

Theorem 4. *If, in X', the sequence (f_n) converges to f_0 and the sequence (g_n) converges to g_0, we have*

$$(f_n+g_n) \to f_0+g_0.$$

Proof. If $\varepsilon > 0$, then, for n sufficiently large, we have $\|f_n-f_0\| \leqq \frac{\varepsilon}{2}$, $\|g_n-g_0\| \leqq \frac{\varepsilon}{2}$, and hence

$$\|(f_n+g_n)-(f_0+g_0)\| = \|(f_n-f_0)+(g_n-g_0)\| \leqq \|f_n-f_0\| + \|g_n-g_0\| \leqq \varepsilon.$$

Theorem 5. *If $(x_n) \to x_0$ in X and $(f_n) \to f_0$ in X', then $(f_n(x_n)) \to f_0(x_0)$ in* $\mathbf{R}$.

Proof. Put $g_n = f_n-f_0$ and $y_n = x_n-x_0$; if $\varepsilon > 0$, then, for n sufficiently large, $\|g_n\| \leqq \varepsilon$, $\|y_n\| \leqq \varepsilon$ and therefore

$$\begin{aligned}|\langle f_n, x_n\rangle-\langle f_0, x_0\rangle| &= |\langle f_0+g_n, x_0+y_n\rangle-\langle f_0, x_0\rangle|\\ &= |\langle f_0, x_0\rangle+\langle f_0, y_n\rangle+\langle g_n, x_0\rangle+\langle g_n, y_n\rangle-\langle f_0, x_0\rangle|\\ &\leqq |\langle f_0, y_n\rangle| + |\langle g_n, x_0\rangle| + |\langle g_n, y_n\rangle| \leqq \|f_0\|\,\varepsilon+\|x_0\|\,\varepsilon+\varepsilon^2.\end{aligned}$$

Hence $\quad (\langle f_n, x_n\rangle) \to \langle f_0, x_0\rangle$.

Theorem 6. *In X', we have $(f_n) \to f_0$ if and only if the sequence $(f_n(x))$ converges uniformly to $f_0(x)$ in the unit ball B of X.*

Proof. If $(f_n) \to f_0$ in X', then, given $\varepsilon > 0$, there exists a number m such that

$$n \geqq m \quad \Rightarrow \quad \|f_n-f_0\| \leqq \varepsilon.$$

For all x in B, we have

$$|f_n(x)-f_0(x)| \leqq \|f_n-f_0\| \cdot \|x\| \leqq \varepsilon$$

whenever $n \geqq m$ and therefore $(f_n(x))$ converges uniformly in B in the sense defined in Example 3, § 2, Chapter IV.

Conversely, suppose that $(f_n(x)) \to f_0(x)$ uniformly in the unit ball B. Then, given $\varepsilon > 0$, there exists a number m such that

$$n \geqq m \quad \Rightarrow \quad \|f_n-f_0\| = \sup_{x \in B} |f_n(x)-f_0(x)| \leqq \varepsilon.$$

Therefore $(f_n) \to f_0$.

§ 2. Topological vector spaces

A **topological vector space**[1] is a vector space X together with a topology, such that the following conditions are satisfied:

(1) The single-valued mapping σ of $X \times X$ into X given by $\sigma(x, y) = x+y$, is continuous; in other words, for each neighbourhood $V(x_0+y_0)$, there exist neighbourhoods $U_1(x_0)$ and $U_2(y_0)$ such that

$$\begin{cases} x \in U_1(x_0) \\ y \in U_2(y_0) \end{cases} \quad \Rightarrow \quad x+y \in V(x_0+y_0).$$

(2) The single-valued mapping τ of $R \times X$ into X, given by $\tau(\lambda, x) = \lambda x$, is continuous; in other words, for each neighbourhood $V(\lambda_0 x_0)$, there exists a number η and a neighbourhood $U(x_0)$ such that

$$\begin{cases} |\lambda-\lambda_0| \leqq \eta \\ x \in U(x_0) \end{cases} \quad \Rightarrow \quad \lambda x \in V(\lambda_0 x_0).$$

EXAMPLE 1. A normed space X is a topological vector space with respect to the topology defined in § 1 (sometimes called the **strong topology** of X).

EXAMPLE 2. Let X be a normed space and let X' be its dual (that is, the set of continuous numerical linear functions on X). Let Φ be a finite subset of X'. Given $\varepsilon > 0$, write

$$N_\varepsilon^\Phi = \{x \,/\, x \in X, \ |f(x)| \leqq \varepsilon \text{ for all } f \in \Phi\}.$$

We can verify that, as Φ and ε vary, the sets of the form

$$N_\varepsilon^\Phi(x) = x+N_\varepsilon^\Phi$$

[1] The idea of a topological vector space was introduced by A. Kolmogoroff (*Studia Math.*, vol. 5, 1934, p. 29) in order to generalise the space $\mathbf{R}^n$, Hilbert space (used in quantum theory), Banach spaces (which are becoming increasingly important in probability theory), etc. In this chapter, we use some of the results of Chapter VIII, but more often we give independent proofs. (We make use of Zorn's theorem; the results in the previous chapter were obtained in a more elementary fashion.)

constitute a fundamental base of neighbourhoods for a topology in X, called the **weak topology** of X; and that X, together with the weak topology, is a topological vector space.

We observe that the mapping of X into X given by $f(x) = x+x_0$ and the inverse mapping such that $f^{-1}(x) = x-x_0$ are continuous; thus f is a homeomorphism. If U is an open neighbourhood of 0, the set $f(U) = U+x_0$ is an open set; moreover, since $x_0 \in U+x_0$, the set $U+x_0$ is an open neighbourhood of x_0.

Conversely, it is easily seen that each open neighbourhood of x_0 is of the form $U+x_0$. We shall denote open neighbourhoods of 0 by U or V and shall write $U(x_0)$ and $V(x_0)$ for $U+x_0$ and $V+x_0$ respectively.

In what follows (to § 4) we shall suppose that the space X satisfies the following axiom:

TOTALITY AXIOM: *for each point $x_0 \in X$ such that $x_0 \neq 0$, there exists a continuous linear function $f \in X'$ such that $f(x_0) \neq 0$; in other words, X' is a total family.*

The spaces which we meet in analysis satisfy this axiom; without it, we would not be able to develop a satisfactory theory. We observe that the totality axiom implies the separation axiom; if $x_0 \neq 0$, there exists a neighbourhood U of 0 and a neighbourhood $V(x_0)$ of x_0 such that U and $V(x_0)$ are disjoint. To prove this, it is sufficient to consider a function $f \in X'$ such that $f(x_0) = \alpha > 0$ and to write

$$U = \left\{x \mid x \in X; \qquad f(x) < \frac{\alpha}{2}\right\},$$

$$V(x_0) = \left\{x \mid x \in X; \qquad f(x) > \frac{\alpha}{2}\right\}.$$

PROPERTY 1. *Amongst the open neighbourhoods of 0 there exists a fundamental base of starred and symmetric open neighbourhoods of 0.*

A set A is **starred** (see § 7, Chapter I), if

$$a \in A, \quad \lambda \in [0, 1] \quad \Rightarrow \quad \lambda a \in A.$$

Further, A is **symmetric** with respect to 0 (see § 5, Chapter VII) if

$$a \in A \quad \Rightarrow \quad -a \in A.$$

Let V be a neighbourhood of 0; there exists a neighbourhood U of 0 and a number η such that

$$\begin{cases} x \in U \\ |\lambda| \leqq \eta \end{cases} \Rightarrow \quad \lambda x \in V.$$

We consider the set

$$U_V = \bigcup_{|\lambda| \leqq \eta} \lambda U.$$

(1) The set U_V is open. For, if $\lambda \neq 0$, the mapping given by $f(x) = \lambda x$ is a homeomorphism (cf. page 161) and therefore λU is an open set; hence U_V is a union of open sets.

(2) The set U_V is starred. For, if $x_0 \in U_V$, then there exists a number μ, satisfying $|\mu| \leqq \eta$, such that $x_0 \in \mu U$. Hence

$$\lambda \in [0, 1] \quad \Rightarrow \quad \begin{cases} \lambda x_0 \in (\lambda\mu)U \\ |\lambda\mu| \leqq \eta \end{cases} \quad \Rightarrow \quad \lambda x_0 \in U_V.$$

(3) The set U_V is symmetric—this is immediate.

(4) As V varies, the sets U_V form a fundamental base of neighbourhoods. For $U_V \subset V$, by the definition of U_V.

Property 2. *Every open neighbourhood of* 0 *has* 0 *as an internal point.*

Suppose that $x_0 \in X$ and that U is an open neighbourhood of 0. Since the mapping f given by $f(\lambda) = \lambda x_0$ is continuous at the point $\lambda = 0$, there exists a number η such that

$$|\lambda| \leqq \eta \quad \Rightarrow \quad \lambda x_0 = f(\lambda) \in U.$$

This shows that 0 is an internal point of the set U, according to the definition given above (§ 5, Chapter VII).

Property 3. *For each neighbourhood U, there exists a neighbourhood V such that $V+V \subset U$.*

Since the mapping σ such that $\sigma(x, y) = x+y$ is continuous, there exist neighbourhoods V_1 and V_2 such that

$$\begin{cases} x \in V_1 \\ y \in V_2 \end{cases} \quad \Rightarrow \quad x+y \in U.$$

If we put $V = V_1 \cap V_2$, we have $V+V \subset U$.

Property 4. *If U is a neighbourhood of* 0 *and $\lambda \neq 0$, the set λU is a neighbourhood of* 0.

This follows from the fact that $f(x) = \lambda x$ determines a homeomorphism.

Remark 1. Properties (1), (2), (3) and (4) are characteristic of a base of neighbourhoods of 0 for a topological vector space; in fact we can show that if these properties are satisfied for a family $(N_i \,/\, i \in J)$ of neighbourhoods of 0, the neighbourhoods $N_i + x = N_i(x)$ define a topology on X and, with this topology, X is a topological vector space.

REMARK 2. If U is a symmetric neighbourhood of 0, we have $U = -U$; because of Theorem 2, § 1, Chapter VII, we have

$$\begin{aligned} x \in U(y) &\Rightarrow \{x\} \cap (U+y) \neq \varnothing \Rightarrow \{0\} \cap (U+y-x) \neq \varnothing \\ &\Rightarrow \{-y\} \cap (U-x) \neq \varnothing \Rightarrow \{y\} \cap (U+x) \neq \varnothing \Rightarrow y \in U(x). \end{aligned}$$

Thus $x \in U(y)$ *is equivalent to* $y \in U(x)$.

Theorem 1. *Let* $\lambda \in \mathbf{R}$; *if* F *is a closed set, then* λF *is closed; if* G *is open and* $\lambda \neq 0$, *then* λG *is open; if* K *is compact, then* λK *is compact.*

Proof. If $\lambda \neq 0$, the mapping given by $f(x) = \lambda x$ is continuous and so is its inverse $f^{-1}(x) = (1/\lambda)x$. Therefore f is a homeomorphism and so $f(F)$, $f(G)$ and $f(K)$ are closed, open and compact respectively. In the case of F and K, the theorem is trivial if $\lambda = 0$.

Theorem 2. *If* A *is a subset of* X, *we have*

$$\bar{A} = \bigcap_{U} (U+A).$$

Proof. Let $\mathscr{B}$ be a fundamental base of symmetric neighbourhoods. We have

$$\begin{aligned} \bar{A} &= \{x \mid U(x) \cap A \neq \varnothing \text{ for all } U \in \mathscr{B}\} = \\ \ldots &= \{x \mid \{x\} \cap (A+U) \neq \varnothing \text{ for all } U \in \mathscr{B}\} \\ &= \bigcap_{U \in \mathscr{B}} (U+A) = \bigcap_{U} (U+A). \end{aligned}$$

Theorem 3. *The family formed by the closures of the open neighbourhoods of* 0 *is a fundamental base of neighbourhoods of* 0.

Proof. For each neighbourhood V, there exists a neighbourhood U such that

$$U+U \subset V.$$

Therefore, by the preceding theorem,

$$\bar{U} = \bigcap_{W} (W+U) \subset U+U \subset V.$$

Hence the sets $\bar{U}$ form a fundamental base of neighbourhoods of 0; in particular, *a topological vector space is regular* (see § 5, Chapter IV).

Theorem 4. *If* K *is a compact set and* F *is a closed set and if* $K \cap F = \varnothing$, *there exists a neighbourhood* U *such that*

$$(U+K) \cap (U+F) = \varnothing.$$

Proof. Let $x \in K$; we have

$$x \notin F = \bar{F} = \bigcap_{U} (U+F).$$

Therefore there exists a neighbourhood U_x such that

$$x \notin U_x+F.$$

Let V_x be a symmetric neighbourhood such that $V_x+V_x \subset U_x$; we have

$$\{x\} \cap (V_x+V_x+F) = \emptyset.$$

Therefore

$$(x+V_x) \cap (V_x+F) = \emptyset.$$

The sets $(x+V_x \,/\, x \in K)$ form an open covering of K; therefore there exists a finite open covering: $x_1+V_1, x_2+V_2, \ldots, x_n+V_n$. Writing $V = \bigcap_{i=1}^{n} V_i$, we have

$$K \cap (V+F) = \emptyset.$$

Let W be a symmetric neighbourhood such that $W+W \subset V$; we have

$$K \cap (W+W+F) = \emptyset$$

and so

$$(K+W) \cap (W+F) = \emptyset$$

as required.

Corollary. *If G is an open set and K is a compact set such that $G \supset K$, there exists a neighbourhood U such that*

$$U+K \subset G.$$

Proof. The set $-G$ is a closed set not meeting K, and so there exists a neighbourhood U such that

$$(K+U) \cap (-G+U) = \emptyset.$$

Hence we have

$$(K+U) \cap (-G) = \emptyset$$

and therefore

$$K+U \subset G.$$

Theorem 5. (i) *If G is an open set and A is any set, then $G+A$ is open;* (ii) *if F is a closed set and K is a compact set, then $F+K$ is a closed set:* (iii) *if K and K' are compact sets, then $K+K'$ is a compact set.*

Proof. (i) Since $G+a$ is open, so is

$$G+A = \bigcup_{a \in A} (G+a).$$

(ii) Suppose that $x_0 \notin F+K$; then

$$(x_0-F) \cap K = \emptyset.$$

The set x_0-F is closed (for the mapping f such that $f(x) = x_0-x$ is a homeomorphism). Hence, by Theorem 4, there exists a neighbourhood U such that

$$(x_0-F+U) \cap K = \emptyset.$$

Therefore

$$(x_0+U) \cap (K+F) = \emptyset$$

and so there exists a neighbourhood $U(x_0) = U+x_0$ not meeting $K+F$, whence $K+F$ is closed.

(iii) In $X \times X$, the set $K \times K'$ is compact (by Tychonoff's theorem). Since the mapping σ of $X \times X$ into X defined by $\sigma(x, y) = x+y$ is continuous, the set $K+K' = \sigma(K \times K')$ is compact.

Theorem 6. *If E is a vector subspace of X, its closure $\bar{E}$ is also a vector subspace of X.*

Proof. The mapping σ given by $\sigma(x, y) = x+y$ is continuous and satisfies $\sigma(E \times E) \subset E$; therefore by Corollary 1 on page 57, we have $\sigma(\overline{E \times E}) \subset \bar{E}$. But, by Theorem 1, § 9, Chapter IV, we have $\bar{E} \times \bar{E} = \overline{E \times E}$; therefore

$$\sigma(\bar{E} \times \bar{E}) \subset \bar{E}$$

and so

$$\begin{cases} x \in \bar{E} \\ y \in \bar{E} \end{cases} \quad \Rightarrow \quad x+y \in \bar{E}.$$

The mapping defined by $\tau(\lambda, x) = \lambda x$ is continuous in $\mathbf{R} \times X$ and satisfies

$$\tau(\mathbf{R} \times E) \subset E;$$

therefore

$$\tau(\mathbf{R} \times \bar{E}) = \tau(\bar{\mathbf{R}} \times \bar{E}) = \tau(\overline{\mathbf{R} \times E}) \subset \bar{E}$$

and so

$$\begin{cases} x \in \bar{E} \\ \lambda \in \mathbf{R} \end{cases} \quad \Rightarrow \quad \lambda x \in \bar{E}.$$

Hence $\bar{E}$ is a vector subspace.

Theorem 7. *Let f be a numerical linear function in X. Then the planes $E_f^\alpha = \{x \,/\, f(x) = \alpha\}$ are all closed if and only if f is continuous in X.*

Proof. If f is a continuous function, the set

$$E_f^\alpha = \{x \,/\, f(x) \leqq \alpha\} \cap \{x \,/\, f(x) \geqq \alpha\}$$

is closed.

If the function f is not continuous at a point x_0 then there exists a number ε such that

$$(\forall U;\ U \in \mathscr{B}):\ f(x_0+U) \not\subset [a-\varepsilon,\ a+\varepsilon],$$

where $f(x_0) = a$. Since $\mathscr{B}$ is the family of symmetric neighbourhoods and f is linear, we have

$$(\forall U;\ U \in \mathscr{B}) : f(x_0+U) \supset [a-\varepsilon,\ a+\varepsilon].$$

Then the plane $P = \left\{ x \,/\, f(x) = a+\frac{\varepsilon}{2} \right\}$ is not closed, since $x_0 \notin P$ and every neighbourhood of x_0 meets P.

Theorem 8. *A plane P which is not closed satisfies $\bar{P} = X$ (we then say that it is 'everywhere dense').*

Proof. Let P be a plane, and suppose that $0 \in P$ (if this is not the case, we make a translation). If P is not closed, there exists a point a such that

$$a \in \bar{P},\ a \notin P.$$

Let D_a be the straight line through 0 and a; since $\bar{P}$ is, by Theorem 6, a vector space, we have

$$\bar{P} \supset D_a+P.$$

But, since P is a plane, we have $D_a+P = X$ and so $\bar{P} = X$.

Theorem 9. *If G is a non-empty open set situated on one side of a plane E_f^α, so that $G \subset \{x \,/\, f(x) \geqq \alpha\}$, then we have*

$$G \subset \{x \,/\, f(x) > \alpha\}.$$

Proof. Suppose that the result is false. Then there exists a point $a \in G$ such that $f(a) = \alpha$ and a point $b \in X$ such that $f(b) = \alpha+1$ (for f is linear and not identically zero). The straight line through a and b is

$$D = \{x \,/\, x = \lambda b+(1-\lambda)a,\ \lambda \in \mathbf{R}\}.$$

If x runs along this line in one direction, the value of $f(x)$ increases continuously, for we have

$$f(x) = \lambda f(b)+(1-\lambda)f(a) = \lambda(\alpha+1)+(1-\lambda)\alpha = \lambda+\alpha.$$

Since a is an internal point of G, there exist points x_1 and x_2 such that

$$x_1, x_2 \in G,\quad f(x_1) > \alpha,\quad f(x_2) < \alpha.$$

But this contradicts the hypothesis $G \subset \{x \,/\, f(x) \geqq \alpha\}$ and the theorem follows.

§ 3. General properties of convex sets

In a topological vector space X, we can prove certain classical (and very convenient) theorems concerning open or closed convex sets.

Theorem 1. *If C is a convex subset of X and if $a \in \mathring{C}$, $b \in \bar{C}$, the interior $\mathring{C}$ contains the set*

$$[a, b[\, = \{x \,/\, x = pa+qb, p > 0, q \geqq 0, p+q = 1\}.$$

Proof. Suppose that $x_0 \in \,]a, b[$; we must show that x_0 is an interior point of the set C. Consider the transformation $x \to y = x_0+\lambda(x-x_0)$ (often called a 'homothety of centre x_0 and ratio λ'). This leaves the point x_0 fixed and, if we choose λ to be such that $b = x_0+\lambda(a-x_0)$, it transforms a into b (we then have $\lambda < 0$).

Let $U(a)$ be an open neighbourhood of a, contained in C. The above homothety transforms $U(a)$ into an open neighbourhood $V(b)$; in $V(b)$, there exists a point v such that $v \in C \cap V(b)$ (for $b \in \bar{C}$). The point v is the image under the homothety of a point u of $U(a)$, so that

$$v = x_0+\lambda(u-x_0).$$

We also have

$$v-x_0 = \lambda(u-x_0) = \lambda(u-v)+\lambda(v-x_0)$$

whence

$$v-x_0 = \frac{\lambda}{\lambda-1}(v-u).$$

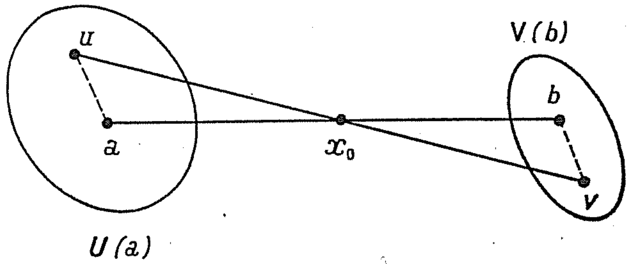


FIG. 48

We have $\dfrac{\lambda}{\lambda-1} > 0$, $\dfrac{\lambda}{\lambda-1} < 1$; the homothety of centre v and of ratio $\dfrac{\lambda}{\lambda-1}$ transforms u into x_0 and $U(a)$ into an open neighbourhood $W(x_0)$. Since $v \in C$ and $U(a) \subset C$, we have

$$W(x_0) \subset C.$$

Therefore the point x_0 is interior to C.

COROLLARY. *The interior $\mathring{C}$ of a convex set C is a convex set.*

Proof. If $x \in \mathring{C}$, $y \in \mathring{C}$, then in particular $y \in \bar{C}$ and so

$$[x, y] \subset \mathring{C}.$$

Theorem 2. *The closure $\bar{C}$ of a convex set C is a convex set.*

Proof. Let p, q be two numbers such that $p > 0, q > 0, p+q = 1$, and let ρ be the single-valued mapping of $X \times X$ into X defined by

$$\rho(x, y) = px+qy.$$

Then ρ is continuous and maps $C \times C$ into C; therefore, by Corollary 1 to the theorem of § 3, Chapter IV, we have

$$\rho(\bar{C} \times \bar{C}) = \rho(\overline{C \times C}) \subset \bar{C}.$$

Hence

$$\begin{cases} x \in \bar{C} \\ y \in \bar{C} \end{cases} \Rightarrow \quad px+qy \in \bar{C}.$$

This is valid for each pair $(p, q) \in \mathbf{P}_2$ and so $\bar{C}$ is convex.

Corollary. *The closed-convex closure of a set A is equal to the topological closure of $[A]$.*

This follows immediately from the theorem and Theorem 4, § 7, Chapter I.

On the other hand, we observe that if F is a closed set, its convex closure $[F]$ is not necessarily a closed set; but if F is a finite set, we can see, as in the proof of Theorem 1 of § 2, Chapter VIII, that $[F]$ is closed.

Theorem 3. *If G is an open set, its convex closure $[G]$ is also an open set.*

Proof. Let $x \in [G]$; we have

$$x = p_1a_1+p_2a_2+ \dots +p_na_n; \quad a_1, a_2, \dots, a_n \in G;$$
$$(p_1, p_2, \dots, p_n) \in \mathbf{P}_n; \quad n \in \mathbf{N}.$$

The set $p_1G+p_2G+ \dots +p_nG$ is open, contains x and is contained in $[G]$; the set $[G]$ is therefore a neighbourhood of each of its points and so is open.

First separation theorem. *If A and B are two disjoint non-empty convex sets and A admits an interior point, then there exists a closed plane E_f^α which separates A and B, so that*

$$\sup_{x \in A} f(x) \leqq \alpha \leqq \inf_{y \in B} f(y).$$

This follows at once from the Hahn-Banach Theorem, proved in § 6, Chapter VII, together with Theorem 8 of § 2.

Corollary 1. *If C is a convex set having an interior point and A is a non-empty linear variety, and if $C \cap A = \emptyset$, there exists a closed plane E_f^α containing A and such that C is on one side of E_f^α.*

Proof. By the theorem, there exists a closed plane P separating A and C. If $a \in A$, the plane $P+a$ contains A and the set C is on one side of this plane.

COROLLARY 2. *If C is a closed convex set having at least one interior point, every frontier point of C (and, in particular, every extreme point of C) belongs to a closed plane of support of C.*

Proof. If $x_0 \in C - \mathring{C}$, we have $x_0 \notin \mathring{C}$, and we can separate $\{x_0\}$ and $\mathring{C}$ by a closed plane E_f^α such that $f(x_0) \leqq \alpha$. The closed half-space $\{x \,/\, f(x) \geqq \alpha\}$ contains $\mathring{C}$ and so contains its closure C; therefore it contains x_0 and hence $f(x_0) = \alpha$. Thus E_f^α is a plane of support of C containing x_0.

§ 4. Separation by convex functions

Fundamental theorem. *Let $C = \{c_i \,/\, i \in I\}$ be a compact convex subset in a topological vector space X and let Φ be a family of lower semi-continuous convex functions. If, for each point c_i, there exists a function $\phi_i \in \Phi$ such that $\phi_i(c_i) > 0$, then there exists a function ϕ_0 such that*

$$\phi_0 = \sum_{i=1}^{m} p_i \phi_i, \quad \phi_1, \phi_2, \ldots, \phi_m \in \Phi, \quad (p_1, p_2, \ldots, p_m) \in \mathbf{P}_m;$$
$$\inf_{x \in C} \phi_0(x) > 0.$$

The proof of this theorem is similar to that of the second fundamental theorem of § 7, Chapter VIII.

Let Φ be a family of lower semi-continuous convex functions and suppose that

$$\begin{cases} \phi_1, \phi_2 \in \Phi \\ (p_1, p_2) \in \mathbf{P}_2 \end{cases} \quad \Rightarrow \quad p_1\phi_1 + p_2\phi_2 \in \Phi.$$

We say that a set C in a vector space X is **regularly convex**(1) with respect to Φ if to each point a not belonging to C there corresponds a function $\phi_a \in \Phi$ such that

$$\sup_{x \in C} \phi_a(x) \leqq 0 < \phi_a(a).$$

From the definition, it follows that the full set X is regularly convex; it is convenient to regard the empty set as also being regularly convex.

Let E' be a vector subspace of the dual space X'; we say that a set C is regularly convex with respect to the space E' if it is regularly convex with respect to

$$\Phi = \{f + \alpha \,/\, f \in E'; \quad \alpha \in \mathbf{R}\}.$$

(1) The idea of 'regularly convex functions' was introduced by M. Krein and D. Smulian (*Ann. Math.*, vol. 41, 1940, p. 556). Here it is introduced in a wider sense, to enable us to apply it to certain non-linear problems. For this purpose we use a previous work (*Bull. ed la Soc. Math. France*, vol. 82, 1954, p. 301) in which we prove the general theorem of separation by convex functions. We note that certain theorems (§ 4) are expressed here with less restrictive hypothesis than in the classical work of Bourbaki (*Les structures fondamentales de l'Analyse*, Livre V, Chapters I, II).

EXAMPLE. If X' is the family of numerical linear functions defined in $\mathbf{R}^n$, a set $C \subset \mathbf{R}^n$ is regularly convex *with respect to the family* X' if and only if it is a closed convex cone with vertex 0; a set $D \subset \mathbf{R}^n$ is regularly convex *with respect to the space* X' if and only if it is a closed convex set (see § 1, Chapter VIII).

Theorem 1. *A set which is regularly convex with respect to Φ is convex and closed.*

Proof. If C is regularly convex, then to each point $a \notin C$, there corresponds a function ϕ_a such that

$$\sup_{x \in C} \phi_a(x) \leqq 0 < \phi_a(a).$$

Therefore

$$C = \bigcap_{a \notin C} \{x \mid \phi_a(x) \leqq 0\}.$$

The set $\{x \mid \phi_a(x) \leqq 0\}$ is closed, for ϕ_a is lower semi-continuous; it is also convex, for ϕ_a is a convex function. Hence C is both closed and convex.

Theorem 2. *The intersection of any collection of sets which are regularly convex with respect to Φ is also regularly convex with respect to Φ.*

Proof. Let C_i $(i \in I)$ be regularly convex sets and put $C = \bigcap_i C_i$. If $a \notin C$, there exists an index k such that $a \notin C_k$ and so there exists a function ϕ in Φ such that

$$\sup_{x \in C} \phi(x) \leqq \sup_{x \in C_k} \phi(x) \leqq 0 < \phi(a).$$

Therefore C is regularly convex.

General separation theorem. *Let X be a topological vector space and let K be a non-empty compact convex set in X. If C is a non-empty set regularly convex with respect to Φ and if $C \cap K = \varnothing$, then there exists a function ϕ_0 in Φ such that*

$$\sup_{x \in C} \phi_0(x) \leqq 0 < \inf_{y \in K} \phi_0(y).$$

Proof. Suppose that $a \in K$; then there exists a function $\phi_a \in \Phi$ such that

$$\sup_{x \in C} \phi_a(x) \leqq 0 < \phi_a(a).$$

Therefore, by the fundamental theorem, there exists a function ϕ_0 of the form $\phi_0 = \sum_{i=1}^{m} p_i \phi_{a_i}$, where $(p_1, p_2, \ldots, p_m) \in \mathbf{P}_m$, such that

$$\inf_{y \in K} \phi_0(y) > 0.$$

Since Φ is assumed to be convex, we have $\phi_0 \in \Phi$; furthermore, we have

$$\sup_{x \in C} \phi_0(x) \leqq \sum_{i=1}^{m} p_i \sup_{x \in C} \phi_{a_i}(x) \leqq 0.$$

Therefore the function ϕ_0 satisfies the required conditions.

Smulian's theorem. *Let E' be a total vector subspace of the dual space X'. Then a compact convex set $K \subset X$ is regularly convex with respect to the space E'.*

Proof. Suppose that $a \notin K$. The set $\{a\}$ is regularly convex with respect to the space E', because E' is a total space and so

$$x \neq a \Rightarrow x-a \neq 0 \Rightarrow (\exists_{E'} f) : f(x-a) > 0 \Rightarrow (\exists_{E'} f) : f(a) < f(x).$$

Hence, by the preceding theorem, there exists a linear affine function ϕ such that

$$\phi(a) \leqq 0 < \inf_{y \in K} \phi(y).$$

Therefore, putting $\phi'(x) = -\phi(x) + \varepsilon$, we have

$$\sup_{y \in K} \phi'(y) \leqq 0 < \phi'(a)$$

and so K is regularly convex with respect to Φ.

Corollary. *If E' is a total subspace of X' and if K and K' are two disjoint compact convex sets, then we can separate K and K' strictly by a plane E_f^α such that $f \in E'$.*

Theorem 3. *In a space X, a compact convex set K is regularly convex with respect to the dual space X'.*

Proof. By the axiom of totality (see page 237) the space X' is total and so this result follows at once from Smulian's theorem.

Corollary. *If K and K' are two disjoint non-empty compact convex sets, there exists a closed plane E_f^α which separates K and K' strictly.*

Theorem 4. *In a space X, a closed convex set C having at least one interior point is regularly convex with respect to the dual space X'.*

Proof. Suppose that $a \notin C$ and $x_0 \in \mathring{C}$. If D^+ is the half-line issuing from x_0 and passing through a, then $D^+ \cap C$ is a closed interval $[x_0, y]$ and y is a frontal point of C. By Corollary 2 in § 3, there exists a closed plane E_f^α, passing through y, such that

$$\sup_{x \in C} f(x) \leqq \alpha.$$

Since $x_0 \in \overset{\circ}{C}$, we have $f(x_0) < \alpha$ (by Theorem 9, § 2). Therefore, since f is linear and $f(y) = \alpha$, we have $f(a) > \alpha$, whence

$$\sup_{x \in C} (f(x)-\alpha) \leqq 0 < (f(a)-\alpha).$$

Thus C is regularly convex with respect to the dual space X'.

COROLLARY. *If K is a non-empty compact convex set, if C is a closed convex set admitting at least one interior point and if $C \cap K = \varnothing$, there exists a closed plane E_f^α which separates C and K strictly.*

Plane of Support Theorem. *A non-empty compact convex set C admits at least one extreme point and every plane of support of C contains an extreme point of C.*

Proof. Let $\mathscr{K}$ be the collective family of non-empty compact convex sets K contained in C, such that

$$[a, b] \subset C, \quad]a, b[\cap K \neq \varnothing \quad \Rightarrow \quad [a, b] \subset K.$$

Since C satisfies this condition, $\mathscr{K}$ is non-empty; it can be ordered by means of the relation $\subset$. Furthermore, the intersection of a totally ordered sub-family of $\mathscr{K}$ belongs to $\mathscr{K}$. Hence, by Zorn's Theorem, there exists a set K_0 in $\mathscr{K}$ such that

$$K \subset K_0, \quad K \in \mathscr{K} \quad \Rightarrow \quad K = K_0.$$

Suppose that $x_0 \in K_0$; if there exists a point y_0 in K_0 such that $y_0 \neq x_0$, there exists a continuous linear function f such that $f(y_0) < f(x_0)$, by the totality axiom. Put

$$\alpha = \max_{x \in K_0} f(x).$$

The set $E_f^\alpha \cap K_0$ is non-empty, compact and convex and is contained in C; also

$$\left.\begin{array}{r} [a, b] \subset C, \\ ]\, a, b\, [\cap (E_f^\alpha \cap K_0) \neq \varnothing \end{array}\right\} \quad \Rightarrow \quad [a, b] \subset E_f^\alpha \cap K_0.$$

Therefore $E_f^\alpha \cap K_0 \in \mathscr{K}$; since $E_f^\alpha \cap K_0 \subset K_0$, we have $E_f^\alpha \cap K_0 = K_0$; but this is impossible, since $y_0 \in K_0$ and $y_0 \notin E_f^\alpha \cap K_0$.

Hence $K_0 = \{x_0\}$ and so x_0 is an extreme point of C.

If E_f^α is a plane of support of C, the set $E_f^\alpha \cap C$ is compact and convex and so admits an extreme point x_0; it is clear that x_0 is also an extreme point of C.

Theorem of Krein and Milman. *A non-empty compact convex set C is equal to the closed-convex closure of the profile $\ddot{C}$ of C.*

Proof. Let $\bar{c}[\ddot{C}]$ be the closed-convex closure of $\ddot{C}$, namely the intersection of all the closed convex sets containing $\ddot{C}$. We have $C \supset \ddot{C}$ and so

$$C = \bar{c}[C] \supset \bar{c}[\ddot{C}].$$

Therefore, if $C \neq \bar{c}[\ddot{C}]$, there exists a point x_0 such that

$$x_0 \in C,\ x_0 \notin \bar{c}[\ddot{C}].$$

The set $\bar{c}[\ddot{C}]$ is non-empty (by the preceding theorem), compact (because it is closed and contained in C) and convex; therefore, by Theorem 3, $\bar{c}[\ddot{C}]$ is regularly convex with respect to X' and so there exists a continuous linear function f such that

$$\sup \{f(x) \,/\, x \in \bar{c}[\ddot{C}]\} < f(x_0).$$

Put $\alpha = \sup_{x \in C} f(x)$; since $x_0 \in C$, we have $f(x_0) \leqq \alpha$, whence $E_f^\alpha \cap \bar{c}[\ddot{C}] = \emptyset$. But this is impossible, since E_f^α is a plane of support of C and so, by the preceding theorem, it meets $\ddot{C}$. It follows that $C = \bar{c}[\ddot{C}]$.

§ 5. Locally convex spaces

We say that a space X is **locally convex** if it is separated and if there exists a fundamental base of convex neighbourhoods for the point 0.

EXAMPLE 1. A normed space X, with its strong topology, is a locally convex space. To prove this, consider the fundamental base of neighbourhoods $B_\varepsilon(0) = \{x \,/\, \| x \| \leqq \varepsilon\}$, where $\varepsilon > 0$; if $(p, q) \in \mathbf{P}_2$, we have

$$\begin{cases} x \in B_\varepsilon(0) \\ y \in B_\varepsilon(0) \end{cases} \Rightarrow \quad \begin{cases} \| x \| \leqq \varepsilon \\ \| y \| \leqq \varepsilon \end{cases} \Rightarrow$$
$$\Rightarrow \quad \| px+qy \| \leqq p\varepsilon+q\varepsilon = \varepsilon \quad \Rightarrow \quad px+qy \in B_\varepsilon(0).$$

The neighbourhood $B_\varepsilon(0)$ is therefore convex.

EXAMPLE 2. A normed space X, with its weak topology, is a locally convex space. To prove this, consider the fundamental base of neighbourhoods

$$N_\varepsilon^\Phi = \{x \,/\, |f(x)| \leqq \varepsilon \text{ for all } f \in \Phi\},$$

where Φ is a finite subset of the dual X' and $\varepsilon > 0$. The set N_ε^Φ is convex, because it is the intersection of closed half-spaces.

EXAMPLE 3. Let ϕ be a numerical function defined in $\mathbf{R}^n$; the smallest closed set outside which it is identically zero is called the *support* of ϕ. Let $\mathscr{D}_m$ denote the vector space formed by all the functions ϕ having a compact support and admitting continuous partial derivatives up to the mth order. If $\phi \in \mathscr{D}_m$, we write

$$\| \phi \|_m = \sup_{p_1+p_2 \ldots +p_n \leqq m} \ \sup_x \left| \frac{\partial^{p_1+p_2+\ldots p_n}}{(\partial x^1)^{p_1} \ldots (\partial x^n)^{p_n}} \phi(x) \right|.$$

We say that the functions $\phi_1, \phi_2, \ldots$ converge to 0 in $\mathscr{D}_m$ if their supports remain in a fixed compact set K and if

$$\lim_{k=\infty} \| \phi_k \|_m = 0.$$

We see at once that $\mathscr{D}_m$ is an L^*-space (cf. § 2, Chapter IV) and that no proper topology can give the same convergence. Nevertheless we can consider the set $\mathscr{D}_m^K$ of functions in $\mathscr{D}_m$ which have their supports in the compact set K; convergence in $\mathscr{D}_m^K$ can then be defined by means of the topology of a locally convex space, with neighbourhoods of the origin of the form

$$U_\varepsilon = \{\phi \,/\, \phi \in \mathscr{D}_m,\ \| \phi \|_m \leqq \varepsilon,\ \phi \text{ has a support in } K.\}$$

EXAMPLE 4. If we put $m = \infty$, we obtain the space of infinitely differentiable functions with compact support, which we denote by $\mathscr{D}$. The topological dual $\mathscr{D}'$ is the **space of distributions**, introduced by L. Schwartz in order to generalise the idea of function in certain functional equations. We can show that the space of distributions is locally convex.

REMARK. *A locally convex space satisfies the totality axiom.* To prove this, it is necessary to show that if $x_0 \neq 0$, there exists a function $f \in X'$ such that $f(x_0) \neq 0$.

Suppose that $x_0 \neq 0$; then there exists a convex open set U such that $0 \notin U+x_0$. The set $U+x_0$ is convex, open and does not meet the convex set $\{0\}$. Therefore, by the first separation theorem, we can find a continuous linear function f such that

$$0 = f(0) \leqq \inf_{x \in U+x_0} f(x).$$

Moreover, since $U+x_0$ is an open set we have

$$(\forall x;\ x \in U+x_0) : f(x) > 0$$

by Theorem 9, § 2. Hence $f(x_0) > 0$.

Theorem 1. *In a locally convex space X, a necessary and sufficient condition for a set C to be regularly convex with respect to the dual space X' is that it is convex and closed.*

Proof. Let C be a closed convex set. If $a \notin C$, there exists a convex open neighbourhood U such that $(U+a) \cap C = \varnothing$. Then the sets $U+a$ and C can be separated by a plane E_f^α, so that

$$\sup_{x \in C} f(x) \leqq \alpha \leqq \inf_{y \in U+a} f(y).$$

Put $\phi(x) = f(x) - \alpha$. The set $U+a$ is open and so we have

$$\sup_{x \in C} \phi(x) \leqq 0 \leqq \inf_{y-a \in U} \phi(y) < \phi(a).$$

Therefore C is regularly convex.

Conversely, every regularly convex set is convex and closed by Theorem 1, § 4.

COROLLARY (Second separation theorem). *If K is a non-empty compact convex set in a locally convex space X and if C is a closed convex set disjoint from K, there exists a closed plane E_f^α which separates C and K strictly.*

We now consider mappings of a locally convex space into itself.

Theorem 2. *If N is a closed neighbourhood of the origin, the mapping determined by $N(x) = N+x$ is closed* (cf. page 111).

Proof. Suppose that $y_0 \notin N(x_0)$. The set $\{y_0\}$ is compact and the set $N(x_0)$ is closed. Therefore, by Theorem 4, § 2, there exists a neighbourhood U such that

$$(y_0+U) \cap (x_0+N+U) = \varnothing.$$

In other words, we have

$$N(x) \cap (y_0+U) = \varnothing$$

for all x in $U+x_0$. Hence the mapping is closed.

Fixed point theorem (Tychonoff, Kakutani, Ky Fan). *Let C be a non-empty compact convex set in a locally convex space X. If Γ is a u.s.c. mapping of C into C and if, for all x, the set Γx is convex and non-empty, then there exists a point x_0 in C such that*

$$x_0 \in \Gamma x_0.$$

Proof. Let N be a closed symmetric neighbourhood and consider the set

$$F_N = \{x \,/\, x \in (\Gamma x+N) \cap C\}.$$

(1) We first show that F_N is closed. Since the mapping determined by $N(x) = N+x$ is closed, the mapping determined by $M(x) = N(x) \cap C$ is u.s.c. and also the mapping determined by $M\Gamma x = (\Gamma x+N) \cap C$ is u.s.c., so that

$$(1 \cap M\Gamma)x = (\Gamma x+N) \cap C \cap \{x\}$$

determines a u.s.c. mapping. The set $(1 \cap M\Gamma)^+ \varnothing$ is therefore open and its complement, which is the set F_N, is closed.

(2) We now show that F_N is non-empty. Since C is compact, there exist points $x_1, x_2, \ldots, x_n$ such that $C \subset \bigcup_{i=1}^{n} (x_i+N)$. By Theorem 1, § 2, Chapter VIII (which is valid in general topological vector spaces as well as

in $\mathbf{R}^n$), we know that the convex closure $K = [x_1, x_2, \ldots, x_n]$ is a compact set, having dimension at most $n-1$. We can then consider in $\mathbf{R}^{n-1}$ the mapping of K into K given by $B(x) = (\Gamma x + N) \cap K$; the set $B(x)$ is convex since N is convex and is non-empty since N is a symmetric neighbourhood, whence

$$\Gamma x \subset C \subset K+N \quad \Rightarrow \quad \Gamma x \cap (K+N) \neq \varnothing \quad \Rightarrow \quad (\Gamma x+N) \cap K \neq \varnothing.$$

Therefore, by Kakutani's theorem (§ 2, Chapter VIII), there exists a point x_0 such that

$$x_0 \in (\Gamma x_0 + N) \cap K.$$

Hence $x_0 \in (\Gamma x_0 + N) \cap C$ and $F_N \neq \varnothing$.

(3) If M and N are two symmetric closed neighbourhoods, we have

$$F_M \cap F_N \supset F_{M \cap N}.$$

Then, because C is compact,

$$\bigcap_N F_N \neq \varnothing,$$

by the finite intersection axiom. If x_0 is a point of this intersection, then we have $x_0 \in \Gamma x_0$ (for otherwise $x_0 \notin F_N$ for some neighbourhood N).

COROLLARY (Schauder's theorem). *Let C be a non-empty compact convex subset of a locally convex space X and let σ be a single-valued mapping of C into C. Then there exists a point x_0 in C such that $\sigma x_0 = x_0$.*

§ 6. Banach spaces: strong convergence

A complete normed space is called a **Banach space**; in such a space, every Cauchy sequence (x_n) is convergent (see page 88).

EXAMPLE 1. If $p \geqq 1$, the space L_p consisting of sequences $x = (x^n)$ such that $\sum_{n=1}^{\infty} | x^n |^p < +\infty$ is a normed space, with

$$\| x \| = \Big(\sum_{n=1}^{\infty} | x^n |^p\Big)^{\frac{1}{p}}.$$

We shall show that it is a Banach space. Let $(x_1, x_2, x_3, \ldots)$ be a Cauchy-convergent sequence. Then, for all $\varepsilon > 0$, there exists a number n such that

$$(\forall k) : \| x_{n+k} - x_n \| \leqq \varepsilon.$$

Therefore

$$(\forall k) : | x_{n+k}^j - x_n^j |^p \leqq \sum_i | x_{n+k}^i - x_n^i |^p \leqq \varepsilon^p.$$

Thus the sequence (x_n^j) in $\mathbf{R}$ is Cauchy-convergent and therefore converges to an element x_0^j in $\mathbf{R}$.

We now show that $x_0 = (x_0^1, x_0^2, \ldots) \in L_p$. In fact, for any q, we have

$$\sum_{i=1}^{q} | x_n^i - x_{n+k}^i |^p \leqq \varepsilon^p.$$

Letting $k \to +\infty$, we get

$$\sum_{i=1}^{q} | x_n^i - x_0^i |^p \leqq \varepsilon^p.$$

Letting $q \to +\infty$, we see that $(x_n - x_0) \in L_p$; therefore, since L_p is a vector space, we have $x_0 = x_n - (x_n - x_0) \in L_p$.

Finally, we show that $(x_n) \to x_0$. In fact, we have just seen that

$$\sum_{i=1}^{\infty} | x_{n+k}^i - x_0^i |^p \leqq \varepsilon^p$$

and so we have

$$m \geqq n \quad \Rightarrow \quad \| x_m - x_0 \| \leqq \varepsilon.$$

EXAMPLE 2. Consider the space $\mathscr{L}_p (p \geqq 1)$ consisting of numerical functions ϕ defined on $[0, 1]$ (to within a set of measure zero) and such that

$$\| \phi \| = \left(\int_0^1 | \phi(t) |^p \, dt \right)^{\frac{1}{p}} < +\infty.$$

We can verify as above that $\mathscr{L}_p$ is a Banach space.

EXAMPLE 3. Consider a function ϕ defined on $[0, 1]$ and a sub-division $\tau = (t_0, t_1, \ldots, t_k)$ of $[0, 1]$ such that $0 = t_0 \leqq t_1 \leqq \ldots \leqq t_k = 1$. Write

$$v(\tau) = \sum_{i=1}^{k} | \phi(t_i) - \phi(t_{i-1}) |.$$

The number

$$\| \phi \| = \sup v(\tau)$$

is called the **variation** of the function ϕ. Let $\mathscr{V}$ be the set of functions ϕ such that

(1) $\| \phi \| < +\infty$,
(2) $\phi(0) = 0$,
(3) $\phi(t) = \frac{1}{2} [\lim_{\varepsilon \to 0} \phi(t+\varepsilon) + \lim_{\varepsilon \to 0} \phi(t-\varepsilon)]$ if $t \in]0, 1[$.

We can verify that $\mathscr{V}$ is a Banach space.

EXAMPLE 4. Let $\mathscr{C}$ be the space of numerical functions ϕ defined and continuous on $[0, 1]$, together with the norm

$$\| \phi \| = \max_t | \phi(t) |.$$

We can see as in example 1 that $\mathscr{C}$ is a Banach space.

Theorem 1. *If X is a Banach space, its dual X' is also a Banach space.*

Proof. We have seen that the dual X' of a normed space X is also normed, and has the norm

$$\| f \| = \sup_{x \in X} \frac{|f(x)|}{\| x \|} = \sup_{\| y \| \leqq 1} |f(y)|.$$

Let (f_n) be a Cauchy-convergent sequence in X'. Then, given $\varepsilon > 0$, there exists an integer n such that

$$(\forall k) : \| f_{n+k} - f_n \| \leqq \varepsilon.$$

Therefore

$$(\forall k) : |f_{n+k}(x) - f_n(x)| \leqq \| f_{n+k} - f_n \| \cdot \| x \| \leqq \varepsilon \| x \|.$$

This shows that, for all x, the sequence $((f_n(x))$ converges; suppose that it converges to $g(x)$. Since this convergence is uniform in the unit ball, the function g is continuous in the unit ball, by Weierstrass' theorem. Since g is linear, it is bounded in the unit ball and so is continuous in the whole of X, whence $g \in X'$. Finally, we have

$$(\forall x; \| x \| \leqq 1) : |f_n(x) - g(x)| \leqq \varepsilon.$$

Therefore $\| f_n - g \| \leqq \varepsilon$ and (f_n) is a convergent sequence.

We say that two Banach spaces X and Y are **in duality** if to each $x \in X$ and each $f \in Y$, there corresponds a number $\langle f, x \rangle$ such that

$$\begin{array}{lll}
(1) & \langle f, \lambda x + \lambda' x' \rangle = \lambda \langle f, x \rangle + \lambda' \langle f, x' \rangle, & \\
(2) & \langle \lambda f + \lambda' f', x \rangle = \lambda \langle f, x \rangle + \lambda' \langle f', x \rangle, & \\
(3) & (\forall_X x) \langle f, x \rangle = 0 \quad \Rightarrow & f = 0, \\
(4) & (\forall_Y f) \langle f, x \rangle = 0 \quad \Rightarrow & x = 0.
\end{array}$$

The number $\langle f, x \rangle$ is called the **scalar product** of f and x.

Clearly the Banach space X and its dual X' are in duality, if, for $f \in X'$ and $x \in X$ we write $\langle f, x \rangle = f(x)$. Conditions (1) and (2) are easily verified; condition (3) follows from the definition of the function $f = 0$. Condition (4) is satisfied because, if $x \neq 0$, there exists a function $f \in X'$ such that $f(x) \neq 0$ (totality axiom).

If X is a Banach space which is in duality with itself and if

$$\langle x, x \rangle^{\frac{1}{2}} = \| x \|,$$

then we say that X is an (abstract) **Hilbert space.**

EXAMPLE 1. Suppose that $p > 1, q > 1, \frac{1}{p}+\frac{1}{q} = 1$ and that $h \in \mathscr{L}_p, g \in \mathscr{L}_q$. Put

$$\langle h, g\rangle = \int_0^1 g(t)h(t)\,\mathrm{d}t \qquad \text{(Lebesgue integral)}$$

By the Hölder inequality, we have $\langle h, g\rangle < +\infty$ and $\langle h, g\rangle$ is a scalar product. We can show (Riesz' theorem) that $\langle h, g\rangle$ is a continuous linear function in g and conversely that any continuous linear function in g can be expressed in the form $\langle h, g\rangle$ where $h \in \mathscr{L}_p$; moreover, the norm of h in $\mathscr{L}_p$ is equal to the norm of the function $\langle h, g\rangle$ in the dual of $\mathscr{L}_q$. We can then agree to identify the dual of $\mathscr{L}_q$ with $\mathscr{L}_p$.

EXAMPLE 2. Let $g \in \mathscr{C}$ and $\alpha \in \mathscr{V}$. Put

$$\langle \alpha, g\rangle = \int_0^1 g(t)\,\mathrm{d}\alpha\,(t) \qquad \text{(Stieljtes-Lebesgue integral)}$$

Then $\langle \alpha, g\rangle$ is a scalar product and it is a continuous linear function in g; conversely, every continuous linear function in g can be written $\langle \alpha, g\rangle$, where α is a function of bounded variation.

The norm of α in $\mathscr{V}$ is again equal to the norm of the function $\langle \alpha, g\rangle$ in the dual of $\mathscr{C}$. We can therefore agree to identify the dual of $\mathscr{C}$ with $\mathscr{V}$.

EXAMPLE 3. Suppose that $x = (x^n) \in L_2, y = (y^n) \in L_2$. Put

$$\langle x, y\rangle = \sum_{n=1}^{\infty} x^n y^n$$

As we have seen (§ 9, Chapter VIII), $\langle x, y\rangle$ is a finite number. Thus it determines a scalar product, and we have

$$\langle x, x\rangle^{\frac{1}{2}} = \| x \|.$$

Hence L_2 is a Hilbert space.

Theorem of norms. *Let E be a vector subspace of X and let ϕ be a continuous linear function defined in the space E. Then there exists a linear and continuous function f defined in X such that*

(1) $f(x) = \phi(x)$ if $x \in E$,
(2) $\| f \| = \| \phi \|_E =$ norm of ϕ in E.

Proof. Put $\alpha = \| \phi \|_E$ and consider the open ball

$$G = \left\{x \,/\, \| x \| < \frac{1}{\alpha}\right\}$$

in X. The set $A = \{x \,/\, x \in E, \phi(x) = 1\}$ is a linear variety in X; in $G \cap E$, we have

$$| \phi(x) | \leqq \alpha \| x \| < \alpha \cdot \frac{1}{\alpha} = 1.$$

Then A does not meet $G \cap E$ and so A does not meet G. By Corollary 1 to the first separation theorem (§ 3), there exists a closed plane $\{x \,/\, f(x) = 1\}$ which contains A and for which G is on one side.

We have $\phi(x) = f(x)$ in A; therefore, since f and ϕ are linear in E, we have $\phi(x) = f(x)$ in E.

Since G is an open set, on one side of the plane of equation $f(x) = 1$, then, by Theorem 9, § 2, B does not meet this plane, whence

$$f(x) = 1 \quad \Rightarrow \quad x \notin G \quad \Rightarrow \quad \| x \| \geqq \frac{1}{\alpha}.$$

Therefore, since f is linear, we have

$$f(x) \leqq \alpha \cdot \| x \|$$

in the whole space. Hence

$$|f(x)| = \max \{f(x), f(-x)\} \leqq \max \{\alpha \| x \|, \alpha \| -x \| \} = \alpha \| x \|$$

and so

$$\| f \| = \sup_{x \in X} \frac{|f(x)|}{\| x \|} \leqq \alpha.$$

On the other hand, we have

$$\| f \| = \sup_{x \in X} \frac{|f(x)|}{\| x \|} \geqq \sup_{x \in E} \frac{|f(x)|}{\| x \|} = \alpha$$

and therefore $\| f \| = \alpha = \| \phi \|_E$.

COROLLARY. *If $a \in X$, $a \neq 0$, there exists a continuous linear function f such that $f(a) = \| a \|$ and $\| f \| = 1$.*

Proof. Consider the linear function defined by $\phi(x) = \phi(\lambda a) = \lambda \| a \|$ on the subspace $E = \{\lambda a \,/\, \lambda \in \mathbf{R}\}$. We have

$$\| \phi \|_E = \sup \frac{| \phi(\lambda a) |}{\| \lambda a \|} = 1.$$

Therefore there exists a continuous linear function f such that $\| f \| = 1$ and such that $f(a) = \phi(a) = \| a \|$.

The continuous linear mappings Φ of X' into $\mathbf{R}$ form a space X'', the dual of X'; we call X'' the **bi-dual** of X. If $x \in X$, the mapping Φ_x such that $\Phi_x(f) = f(x)$ for all f in X', is a linear mapping in X' satisfying

$$|\Phi_x(f)| = |f(x)| \leqq \| x \| \cdot \| f \|.$$

Therefore, by Theorem 2, § 1, Φ_x is a continuous mapping and so $\Phi_x \in X''$. Moreover, we have

$$\| \Phi_x \| = \sup_{\| f \| \leqq 1} |\Phi_x(f)| \leqq \sup_{\| f \| \leqq 1} (\| x \| \cdot \| f \|) \leqq \| x \|.$$

By the above corollary, there exists a function f such that $\| f \| = 1$, $f(x) = \| x \|$, whence

$$\| \Phi_x \| = \sup_{\| f \| \leqq 1} |f(x)| \geqq \| x \|.$$

Therefore $\| \Phi_x \| = \| x \|$; this enables us to agree to identify the element Φ_x of X'' with the point x of X and to write $X \subset X''$.

If $X = X''$, we say that the space X is **reflexive**; in other words, for each continuous linear function Φ_0 in X' there exists a point x_0 of X such that

$$(\forall f;\, f \in X') : \Phi_0(f) = f(x_0).$$

EXAMPLE. The space $\mathscr{L}_2$ is reflexive. However, the space $\mathscr{C}$ is not reflexive, for

$$\mathscr{C}'' = \mathscr{V}' \supset\supset \mathscr{C}.$$

Theorem 2. *If $e_1, e_2, \ldots, e_n$ are n linearly independent vectors, there exists a number $\eta > 0$ such that, for every n-tuple of real numbers $(\lambda^1, \lambda^2, \ldots, \lambda^n)$, we have*

$$|\lambda^1| + |\lambda^2| + \ldots + |\lambda^n| \leqq \eta \, \| \lambda^1 e_1 + \lambda^2 e_2 + \ldots + \lambda^n e_n \|.$$

Proof. Suppose that the theorem is false. Consider an n-tuple $\lambda = (\lambda^1, \lambda^2, \ldots, \lambda^n)$ as a point of $\mathbf{R}^n$ with the norm

$$\| \lambda \| = |\lambda^1| + |\lambda^2| + \ldots + |\lambda^n|$$

(equivalent to the Euclidean norm). Then

$$\inf_{\| \lambda \| = 1} \| \sum_{i=1}^{n} \lambda^i e_i \| = 0.$$

Let $\lambda_k = (\lambda_k^1, \lambda_k^2, \ldots, \lambda_k^n)$ be a vector of $\mathbf{R}^n$ such that $\| \lambda_k \| = 1$ and such that

$$\| x(k) \| = \| \sum_{i=1}^{n} \lambda_k^i e_i \| \leqq \frac{1}{k}.$$

The sequence whose general term is $x(k)$ converges to 0.

Since the set $\{\lambda \,/\, \lambda \in \mathbf{R}^n, \| \lambda \| = 1\}$ is compact in $\mathbf{R}^n$, there exists a sub-sequence (λ_{k_n}) of the λ_k converging to a vector λ_0 such that $\| \lambda_0 \| = 1$; putting $x_0 = \lambda_0^1 e_1 + \lambda_0^2 e_2 + \ldots + \lambda_0^n e_n$, we have

$$\| x(k_n) - x_0 \| \leqq \sum_{i=1}^{n} | \lambda_{k_n}^i - \lambda_0^i | \cdot \| e_i \|.$$

Hence the sequence $x(k_n)$ converges to x_0 and, since $x_0 \neq 0$, we have obtained a contradiction.

COROLLARY 1. *If a Banach space has dimension n, its norm is equivalent to the Euclidean norm of* $\mathbf{R}^n$.

Proof. Suppose that there exist n linearly independent vectors $e_1, e_2, \ldots, e_n$ such that each point x can be written in the form

$$x = \sum_{i=1}^{n} x^i e_i; \quad x^1, x^2, \ldots x^n \in \mathbf{R}.$$

We can always suppose that $\| e_i \| = 1$ for all i, whence

$$| x^1 | + | x^2 | + \ldots + | x^n | \leqq \eta \| x \| \leqq \eta (| x^1 | + | x^2 | + \ldots + | x^n |).$$

Therefore, in the sense defined in § 1, Chapter V, the norm $\| x \|$ is equivalent to the norm defined by $| x^1 | + | x^2 | + \ldots + | x^n |$, which is itself equivalent to the Euclidean norm (see the example in § 1, Chapter V).

COROLLARY 2. *The subspace E generated by the linearly independent vectors* $e_1, e_2, \ldots, e_m$ *is a closed set.*

Proof. Let the points $x_n = \sum_{i=1}^{m} \lambda_n^i e_i$ be such that $(x_n) \to x_0$, where $x_0 \in X$; we shall prove that $x_0 \in E$.

Given $\varepsilon > 0$, there exists a number n such that

$$\begin{aligned} | \lambda_{n+k}^1 - \lambda_n^1 | &\leqq | \lambda_{n+k}^1 - \lambda_n^1 | + | \lambda_{n+k}^2 - \lambda_n^2 | + \ldots + | \lambda_{n+k}^m - \lambda_n^m | \\ &\leqq \eta \| x_{n+k} - x_n \| \leqq \varepsilon. \end{aligned}$$

Therefore the sequence (λ_n^1) in $\mathbf{R}$ converges to a point λ_0^1. Since X is a topological vector space, we have

$$(x_n) = (\sum_i \lambda_n^i e_i) \to \sum_i \lambda_0^i e_i$$

and therefore

$$x_0 = \lambda_0^1 e_1 + \lambda_0^2 e_2 + \ldots + \lambda_0^m e_m \in E.$$

Theorem 3. *If the ball* $B = \{x / \| x \| \leqq 1\}$ *in a Banach space X is a compact set, the X has finite dimension.*

Proof. Suppose that X is not generated by a finite number of vectors. Let $x_1 \in B$ be such that $x_1 \neq 0$. The vector subspace $E_1 = s[x_1]$ is strictly contained in X and so there exists a point $y_1 \notin E_1$; since E_1 is closed (by the

Corollary 2 to Theorem 2), the distance $\delta(y_1, E_1)$ from y_1 to the set E_1 is such that

$$\delta(y_1, E_1) = \inf_{z \in E_1} \| y_1 - z \| = \delta > 0.$$

Let z_1 be a point of E_1 such that $\| y_1 - z_1 \| \leqq 2\delta$ and put

$$x_2 = \frac{y_1 - z_1}{\| y_1 - z_1 \|}.$$

We have $x_2 \in B$ and, for all x in E_1, we have

$$\| x_2 - x \| = \frac{1}{\| y_1 - z_1 \|} \| (y_1 - z_1) - (\| y_1 - z_1 \|) x \| \geqq \frac{1}{2\delta} \inf_{z \in E_1} \| y_1 - z \| = \frac{1}{2}.$$

With the linear variety $E_2 = s[x_1, x_2]$ in place of E_1, we can determine a point x_3 in a similar manner, with

$$x_3 \in B,\ x_3 \neq 0,\ \| x_3 - x_1 \| \geqq \frac{1}{2},\ \| x_3 - x_2 \| \geqq \frac{1}{2}.$$

Continuing this process, we determine the points of a sequence (x_n); by Cauchy's criterion, this sequence has no point of accumulation. But $x_n \in B$ and B is compact; therefore we have a contradiction.

§ 7. Banach spaces: weak convergence

We now consider a sequence $(f_1, f_2, \ldots) = (f_n)$ in the dual X' of a Banach space X. We say that such a sequence **converges strongly** if it converges in the sense of the norm $\| f \|$, and we write $(f_n) \to g$.

On the other hand, we say that a sequence (f_n) **converges weakly** if, for all $x \in X$, the sequence $(f_n(x))$ is convergent in $\mathbf{R}$. Let $g(x)$ be the limit of $(f_n(x))$; the function g so defined is called the **weak limit** of (f_n) and we write $(f_n) \rightharpoonup g$.

We remark that weak convergence to $g = 0$ can also be defined, in a similar manner, by the topology of a locally convex space; it is sufficient to take the sets of the form

$$N_\varepsilon^E = \{ f / f \in X';\ | f(x) | \leqq \varepsilon \text{ for all } x \in E \}$$

as a base of neighbourhoods, where $\varepsilon > 0$ and E is a finite subset of X.

Osgood's Lemma. *If $g_1, g_2, g_3, \ldots$ are functions defined on an open set G of X, if $| g_n |$ is lower semi-continuous and if the sequence $(g_n(x))$ converges for all x in G, there exists a ball $B \subset G$ in which the $| g_n(x) |$ are uniformly bounded: that is,*

$$(\forall n)\ (\forall_B x) : \ | g_n(x) | < \alpha.$$

Proof. Suppose that the result is not true. Let $\alpha_1, \alpha_2, \ldots$ be numbers which increase indefinitely, let $\varepsilon_1, \varepsilon_2, \ldots$ be numbers which tend to 0 and let $m_1, m_2, \ldots$ be integers which increase indefinitely. Let B_0 be a ball contained in G; there exists an integer $i_1 \geqq m_1$, a point $x_1 \in \mathring{B}_0$ such that

$$|g_{i_1}(x_1)| \geqq \alpha_1 + 1.$$

Let λ_1 be a number such that $\lambda_1 \leqq \varepsilon_1$, $B_{\lambda_1}(x_1) \subset B_0$ and

$$x \in B_{\lambda_1}(x_1) \quad \Rightarrow \quad |g_{i_1}(x)| \geqq \alpha_1$$

(such a number λ_1 exists because of the lower semi-continuity of $g_{i_1}(x)$).

By hypothesis, there exists a number $i_2 \geqq m_2$ and a point $x_2 \in \mathring{B}_{\lambda_1}(x_1)$ such that

$$|g_{i_2}(x_2)| \geqq \alpha_2 + 1.$$

Let λ_2 be a number such that $\lambda_2 \leqq \varepsilon_2$, $B_{\lambda_2}(x_2) \subset B_{\lambda_1}(x_1)$ and

$$x \in B_{\lambda_2}(x_2) \quad \Rightarrow \quad |g_{i_2}(x)| \geqq \alpha_2.$$

In a similar manner, we can determine a number $i_3 \geqq m_3$ and a ball $B_{\lambda_3}(x_3)$ such that

$$x \in B_{\lambda_3}(x_3) \quad \Rightarrow \quad |g_{i_3}(x)| \geqq \alpha_3.$$

Continuing in this way, we obtain a sequence $x_1, x_2, x_3, \ldots$ This sequence converges, because, if $\varepsilon > 0$, there exists a number n such that

$$(\forall k) : \| x_{n+k} - x_n \| \leqq \varepsilon.$$

We therefore have $(x_n) \to x_0$; since the ball $B_{\lambda_k}(x_k)$ is a complete set (it is a closed set contained in a complete space) and since $(x_{k+n}) \to x_0$ and $x_{k+n} \in B_{\lambda_k}(x_k)$, we have $x_0 \in B_{\lambda_k}(x_k)$ for all k. Therefore at the point x_0, the sequence $(g_n(x_0))$ does not converge. This is contrary to hypothesis; therefore the theorem is proved.

Theorem of Osgood and Banach. *If (f_n) is a sequence in X' and for each point x of an open set G we have $(f_n(x)) \to f_0(x)$, then the f_n are bounded in norm.*

Proof. By the lemma, there exists a number α and a ball $B_\lambda(x_0)$ such that

$$(\forall n)\ (\forall x;\ x \in B_\lambda(x_0)) : |f_n(x)| \leqq \alpha.$$

Therefore, if $\| y \| \leqq 1$, we have

$$|f_n(y)| \leqq \left| f_n\left(\frac{x_0 - \lambda y}{\lambda}\right) \right| + \left| f_n\left(\frac{x_0}{\lambda}\right) \right| \leqq \frac{\alpha}{\lambda} + \frac{\alpha}{\lambda}$$

and so

$$\| f_n \| = \sup_{\| y \| \leqq 1} |f_n(y)| \leqq \frac{2\alpha}{\lambda}.$$

COROLLARY. *If* (f_n) *is a weakly convergent sequence, the* f_n *are bounded in norm: that is,*

$$\| f_n \| \leqq \alpha \text{ for all } n.$$

It is useful to extend the idea of weak convergence of a sequence to any filtered family; the above corollary enables us to make this generalisation.

Let X be a Banach space and X' its dual. Let $(f_i) = (f_i \mid i \in I, \mathscr{B})$ be a family together with a filter base $\mathscr{B}$ on a set I (cf. § 4, Chapter IV). The filter sections, namely the subsets of I belonging to the family $\mathscr{B}$, will be denoted by S, S' etc. We say that (f_i) **converges weakly** if

(1) the f_i are partially bounded by a number α: that is,

$$(\exists S) : i \in S \quad \Rightarrow \quad \| f_i \| \leqq \alpha,$$

(2) for all x, the family $(f_i(x))$ is convergent in $\mathbf{R}$.

Let $g(x)$ be the limit in $\mathbf{R}$ of the family $(f_i(x))$. The function g so defined is called the **weak limit** of (f_i). We also say that (f_i) converges weakly to g and we write $(f_i) \rightharpoonup g$.

Theorem 1 (Banach-Steinhaus). *If* $(f_i) \rightharpoonup g$, *then* $g \in X'$ *and also*

$$\| g \| \leqq \underline{\mathrm{Lim}} \, (\| f_i \|).$$

Proof. The function g is linear, because

$$g(\lambda x) = \lim (f_i(\lambda x)) = \lambda \lim (f_i(x)),$$
$$g(x_1 + x_2) = \lim (f_i(x_1) + f_i(x_2)) = \lim (f_i(x_1)) + \lim (f_i(x_2)).$$

Let $\alpha = \underline{\mathrm{Lim}} \, (\| f_i \|)$ be the smallest of the cluster points of the family $(\| f_i \|)$. Since the f_i are partially bounded, $\alpha < +\infty$; moreover, there exists a sub-family (f'_j) such that

$$(\| f'_i \|) \to \alpha.$$

We have

$$| f'_j(x) | \leqq \| f'_j \| \cdot \| x \|.$$

Therefore, taking limits, we get

$$| g(x) | \leqq \alpha \cdot \| x \|.$$

This shows that g is continuous and that

$$\| g \| \leqq \alpha = \underline{\mathrm{Lim}} \, (\| f_i \|).$$

Theorem 2. *If* $(f_i) \to f_0$, *we also have* $(f_i) \rightharpoonup f_0$.

Proof. Suppose that $\varepsilon > 0$; there exists a section S_0 such that

$$i \in S_0 \quad \Rightarrow \quad \| f_i - f_0 \| \leqq \varepsilon,$$

whence

$$\| f_i \| \leqq \| f_i - f_0 \| + \| f_0 \| \leqq \varepsilon + \| f_0 \|.$$

Therefore the f_i are partially bounded. Furthermore, if $x \in X$, we have

$$|f_i(x)-f_0(x)| \leqq \|f_i-f_0\| \cdot \|x\| \leqq \varepsilon \cdot \|x\|$$

whenever $i \in S_0$. Therefore $(f_i(x))$ converges to $f_0(x)$ and so $(f_i) \rightharpoonup f_0$.

Theorem 3. *If* $(f_i) \rightharpoonup f_0$, $(f'_i) \rightharpoonup f_0$, *then* $(f_i+f'_i) \rightharpoonup f_0+f'_0$.

Proof. Let α, α', S, S' be such that

$$\begin{aligned} i \in S &\Rightarrow \|f_i\| \leqq \alpha, \\ i \in S' &\Rightarrow \|f'_i\| \leqq \alpha'. \end{aligned}$$

If S_0 is a section contained in $S \cap S'$, we have

$$i \in S_0 \quad \Rightarrow \quad \|f_i+f'_i\| \leqq \|f_i\| + \|f'_i\| \leqq \alpha+\alpha'$$

and so the $(f_i+f'_i)$ are partially bounded.

Also, for all x, we have

$$(f_i(x)+f'_i(x)) \to f_0(x)+f'_0(x) = [f_0+f'_0](x).$$

Theorem 4. *If* $(f_i) \rightharpoonup f_0$, $(x_i) \to x_0$, *then* $(f_i(x_i)) \to f_0(x_0)$.

Proof. There exists a section S_0 such that

$$\|f_i\| \leqq \alpha, \quad \|x_i-x_0\| \leqq \frac{\varepsilon}{2\alpha}, \qquad |f_i(x_0)-f_0(x_0)| \leqq \frac{\varepsilon}{2}$$

whenever $i \in S_0$. Hence

$$\begin{aligned} |f_i(x_i)-f_0(x_0)| &\leqq |f_i(x_i)-f_i(x_0)| + |f_i(x_0)-f_0(x_0)| \leqq \\ &\leqq \|f_i\| \cdot \|x_i-x_0\| + |f_i(x_0)-f_0(x_0)| \leqq \alpha \cdot \frac{\varepsilon}{2\alpha}+\frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

whenever $i \in S_0$.

We say that a subset F' of X' is **weakly closed** if whenever $(f_i) \rightharpoonup f_0$ and $f_i \in F'$ for all i, we have $f_0 \in F'$. Clearly if F' is weakly closed, it is also closed, for

$$\begin{cases} (f_i) \to f_0 \\ f_i \in F' \end{cases} \quad \Rightarrow \quad \begin{cases} (f_i) \rightharpoonup f_0 \\ f_i \in F' \end{cases} \quad \Rightarrow \quad f_0 \in F'.$$

We say that a subset K' of X' is **weakly compact** if, given any filtered family in K', there exists a sub-family converging weakly to an element of K'. Clearly a compact set K' is also weakly compact.

Alaoglu's theorem. *The unit sphere* $B' = \{f \,/\, f \in X', \|f\| \leqq 1\}$ *is weakly compact.*

Proof. Consider the interval $I_x = [-\|x\|, +\|x\|]$ in $\mathbf{R}$, with the Euclidean topology; the product space $Y = \prod_{x \in X} I_x$ is then a compact topological space, by Tychonoff's theorem (§ 9, Chapter IV). With $f \in B'$, we associate the point $y_f = (f(x) \,/\, x \in X)$ in Y. The set $Y_0 = \{y_f \,/\, f \in B'\}$ is closed in Y, for

$$
\begin{array}{lll}
\begin{array}{l}(y_{f_i}) \to y_0 \\ f_i \in B'\end{array} & \Rightarrow & \left\{\begin{array}{l}(f_i(x)) \to g(x), \\ \| f_i \| \leqq 1,\end{array}\right. \\
\Rightarrow \left\{\begin{array}{l} y_0 = y_g \\ g \in B'\end{array}\right. & \Rightarrow & \left\{\begin{array}{l} g \in X', \\ \| g \| \leqq \underline{\mathrm{Lim}}\,(\| f_i \|) \leqq 1.\end{array}\right.
\end{array}
$$

Therefore Y_0 is compact, because it is a closed set contained in a compact space. Hence every filtered family in B' contains a sub-family which is weakly convergent to a point of B'.

INDEX