JOSÉ SEBASTIÃO E SILVA

TEXTOS DIDÁCTICOS

Volume III

SERVIÇO DE EDUCAÇÃO
FUNDAÇÃO CALOUSTE GULBENKIAN

TEXTOS
DIDÁCTICOS

Página em branco

JOSÉ SEBASTIÃO E SILVA

TEXTOS DIDÁCTICOS

Volume III

SERVIÇO DE EDUCAÇÃO E BOLSAS
FUNDAÇÃO CALOUSTE GULBENKIAN | LISBOA

Reservados todos os direitos de acordo com a lei

Edição da
FUNDAÇÃO CALOUSTE GULBENKIAN
Av. de Berna — Lisboa
1999

ISBN 972-31-0971-9
Depósito Legal n.º 148 805/00

III.1

THEORY OF DISTRIBUTIONS*

* Este texto tem por base apontamentos coligidos por diversos alunos de José Sebastião e Silva na sequência de um curso que realizou em 1958 na Universidade de Maryland, e que posteriormente foram utilizados, e por ele revistos, na Faculdade de Ciências de Lisboa.

CHAPTER I

HEURISTIC INTRODUCTION

1.1. Intuitive antecedents of the theory of distributions

One of the principal aims of the theory of distributions has been the justification of certain methods of calculation and reasoning used empirically by physicians and engineers, for example in the symbolic calculus of electricians (Heaviside school) and in quantum mechanics (Dirac school).

Although these methods were improvided with a logic foundation, they enabled one to obtain with simplicity and elegance the solution of problems for which the classical resources of analysis were, as a rule, less practical and even, sometimes, not efficient. However, the use of the new methods was a permanent defiance to traditional rigor of mathematics and led eventually to contradiction or to unclear situations, which made more and more desirable a rational foundation of those methods.

The history of mathematics offers many examples of similar situations. The creation of differential and integral calculus, the introduction of complex numbers, etc., were imposed by the necessity of rationalizing some heuristic methods of calculation and reasoning which mathematicians and physicists were naturally induced to apply.

Like those historical examples, the goal of the theory of distributions does not reduce to the justification of heuristic methods to avoid more or less contradictory situations. On the contrary, this theory has opened to mathematics, either pure or applied, new possibilities to an extent that is not yet possible to foresee.

The concept of distribution is a generalization of the concept of function in a similar way in which, for example, the concept of complex number is a generalization of the concept of real number. One of the distributions that first appeared outside the classical domain of mathematics is the entity improperly called the "Dirac's δ-function" (or "unitary impulse function" of electricians). According to physicists and engineers this entity should be, in the case of one variable, a function $\delta(x)$ satisfying the two conditions:

$$(1) \qquad \delta(x) = \begin{cases} +\infty & , \text{ if } x = 0 \\ 0 & , \text{ if } x \neq 0 \end{cases}$$

$$(2) \qquad \int_a^b \delta(x)\, dx = 1 \text{ whenever } a < 0 < b.$$

Condition (1) defines, by itself, a function in the real axis – the function that assumes the value $+\infty$ at the point 0 and the value 0 at any point $x \neq 0$. But condition (2) in conjunction with (1), is contrary to any theory of the integral, either classical or modern. In fact the function $\delta(x)$ defined by (1) is integrable, in the Lebesgue sense, on any interval of $I\!R$, but its integral is always zero, since $\delta(x)$ differs from the zero function on $I\!R$ only at the point zero.

It might be attempted to define a new concept of integral in order to satisfy condition (1) along with (2). But this is impossible without renouncing the most elementary properties of the integral. For example, we have $2\delta = \delta$ since $2\delta(x) = 0$ for $x \neq 0$, and $2\delta(x) = +2\infty = +\infty$ for $x = 0$. Therefore, if $a < 0 < b$, we should have by (2)

$$\int_a^b 2\delta(x)\, dx = \int_a^b \delta(x)\, dx = 1.$$

But, on the other hand, in order to maintain the elementary property concerning the integral of a function multiplied by a constant, we should have:

$$\int_a^b 2\delta(x)\,dx = 2\int_a^b \delta(x)\,dx = 2.$$

So the integral would not be well defined if condition (2) would be satisfied.

Hence there exists no function $\delta(x)$ *satisfying at the same time* (1) *and* (2) *in any reasonable theory of integration.*

Another solution that might be tried would consist in changing the ordinary meaning of the symbol $+\infty$, according to a consistent theory in which $+c\infty$ would be distinct from $+\infty$, for any number $c \neq 1$. Indeed, some attempts have been made by using such a theory of "hyper numbers", but to our knowledge no significant progress has been obtained in this direction.

So, only one way out is left to us: *trying to define* δ, *not as a function of a real variable, but as an entity of a new kind.* For that purpose, it is convenient to come back to the intuitive considerations, which led to the preceding pseudo definition of δ.

1.2. Intuitive origin of the δ-concept

Let us consider a distribution of matter on the real axis /*R*. Then, to each bounded interval *I* in /*R* there corresponds a non-negative number *m(I)* giving the amount of mass contained in *I*. In many cases there exist a function *P(x)* of the real variable *x* such that:

1.2.1. $$m(I) = \int_I P(x)\,dx$$

for every bounded interval *I*. Then *for any continuity point x* of this function we have

1.2.2. $$P(x) = \lim_{R\to 0} \frac{m(I_R)}{R}\ , \text{ where } I_R = \left[x - \frac{R}{2},\ x + \frac{R}{2}\right].$$

The value $P(x)$ is said to be the *density* (of the mass distribution) at the point x.

Let us now consider, for example, the mass distribution on $I\!R$ consisting of one single material point of mass 1, placed at the origin, the remaining part of $I\!R$ being unprovided with matter. In this case $m(I)$ is equal to 1 or 0 according as the point 0 belongs to I or not. Assume that there exist a density function of this distribution and denote by $\delta(x)$ that function. Then we should have according to 1.2.2.

$$\delta(x) = \begin{cases} +\infty \ , & \text{if } x = 0 \\ 0 \ , & \text{if } x \neq 0 \end{cases}$$

since

$$m(I_R) = \begin{cases} 1 \ , \text{ for every } \ R > 0 \ \text{ when } \ x = 0 \\ \\ 0 \ , \text{ if } \ x \neq 0 \ \text{ and } \ 0 < R < 2|x| . \end{cases}$$

On the other hand, we should have according to 1.2.1.:

$$m[a,b] = \int_a^b \delta(x)dx = 1 \ , \ \text{ whenever } \ a < 0 < b.$$

So the density function $\delta(x)$ should satisfy conditions (1) and (2) considered in 1.1. *But we have seen that such a function does not exist.*

It might be argued that the concept of a material point is only an abstraction that is never realized in nature. But so is *every* scientific concept: nothing else but a scheme, i.e. a simplified representation that fits more or less successfully to the described situation. One of the tasks of mathematicians is to organize in a consistent theory everybody of schemes sketched by physicists so that the results obtained by such a theory may be logically correct, free from contradictions.

1.3. The theory of measures

Before the theory of distributions, *measure theory* – or better, *the theory of measures* – afforded already a simple and coherent interpretation of the δ-symbol agreeing with preceding intuitive views.

Observe that a mass distribution on $I\!R$ is not given, in general, by a function of the *real variable* x (or a function of the *point* x), but by a function $m(I)$ of the *variable interval* I. Besides, this function is additive, i.e. if I is decomposed into two or more intervals $I_1, \ldots, I_n$, mutually disjoint, then $m(I) = m(I_1) + \cdots + m(I_n)$. *It is even natural to assume that this additive property, holding for any decomposition into a finite number of intervals, extends also to any decomposition into a countable system of intervals.*

Another example of an additive function of a variable interval can be given by a *probability distribution* on $I\!R$, which assigns to each interval I the probability $p(I)$ corresponding to I. Then we always have $0 \leq p(I) \leq 1$.

A third example may be given in certain cases by a distribution of electric charges on $I\!R$. But, there, the number $q(I)$ giving the quantity of electricity contained in each bounded interval is a real number that *may be negative.*

Such concrete models suggested the abstract concept of measure. A measure μ is defined on $I\!R$ if and only if, to each bounded interval I in $I\!R$, there is assigned a real or complex number, denoted by $\mu(I)$ or μI (the μ-measure of I) so that if I is expressed as the union of a countable system of mutually disjoint intervals, $I_1, \ldots, I_n, \ldots$, then the series $\sum \mu(I_n)$ is absolutely convergent and $\mu(I) = \sum \mu(I_n)$. In particular cases, there exists a function $f(x)$ (of the real variable x) such that:

$$\mu(I) = \int_I f, \quad \text{for each bounded interval } I,$$

where $\int_I f$ denotes the integral of f over I in the ordinary sense. *In such cases giving the measure μ is quite equivalent to give the function f.* Then the function f of the variable x is uniquely extended as a function μ of the variable I, so that the measure μ may be identified with the function f. [If the interval I reduces to a single point a such that $f(a) \neq 0$, we must have at the same time $\mu(I) = f(I) = 0$. But this involves no contradiction; it should be remembered that the interval $[a, a]$ is to be distinguished from the point a itself].

In the general case such a density f does not exist (as we have seen with the δ-measure), but the preceding identification suggests, in such cases, the following definition:

1.3.1. $$\mu(I) = \int_I \mu$$

and to say that $\mu(I)$ is the integral of μ over I. Of course these are only new symbols and a new name for the μ-measure of I, but in doing so, we approach successfully the intuitions of physicists. For example, with respect to the $\delta_{(a)}$-measure we have, for each $a \in /R$, the definition:

$$\int_I \delta_{(a)} = \delta_{(a)}(I) = \begin{cases} 1\,, & \text{if } \; a \in I \\ 0\,, & \text{if } \; a \notin I \end{cases}$$

Although $\delta_{(a)}$ is not a function of a real variable, we shall sometimes denote this measure by the functional notation $\delta(x-a)$ used by physicists. This problem of notation will be discussed in chapter III [3.6]. If $a = 0$ we denote $\delta_{(0)}$ only by δ or $\delta(x)$.

1.4. Measures as derivatives of functions of a real variable

Let μ be any measure on $/R$. If we chose any point $c \in /R$ and put

$$F(x) = \begin{cases} \mu[c, x]\,, & \text{if } x \geq c \\ -\,\mu]x, c[\,, & \text{if } x < c \end{cases}$$

then we obviously define a function $F(x)$ of the real variable x on $/R$. If there exists a density function f of the measure μ, we have, in the ordinary sense:

$$F(x) = \int_c^x f(\xi)\, d\xi\,, \quad \text{for any} \quad x \in /R\,,$$

and hence $f(x)$ is the derivative of $F(x)$ at any continuity point of that function.

In general, the density function f may not exist, but the measure μ can be derived from the function F in a precise way that we shall describe in chapter III. *Thus the measure μ is always determined by the function F; on the other hand, two functions, F and G, determine the same measure μ if and only if F – G reduces to a constant function on IR.* That being so, it will be natural to say (by definition) that μ is the derivative of the functions *F, G, ...* and to write:

$$\mu = DF \ , \quad \mu = DG \ , \ \dots .$$

On the other hand it will be natural to say that *F, G, ...* are the **primitives** of μ. For example, a primitive of δ will be the function:

$$H(x) = \begin{cases} \delta[0, x] \ , & \text{if} \quad x \geq 0 \\ \delta]x, 0[\ , & \text{if} \quad x < 0 \end{cases}$$

Hence we have $H(x) = 1$ for $x \geq 0$ and $H(x) = 0$ for $x < 0$ (Heaviside's function), and we may write in the preceding sense:

$$\delta = DH .$$

A similar, but a little more elaborate example, is given by the measure:

$$\mu = 3\delta_{(2)} - 2\delta_{(4)} + \delta_{(5)}$$

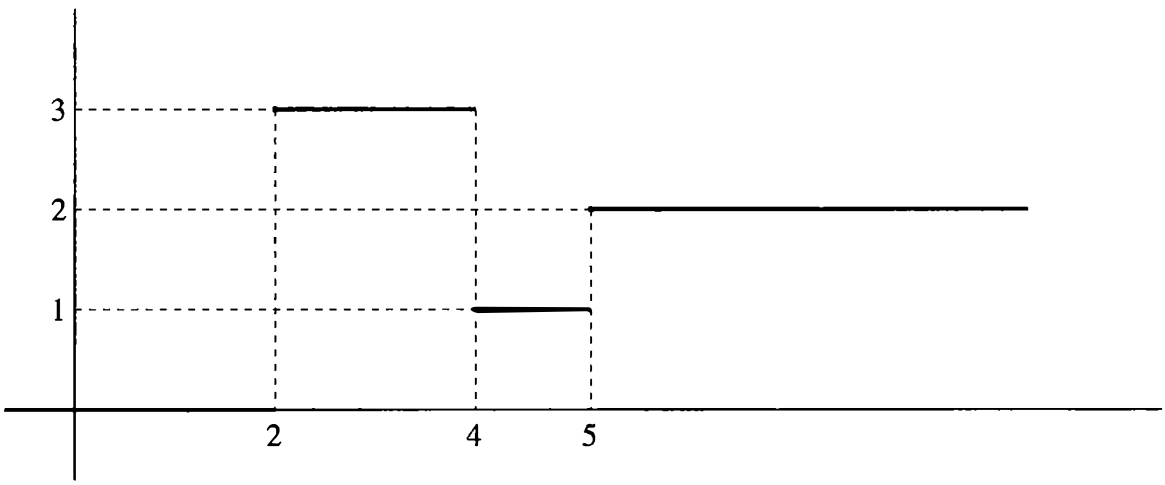
which is defined as follows:

$$\mu(I) = 3\delta_{(2)}(I) - 2\delta_{(4)}(I) + \delta_{(5)}(I) \text{ for any interval } I .$$

A concrete model of this measure is given by a system of these electric charges: $q_1 = 3$, $q_2 = -2$ and $q_3 = 1$, placed at the points 2, 4 and 5 respectively. It is readily seen that a primitive of this measure is:

$$F(x) = 3H(x-2) - 2H(x-4) + H(x-5) .$$

In this case the charges 3, -2 and 1 are given by the jumps of F at the discontinuity points 2, 4 and 5 respectively.



In general, it is proved that *the primitives of measures are the functions of bounded variation.* So the measures may be characterized as the *derivatives (in the generalized sense) of functions of bounded variation.*

1.5 Insufficiency of the theory of measures

We have just seen that the theory of measures affords a quite satisfactory interpretation of physical schemes such as the δ-concept. But there are physical schemes which may not be interpreted in terms of measures. For example, in the symbolic calculus of electricians, as well as in electromagnetic theory, the use of such entities as the derivatives of δ, which are denoted by δ', $\delta'',\ldots$, are frequent.

If δ' were a measure, then it would be the derivative of some function F of bounded variation; i. e.:

$$\delta' = D\delta = DF\,.$$

Hence, if we wanted the usual rules for derivatives to hold, we should have $D(\delta - F) = 0$ and δ should be of the form $\delta = F + C$, where C is a constant. But this is impossible since δ is not a function of a real variable. So δ' (as well as δ'', δ''', ...) can not be considered as being

a measure without renouncing the most elementary differentiation rules.

For many cases, symbols like δ', δ'',..., are only calculation devices, comparable to the imaginary numbers, which desappear after helping to find the solution of a problem. However, in other cases, these symbols are directly used for interpreting physical situations. For example, consider the system formed by two electric charges g and $-g$ placed respectively at the points $-h$ and h of $I\!R$. We then have a charge distribution on $I\!R$ that is represented by the measures $g\delta(x+h)-g\delta(x-h)$. Suppose now that $h \to 0$ and $g \to +\infty$ in such a way that $2gh$ tends to a finite limite P. Then:

$$\lim_{h\to 0} g[\delta(x+h)-\delta(x-h)] = \lim_{h\to 0}\left[2gh\,\frac{\delta(x+h)-\delta(x-h)}{2h}\right] = P\delta'(x).$$

This being so, $P\delta'(x)$ represents a kind of charge distribution that is not a measure; it is called by physicists a *dipole of momentum P* placed at the origin. Analogously, $\delta''(x)$, which may be described as the limit of the two dipole system $[\delta'(x+R)-\delta'(x)]/R$, as $R \to 0$, is interpreted as a quadripole, and so on. But, of course, since such "distributions" can not be identified with measures, they require an adequate mathematical theory *free from internal contradiction,* extending the theory of measures.

Furthermore, it must be observed that the preceding considerations, that we have related for the sake of simplicity to the real axis, $I\!R$, are extensible to any $I\!R^n$ space. For example, if $n=3$, the bounded intervals I are parallelepipeds whose edges are parallel to the three coordinate axes. Then the measure concept may be defined as we did for $I\!R$; in particular the Dirac measures of $I\!R^3$ may also be defined as we did for $I\!R$. For instance, a system of r material points of masses $m_1,\dots, m_r$, placed at r points $a_1,\dots, a_r$ of $I\!R^3$ is represented by the measure:

$$m_1\delta_{(a_1)} + \cdots + m_r\delta_{(a_r)}\,.$$

On the other hand, if x, y, z are coordinate variables, we have then Dirac measures of one variable $\delta(x)$, $\delta(y)$, $\delta(z)$. For example,

$\delta(x)\,\delta(y)$, represents the measure that assigns to each bounded interval I of $/R^3$ the area of the portion of the xy-plane that is contained in I. It is customary to consider the δ-measure on $/R^3$ as the product of the measures $\delta(x)$, $\delta(y)$, $\delta(z)$:

$$\delta(x, y, z) = \delta(x)\,\delta(y)\,\delta(z)\,.$$

Physics and the theory of probability offer many concrete examples of measures that are not functions of one point. For instance such is the case of a charge distribution on $/R^3$, which reduces to the charge on the surface of one conductor in electrostatic equilibrium; then we may have a *superficial density function,* but not a *spatial density function.*.

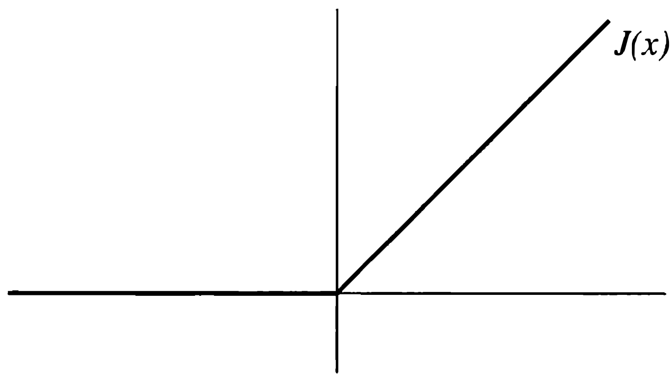
Consider now the symbol $D_z\delta(x, y, z)$. It denotes the derivatives of the δ measure on $/R^3$ with respect to z that may not be identified with a measure and is called by physicists a *dipole* with the *vector momentum* (0, 0, 1) placed at the origin.

Analogously, the symbol $\delta'(z)$ denotes the derivative of $\delta(z)$ with respect to z, which is called a *doublet* on the xy-plan; a plate conductor electrified by induction suggests a scheme of this kind.

These examples and many others that we could draw from several domains of mathematics, whether pure or applied, show the necessity of a theory of distributions.

1.6. Distributions as formal derivatives of continuous functions

Consider again the δ-measure on $/R$. We have seen that $\delta = DH$ where H is the Heaviside function assuming the value 1 for $x \geq 0$ and 0 for $x < 0$. In turn a primitive of H is the function J, such that $J(x) = x$ for $x \geq 0$ and $J(x) = 0$ for $x < 0$. Contrary to the Heaviside function, Y_1 is a continuous function on $/R$.



Observe now that, *in the sense of measure theory, H is* the derivative of J, *but in the ordinary sense,* we have only:

$$H(x) = J'(x)\,, \quad \text{for} \quad x \neq 0$$

since J, has no derivative at 0 (only the right-hand and the left-hand derivatives 1 and 0). Anyway, we have, in the sense of measure theory, $\delta = DH$ and $H = DJ$, hence $\delta = D^2J$. This suggests writing:

$$\delta' = D^3J \quad , \quad \delta'' = D^4J \quad , \ldots$$

so that δ and its derivatives (whatever they may be) come expressed as derivatives of a continuous function on $/R$.

This conclusion holds for any measure μ on $/R$. In fact, a primitive F of μ is always a function of locally bounded variation (hence integrable on any bounded interval), so that a primitive G of F is always a continuous function. Thus we have $\mu = DF = D^2G$ and in general:

$$D^n\mu = D^{n+2}G\,.$$

In this way, we are always brought back to continuous functions, which are in general more manageable than measures or functions of bounded variation. So, the problem of constructing a theory of distributions can be formulated as follows (for the case of a single variable).

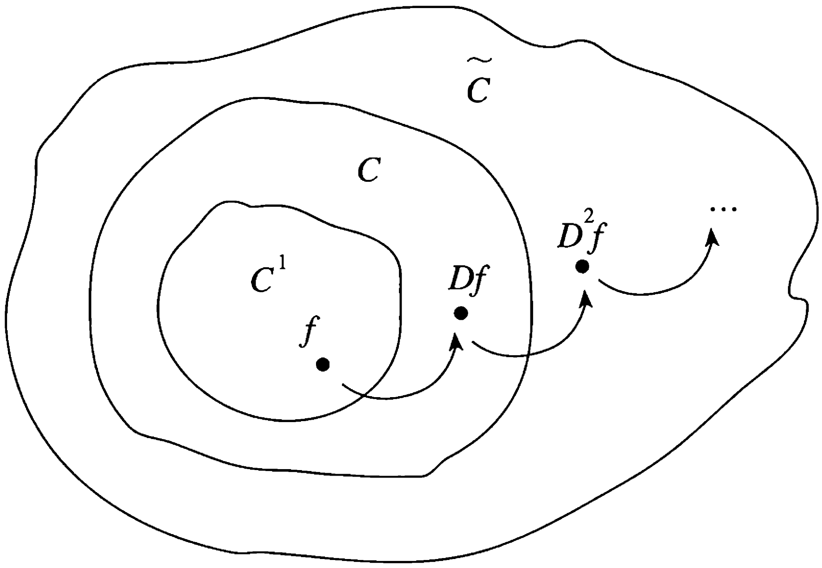
Let us consider the set C of all continuous functions on $/R$ and the set C^1 of all functions f having a continuous derivative, f', on $/R$. We have $C^1 \subset C$, but not $C^1 = C$. So, to each function f in C^1, the derivative operator D assign a function Df (or f') which may not belong to C^1. But if a continuous function does not belong to C^1, then f has no continuous derivative or even *no derivative in the ordinary sense.*

That being so, our purpose is to enlarge the set C, to a set $\tilde{C}$, by adding to it new objects in such a way that the operator D may be extended to the whole set $\tilde{C}$ and that the ordinary formal properties of D may be maintained as much as possible. The elements of $\tilde{C}$ will be called distributions regardless of whether they are functions or

not. If such a set exists, then every continuous function f will have derivatives of all orders

$$Df,\ D^2f, \ldots,\ D^nf, \ldots$$

which must be distributions.



In the next chapter this problem will be formulated more precisely and discussed in detail. For the present it is enough to emphasize that the decisive role in defining and solving this problem will be played by the purpose of maintaining the following property of the derivation operator:

If $Df = 0$, on an interval I, then f is a constant function on I.

This implies the more general property:

If $D^nf = 0$, *on an interval* I, *then* f *is a polynomial of degree less than* n, *on* I.

It would be impossible to found an useful theory of distributions without requiring such a condition. In particular all uniqueness theorems for differential equations will depend upon this property.

Obviously the problem of constructing the system $\tilde{C}$ of all distributions on $/R$ must be extended to distributions on an interval I (or even a more general point-set) in any $/R^n$ space. This will be discussed in chapter II (for $n = 1$) and in chapter VII (for $n > 1$).

CHAPTER II

DISTRIBUTIONS OF ONE VARIABLE: FUNDAMENTAL CONCEPTS.

2.1. Terminology and notation

We shall denote by $/R$ the field of real numbers and by $\mathbb{C}$ the field of complex numbers. If $a,\ b \in /R,$ with $a<b$, then

$$[a,\ b],\ ]a,\ b],\ [a,\ b[,\ ]a,\ b[$$

denote the intervals with extreme points $a,\ b$ defined respectively by the conditions

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

They are respectively **closed**, **closed on the right**, **closed on the left**, and **open**. All of them are **bounded**.

An interval (or more generally any set of points of $/R$) is said to be **compact** if it is closed and bounded. By extension of language, any set that reduces to a single point $a \in /R$ is called a **degenerate interval** that is the interval $[a,\ a]$. However, by an interval we shall always mean a non-degenerate interval, unless the contrary is explicity stated.

On the other hand, the symbols $[a,\ +\infty[$, $]a,\ +\infty[$, $]-\infty,\ a]$, $]-\infty,\ a[$ denote the **unbounded intervals** defined respectively by $x \geq a,\ x > a,\ x \leq a,\ x < a$.

The first two are bounded on the left and the last two are bounded on the right. The first is closed, the second is open, etc.

Finally the unbounded interval $]-\infty, +\infty[$ is the set $/R$ itself; it is both open and closed (in $/R$).

Let I be an interval in $/R$. We denote by $C(I)$ – or by C if there is no danger of mistake – the set of all complex-valued functions $f(x)$ of the real variable x which are defined and continuous on I. More generally, for any integer $n > 0$, we denote by $C^n(I)$ – or simply C^n – the subset of $C(I)$ formed by those functions f, which have a derivative, $f^{(n)}$, of order n continuous on I; in particular $C^0 = C$. The elements of $C^n(I)$ are said to be C^n **functions on *I***.

The term "function" will mean "complex-valued function" wherever the range is not specified.

Instead of $f^{(n)}$ we shall often write $D^n f$. This notation puts in evidence the derivation operator, D, which assigns to each function $f \in C^1$ the function $Df = f' \in C$. So D^n is the n^th^ **power** of D.

On the other hand, the symbol $\Im$ denotes an integration operator defined by the formula:

$$\Im f(x) = \int_c^x f(\xi)\, d\xi\,, \quad \text{for all } f \in C$$

where c is an arbitrary fixed point in I. Then $\Im$ is a mapping of the set C into $C^1 \subset C$ such that

$$D\Im f = f, \quad \text{for all } f \in C.$$

This $\Im$ is a **right-inverse** of D (but not a **left-inverse**). More generally:

2.1.1. $$D^n \Im^n f = f, \quad \text{for all } f \in C,\ n = 0, 1, \ldots$$

2.2. Axiomatic introduction of distributions.[1]

Let I be any interval in $/R$; the system of all distributions on I can be described by the following system of fundamental properties:

---

(1) The word "distributions" is used here with a meaning equivalent to that of "distributions of finite order" according to L. Schwartz. We shall further introduce the concept of "global distribution" which is equivalent to that of "distribution" in the sense of Schwartz.

AXIOM 1. *Every function which is defined and continuous on I is a distribution on I.*

AXIOM 2. *To every distribution f on I there corresponds one and only one distribution on I, which is called the "**derivative of f**" and denoted by Df, in such a way that, if f is a C^1 function, then Df is the derivative of f in the ordinary sense.*

DEFINITION: The derivative of order n of a distribution f, which is denoted by $D^n f$, is defined as follows:

$$D^0 f = f\,,\, D^n f = D(D^{n-1} f)\,,\ \text{for } n = 1, 2, \ldots$$

AXIOM 3. *To every distribution f on I there exists at least one integer $n \geq 0$ and one function F, continuous on I, such that $f = D^n F$.*

AXIOM 4. *If n is an integer, $n \geq 0$, and f, g are two continuous functions on I, then we have $D^n f = D^n g$ if and only if $f - g$ is a polynomial function of degree $< n$.*

We denote by $\mathbb{N}_0$ the set of all integers $n \geq 0$ and by $\mathcal{P}_n$, for each $n \in \mathbb{N}_0$, the set of all polynomial functions of degree $< n$ (restricted to I). Our immediate purpose is to prove that the preceding axioms are:

1° consistent; i.e. there exists at least one structure satisfying the axioms (a model).

2° categorical; i.e. two such models are necessarily isomorphic. This will imply that any statement about distributions on I which is not false is a consequence of the axioms, and eventually of some supplementary definitions that have been introduced in order to simplify the language.

We shall begin with the proof of categoricalness because it leads to a natural proof of consistency.

PROOF of categoricalness – Suppose that there is a model M satisfying the axioms, i.e. a set of objects $f, g, \ldots$, and an operator D such that, if we call these objects the distributions on I and Df, Dg, $\ldots$, the **derivatives** of f, $g, \ldots$, then the axioms are satisfied. The axioms 1 and 2 along with definition 1 imply that, for any couple

(n, F), where $n \in I\!N_0$ and $F \in C$, there exists one and only one distribution $f = D^n F$ (element of M).[2] Conversely, according to axiom 3, for any $f \in M$ there exists at least one couple (n, F) with $n \in I\!N_0$, $F \in C$, such that $f = D^n F$.

However there exists more than one couple (n, F) satisfying this condition. Let (m, G) be any couple such that:

2.2.1. $$D^n F = D^m G \quad (m \in I\!N_0, G \in C),$$

and let k be any integer such that $k \geq m, n$. By axioms 1 and 2, definition 1 and property 2.2.1., we have:

$$D^n F = D^k (\mathfrak{I}^{k-n} F), \quad D^m G = D^k (\mathfrak{I}^{k-m} G)$$

hence

$$D^k (\mathfrak{I}^{k-n} F) = D^k (\mathfrak{I}^{k-m} G)$$

and consequently by axiom 4:

2.2.2. $$\mathfrak{I}^{k-n} F - \mathfrak{I}^{k-m} G \in \mathcal{P}_k .$$

Conversely, axiom 4 shows that 2.2.2. implies 2.2.1. *These two conditions are therefore equivalent.* (For example, if $m \geq n$, we can choose $k = m$, so that condition 2.2.1. is satisfied by all functions G of the form $G = \mathfrak{I}^{m-n} F + P$, where $P \in \mathcal{P}_m$).

Now let us denote by $[n, F]$ the class of all couples (m, G) satisfying 2.2.2., i.e., such that $D^m G = D^n F$, and let us denote by $\tilde{C}$ the set of classes $[n, F]$, with arbitrary $n \in I\!N_0$, $F \in C$. Then the correspondance:

$$[n, F] \to D^n F$$

is obviously a one-to-one mapping of $\tilde{C}$ onto M such that:

2.2.3. $$[n+1, F] \to D(D^n F).$$

---

(2) Remember that we write C instead of $C(I)$ for the sake of simplicity.

In particular, the correspondence:

$$[0, F] \to F$$

is a one-to-one mapping of a subset C^* of $\tilde{C}$ onto C. Therefore, if we *define*:

2.2.4. $$D[n, F] = [n+1, F]$$

and we identify(3) each element $[0, F]$ of C^* with the function F itself by puting $F = [0, F] = [1, \Im F] = \cdots$, then $\tilde{C}$ becomes a second model, consistent with the axioms, isomorphic to M according to 2.2.3. and 2.2.4.

Thus any model M satisfying the axioms is isomorphic to $\tilde{C}$ and therefore, any two models M and M' are isomorphic (remember that the construction of $\tilde{C}$, based on 2.2.2., is independent of the choice of M). ♦

PROOF of consistency – We have just seen that if there is a model M of the system of axioms then the set $\tilde{C}$, described above, exists too and is also a model. We shall now prove, without assuming the existence of any previous model M, that the set $\tilde{C}$ actually exists and gives us a model of the system of axioms.

Let us consider the set $/N_0 \times C$ of all couples (n, F), where $n \in /N_0$ and $F \in C$. Given two such couples (n, F) and (m, G) we shall write:

$$(n, F) \sim (m, G)$$

if and only if there exists an integer $k \geq m, n$, such that:

2.2.5. $$\Im^{k-n} F - \Im^{k-m} G \in \mathcal{P}_k .$$

---

(3) *By identifying* $[0, F]$ *with* F, we mean in reality that the symbol "$[0, F]$" and its equivalents "$[1, \Im F]$", $[2, \Im^2 F], \ldots$, will denote from now the function F, instead of the class of couples $[0, F]$ equivalent to $(0, F)$. Thus the meaning of the symbol $\tilde{C}$ is also changed.

It is easily seen that the relation "~" just defined is *reflexive* and *symmetrical.* We now prove that it is *transitive*. Observe first that if there exists an integer $k \geq m,\ n$ satisfying 2.2.5., so does any other integer r, such that $r \geq m,\ n$. In fact we find that:

$$\mathfrak{I}^{r-n}F - \mathfrak{I}^{r-m}G \in \mathcal{P}_r$$

by applying to both members of 2.2.5. the operator D^{k-r} or $\mathfrak{I}^{r-k}$ according to whether $k \geq r$ or $k < r$. So suppose:

$$(n,\ F) \sim (m,\ G) \text{ and } (m,\ G) \sim (p,\ H).$$

Then, if we choose $r \geq m,\ n,\ p,$ we have:

$$\mathfrak{I}^{r-n}F - \mathfrak{I}^{r-m}G \in \mathcal{P}_r\ ,\quad \mathfrak{I}^{r-m}G - \mathfrak{I}^{r-p}H \in \mathcal{P}_r\,,$$

and hence, by addition:

$$\mathfrak{I}^{r-n}F - \mathfrak{I}^{r-p}H \in \mathcal{P}_r\ ,\ \text{ that is } (n,\ F) \sim (p,\ H).$$

So the relation ~ is an **equivalence relation** and, as such, it determines a partition of the set $/N_0 \times C$ off all couples $(n,\ F)$ into equivalence classes. For each couple $(n,\ F)$, we shall denote by $[n,\ F]$ the class of all couples which are equivalent to $(n,\ F)$ and we shall denote by $\tilde{C}$ the set of all of these classes (the "**quotient**" of $/N_0 \times C$ by ~).

The correspondence $[0,\ F] \rightarrow F$ being a one-to-one mapping of a subset C^* of $\tilde{C}$ onto C, we can identify each element $[0,\ F]$ of C^* with $F \in C$. Now, let us call the elements of $\tilde{C}$ **distributions on *I*.** *So Axiom 1 is satisfied by* $\tilde{C}$.

Moreover, we shall call $[n{+}1,\ F]$ the **derivative** of $[n,\ F]$ and we shall write:

$$D[n,\ F] = [n+1,\ F].$$

According to this definition, there is only one derivative for each $[n,\ F] \in \tilde{C}$. Indeed, suppose $[n,\ F] = [m,\ G]$; this means that

$(n, F) \sim (m, G)$, i.e. $\mathfrak{I}^{k-n}F - \mathfrak{I}^{k-m}G \in \mathcal{P}_k$ for any $k \geq m, n$; hence

$$\mathfrak{I}^{(k+1)-(n+1)}F - \mathfrak{I}^{(k+1)-(m+1)}G \in \mathcal{P}_{k+1},$$

that is $[n+1, F] = [m+1, G]$ which means that $D[n, F] = D[m, G]$.

Moreover if $f \in C^1$, then $D[0, f] = [1, f] = [0, f'] = f'$, since $f - \mathfrak{I}f' \in \mathcal{P}_1$. *So axiom* 2 *is satisfied.*

On the other hand we have:

$$[n, F] = D[n-1, F] = \cdots = D^n[0, F] = D^nF \text{ for every } [n, F] \in \tilde{C}.$$

So axiom 3 *is also satisfied.*

Finally, if $D^nf = D^ng$, with $f, g \in C$, then $[n, f] = [n, g]$, that is $\mathfrak{I}^{k-n}f - \mathfrak{I}^{k-n}g \in \mathcal{P}_k$, for any $k \geq n$. Choosing $k = n$, we see that *axiom* 4 *is also satisfied,* as we have $D^nf = D^ng$ if and only if $f - g \in \mathcal{P}_n$.♦

Conclusion: We have just proved that the set $\tilde{C}$ gives us a model of the preceding system of axioms. We could conceive many other such models, but this would have no essential interest since we have proved that such models are necessarily isomorphic to $\tilde{C}$. The model $\tilde{C}$ itself, after having afforded a simple proof of consistency of axiomatic system, will have no further interest.

From now on, all that matters will be the **rules of calculus of distributions:** that is, *the axioms* 1-4 *and the definitions that will be convenient to add to them, as well as the propositions implied by this system of axioms and definitions.*

In reasoning as well as in calculation, the distributions will be denoted by the notation "D^nf" or by any other that be convenient. But it will no longer be necessary to think of a distribution as a class of couples (n, f). Observe that this situation is quite similar to the one connected with the successive extensions of the number concept.

2.3. Rank of a distribution and further conventions

For each integer $n \geq 0$ we shall denote by $C_n(I)$ – or by C_n, when there will be no danger of mistake – the set of all distributions f on I of the form:

$$f = D^n F, \text{ where } F \in C(I).$$

Observe that $C_0 = C \subset C_1 \subset C_2 \subset \cdots$.

On the other hand, we shall denote by $\mathscr{D}(I)$ – or simply $\mathscr{D}$ – the set of all distributions on I. Thus $\mathscr{D}$ is the union of all the sets $C_n(I)$:

$$\mathscr{D}(I) = \bigcup_{n=0}^{\infty} C_n(I) \quad (C_0 = C),$$

and accordingly we may use the alternative notation C_∞ for $\mathscr{D}$.

2.3.1. We say that a distribution f is of **rank** n if and only if (iff) n is the least integer such that $f \in C_n$.

For example, consider the Dirac δ-distribution which can be defined as follows:

2.3.2.
$$\delta = D^2 J, \text{ with } J = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{for } x \geq 0 \end{cases}.$$

So, $\delta \in C_2$. Suppose there exist a continuous function F, such that $\delta = DF$. Then $DF = D^2 J$, that is $J = \Im F + P$, with $P \in \mathscr{P}_2$. Hence $DJ = F + P'$. But this is impossible as $F + P'$ is continuous and J has no continuous derivative (on $/R$). Consequently the rank of δ is 2.

It follows from this that $\delta^{(n)}$ is of rank $n+2$, for $n=1, 2, \ldots$.

2.3.3. An interval I is said to be the **domain** of a distribution f iff f is a distribution on I; i.e. $f \in \mathscr{D}(I)$. We also say that f is **defined** on I.

2.4. Addition of distributions

The sum $f + g$, of two distributions f, g, on the same interval I, is to be defined so as to guarantee the following properties:

A1. *If* $f, g \in C(I)$, then $f + g$ *is the sum of the functions in the ordinary sense.*

A2. *If* f, $g \in \mathscr{D}(I)$, *then* $f + g \in \mathscr{D}(I)$ *and* $D(f+g) = Df + Dg$.

Suppose:

2.4.1. $f = D^n F,\ g = D^m G$ with $n, m \in \mathbb{N}_0$, $F, G \in C(I)$.

According to the axioms (2.2), it is possible to represent f and g as derivatives of *the same order* of two continuous functions; indeed, taking $r \geq m, n,$ we have:

2.4.2. $f = D^r \tilde{F},\ g = D^r \tilde{G}$, where $\tilde{F} = \Im^{r-n} F,\ \tilde{G} = \Im^{r-m} G.$

Now, conditions A1 and A2 imply

$$f + g = D^r \tilde{F} + D^r \tilde{G} = D^r (\tilde{F} + \tilde{G}).$$

So,

2.4.3. $$f + g = D^n F + D^m G = D^r (\Im^{r-n} F + \Im^{r-m} G).$$

In this way we assign to each couple (f, g) of distributions on I, *at least* one distribution on I, which is denoted by $f + g$. We shall next prove that there is *only* one distribution $f + g$, for each couple (f, g); i.e., we shall prove the sum $f + g$ *does not depend on the representation* 2.4.1. *of f and g*. Indeed, consider another representation:

$$f = D^\nu \Phi, \quad g = D^\mu \Psi, \quad \text{with} \quad \nu, \mu \in \mathbb{N}_0, \quad \Phi, \Psi \in C.$$

Then, taking $p \geq \nu, \mu$ we get:

$$f + g = D^p (\tilde{\Phi} + \tilde{\Psi}), \quad \text{with} \quad \tilde{\Phi} = \Im^{p-\nu} \Phi, \quad \tilde{\Psi} = \Im^{p-\mu} \Psi.$$

Choose now $k \geq r, p$. Then:

$$D^r (\tilde{F} + \tilde{G}) = D^k (F^* + G^*), \text{ with } F^* = \Im^{k-n} F, \ G^* = \Im^{k-m} G.$$

$$D^p (\tilde{\Phi} + \tilde{\Psi}) = D^k (\Phi^* + \Psi^*), \text{ with } \Phi^* = \Im^{k-\nu} \Phi, \ \Psi^* = \Im^{k-\mu} \Psi.$$

But $D^k F^* = D^k (\Im^{k-n} F) = D^n F = f$ and $D^k \Phi^* = D^k (\Im^{k-\nu} \Phi) = D^\nu \Phi = f.$

Then $D^k F^* = D^k \Phi^*$ and $F^* - \Phi^* \in \mathcal{P}_k$.

Analogously $D^k G^* = D^k \Psi^*$ and $G^* - \Psi^* \in \mathcal{P}_k$; hence $(F^* + G^*) - (\Phi^* + \Psi^*) \in \mathcal{P}_k$, which means:

$$D^k(F^* + G^*) = D^k(\Phi^* + \Psi^*)\ ;\ \text{i.e.,}\ \ D^r(\tilde{F} + \tilde{G}) = D^p(\tilde{\Phi} + \tilde{\Psi}).$$

Thus, we have proved that the sum $f + g$ is uniquely defined for each couple (f, g). Besides, it is obvious that conditions A1 and A2 are actually satisfied by addition defined according to 2.4.3.. Hence addition in $\mathcal{D}(I)$ can be defined either implicity by the properties A1 and A2 or explicity by formula 2.4.3.. Moreover, this operation is:

I. *Associative:* $(f+g)+h = f+(g+h),\ \ \forall f, g, h \in \mathcal{D}$.

II. *Commutative:* $f+g = g+f,\ \ \forall f, g \in \mathcal{D}$.

III. *Reversible:* for any two distributions f, g on I, there exists one and only one distribution ξ on I, such that $f + \xi = g$.

To prove these properties, it is sufficient to represent f, g, h as derivatives of the same order of continuous functions and to apply the corresponding properties of addition in C.

The preceding properties I, II, III along with the existence and uniqueness of $f + g$ in $\mathcal{D}$, for all $f, g \in \mathcal{D}$, can be expressed shortly by saying:

2.4.4. *$\mathcal{D}$ is a commutative group with respect to addition.*

2.5. Multiplication by complex numbers

The product, αf, of a complex number α by a distribution f is to be defined so as to guarantee the two following properties:

P1. – *If $f \in C(I)$, then αf is the product of α by f in the ordinary sense.*

P2. – *If $f \in \mathcal{D}(I)$, then $\alpha f \in \mathcal{D}(I)$ and $D(\alpha f) = \alpha(Df)$.*

Suppose $f = D^n F$, with $n \in \mathbb{N}_0$, $F \in C$. Then P1 and P2 imply the explicit definition:

$$\alpha f = \alpha D^n F = D^n(\alpha F),\ \text{with}\ \alpha F \in C(I).$$

Thus to each couple (α, f), where $\alpha \in \mathbb{C}$ and $f \in C(I)$, there is assigned *at least* one distribution on I, which is denoted by αf. It is easily seen that the product αf is *unique* for each couple (α, f); i.e., *does not depend on the representation of* f. Moreover, it is quit trivial to prove that if $f, g \in \mathscr{D}(I)$ and $\alpha, \beta \in \mathbb{C}$, then:

$$\left.\begin{array}{ll} \text{I.} & \alpha(f+g)=\alpha f+\alpha g \\ \text{II.} & (\alpha+\beta)f=\alpha f+\beta f \end{array}\right\} \textit{(distributive laws)}$$

III. $(\alpha\beta)f=\alpha(\beta f)$ (*associative law*)

IV. $1 \cdot f = f$.

We have seen (2.4.4.) that $\mathscr{D}(I)$ is a module, i.e., a commutative group with respect to addition. As usual, this fact along with properties I-IV, can be expressed by saying:

2.5.1. *$\mathscr{D}(I)$ is a vector space over $\mathbb{C}$ (or a complex vector space).*

On the other hand, the conjunction of the properties $D(f+g)= =Df+Dg$ and $D(\alpha f) = \alpha Df$ is equivalent to the property:

$$D(\alpha f+\beta g) = \alpha Df+\beta Dg, \;\; \forall \alpha, \beta \in \mathbb{C};\;\; f, g \in \mathscr{D}(I)$$

and it may be expressed by saying:

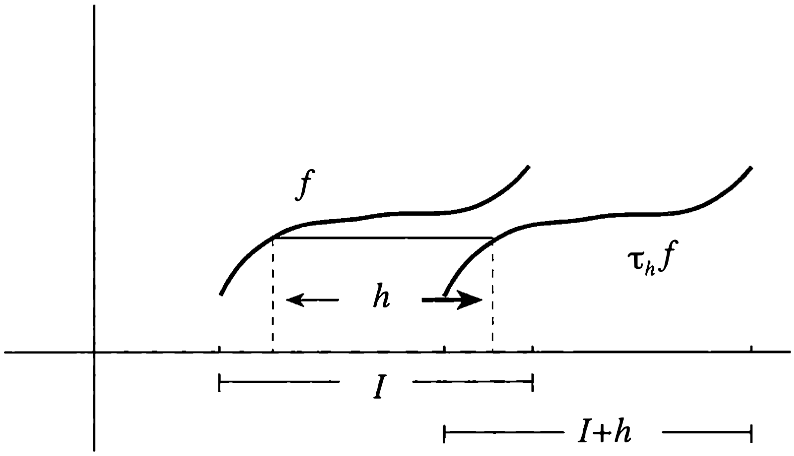
2.5.2. *The operator D is a linear mapping of the space $\mathscr{D}(I)$ into itself.*

We shall further be concerned with the more delicate problem of defining the product of two distributions.

2.6. Translation operators

If $f \in C(I)$, $h \in IR$, then $\tau_h f$ is the function defined as follows:

2.6.1. $$\tau_h f(x) = f(x-h)$$



The graph of $\tau_h f$ is that of f translated by an amplitude h. In particular, the domain of $\tau_h f$ is the interval $J = I + h$. If $h \neq 0$, we have $I=J$ if and only if $I = \mathbb{R}$.

Thus τ_h denotes a one-to-one mapping of $C(I)$ onto $C(J)$, which is natural to call a **translation operator**. Accordingly, we shall call $\tau_h f$ the **h-translate** of f.

Taking 2.6.1. into account it is readily seen that

$$\tau_h (Df) = D(\tau_h f) \text{ if } f \in C^1(I)$$

The extension of the operator τ_h to distributions is defined so as to generalize this property. So we set by *definition*:

$$\tau_h (D^n F) = D^n(\tau_h F), \ \forall n \in \mathbb{N}_0, \ F \in C(I).$$

It is obvious that this formula actually defines a one-to-one mapping τ_h of $\mathscr{D}(I)$ onto $\mathscr{D}(J)$ whose inverse is τ_{-h}. Besides, it is easily seen that for any $h \in \mathbb{R}$, this operation *is linear and interchangeable with D*, that is:

$$\tau_h (Df) = D(\tau_h f) \text{ for any } f \in \mathscr{D}(I).$$

For example, for $\delta = D^2 Y_1$ (cf. 2.3.2.), we have:

$$\tau_h \delta = D^2(\tau_h Y_1), \text{ with } \tau_h Y_1(x) = \begin{cases} x-h, & \text{if } x \geq h \\ 0 & \text{, for } x < h \end{cases}.$$

The distribution $\tau_h \delta$, which we shall also denote by $\delta_{(h)}$ is the **Dirac distribution at the point *h*.**

Remarks about notation. If f is a function and x a point of its domain, the symbol $f(x)$ denotes the value that f assumes at this point. When the point x is not specified, we are dealing with a variable and the expression "function $f(x)$" is generally used instead of "function f". Now, it must be remembered that this is an *abuse* of language which is certainly convenient in many situations, but which can lead to error in other cases, *especially in functional analysis.* In these cases it is advisable to adopt the convention consisting on writing the accent ^ over the variable which is then said to be an *apparent or mute variable.* So the symbols f, $f(\hat{x})$, $f(\hat{t})$, … , become equivalent. For example, the expression $3x^2+x$ is only a *variable dependent* on x; meanwhile the expression $3\hat{x}^2+\hat{x}$ denotes properly the function f defined by $f(x)=3x^2+x$, for all $x \in /R$.

These conventions can be extended to distributions. If f is a distribution on I and x a point of I, then the symbol $f(x)$ has generally no meaning for there is in general no value of a distribution at a point, as we shall see. But it is often convenient to use the symbol $f(\hat{x})$ for denoting the distribution f. *Accordingly, the distribution $\tau_h f$ may be suggestively denoted by $f(\hat{x}-h)$. In particular, we may write $\delta(\hat{x}-a)$ for $\tau_a \delta$ and more generally*

$$\delta^{(n)}(\hat{x}-a) \textit{ instead of } \tau_a \delta^{(n)}.$$

Frequently, we shall write $f(x)$ instead of $f(\hat{x})$ or f. It must be remembered however that this is an *abuse of writing*, which we can admit for the sake of simplicity, *whenever no misunderstanding is possible.*

2.7. Restrictions operators

If $f \in C(I)$, the **restriction** of f to an interval $J \subset I$ is the function f^* whose domain is J and such that:

$$f^*(x)=f(x), \text{ for all } x \in J.$$

We denote by $\rho_J f$ the function f^*, which is the restriction of f to J.

It is obvious that the symbol ρ_J denotes a linear mapping of $C(I)$ into $C(J)$, interchangeable with D, that is $\rho_J(Df)=D(\rho_J f)$, for all $f \in C^1(I)$. It is then natural to put by *definition*:

2.7.1. $$\rho_J(D^n F)=D^n(\rho_J F), \ \forall n \in /N_0, \ F \in C(I).$$

Thus the operator ρ_J becomes a linear mapping of $\mathscr{D}(I)$ into $\mathscr{D}(J)$ such that:

2.7.2. $$\rho_J(Df)=D(\rho_J f), \forall f \in \mathscr{D}(I).$$

Observe, that *the restriction operator ρ_J may reduce the rank of a distribution.* For example, the distribution $sin\,\hat{x}-3\delta+\delta'(\hat{x}-3)$ which is of rank 3 on $/R$ (cf. 2.3), becomes of rank 2 by restriction to $]-\infty, 3[$ and of rank 0 by restriction to $]-\infty, 0[$.

Another property of the restriction operators, which is easily shown, is the following:

2.7.3. *If I, J, K are three intervals such that $I \supset J \supset K$, then*

$$\rho_K f=\rho_K(\rho_J f), \ \forall f \in \mathscr{D}(I).$$

2.8. Collecting principle. Global distributions (or distributions in the sense of Schwartz)

Let I_1 and I_2 be any two intervals in $/R$ (distinct or coincident). Then:

2.8.1. DEFINITION. Two distributions, $f \in \mathscr{D}(I_1)$ and $g \in \mathscr{D}(I_2)$ are said to be **equal on an interval** $J \subset I_1 \cap I_2$ iff $\rho_J f=\rho_J g$. Then we write $f=g$ on J.

2.8.2. LEMMA. *Let I_1, I_2, be two open intersecting intervals in $/R$ and f_1, f_2, two distributions on I_1 and I_2 respectively, such that $f_1=f_2$ on $I_1 \cap I_2$. Then there exists one and only one distribution f on the interval $I_1 \cup I_2$ such that $f=f_1$ on I_1 and $f=f_2$ on I_2.*

PROOF. Suppose $f_1 = D^n F_1$ and $f_2 = D^n F_2$ with $F_1 \in C(I_1)$ and $F_2 \in C(I_2)$. Then $F_1 - F_2$ equals a polynomial P of degree $< n$ on $I_1 \cap I_2$. Therefore if we put $F = F_1$ on I_1 and $F = F_2 + P$ on I_2, we define a continuous function F on $I_1 \cup I_2$, and the distribution $f = D^n F$ satisfies the condition of the lemma. Conversely, a distribution $g = D^q G$ satisfying this condition coincides necessarily with f, as is readily seen. ♦

2.8.3. Collecting principle (1st form). *Let $I_1, \ldots, I_n$ be n open intervals whose union is again an interval I, and let $f_1, \ldots, f_n$ be n distributions on $I_1, \ldots, I_n$ respectively, satisfying the conditions $f_j = f_k$, on $I_j \cap I_k$, whenever $I_j \cap I_k$ is not empty (j, $k = 1, 2, \ldots, n$). Then there exists one and only one distribution f on I such that $f = f_j$* on *I_j for $j = 1, \ldots, n$.*

PROOF. We can suppose the intervals $I_1, \ldots, I_n$ ordered in such a way that if $j < k$, then the left extremity of I_j precedes the left extremity of I_k or, if these extremities are coincident, the right extremity of I_j precedes that of I_k. Then, the successive unions $I_1 \cup I_2$, $(I_1 \cup I_2) \cup I_3$, ... are again intervals, and we can achieve the proof by repeated application of the lemma along with 2.7.3. ♦

2.8.4. Remark. The conclusion is no longer true, if we consider an infinite system of distributions $f_1, f_2, \ldots$ on intervals $I_1, I_2, \ldots$ instead of a finite system. For example, take: $I_n =]-n, n[$ and $f_n = \rho_{I_n}[\delta(\hat{x}) + \delta'(\hat{x}-1) + \cdots + \delta^{(n-1)}(\hat{x}-n+1)]$ for $n = 1, 2, \ldots$. Since the rank of f_n is $n+1$ for each n, *there is no bound for the ranks of* $f_1, f_2, \ldots$, and, therefore, there is no distribution f on */R,* such that $f = f_n$ on I_n, for every n.

But we can extend the collecting principle in the following way:

2.8.5. Collecting principle (2nd form). *Let A be any set of objects. Suppose that to each $\alpha \in A$ is assigned an open interval I_α and a distribution f_α on I_α, in such a way that:*

(i) the union of all these intervals is again an interval I;

(ii) whenever two of these intervals, I_α and I_β, intersect, then $f_\alpha = f_\beta$ on $I_\alpha \cap I_\beta$;

(iii) *there exists an integer γ such that the rank of f_α is $\leq \gamma$ for every $\alpha \in A$.*

Then, there exists one and only one distribution f on I such that $f = f_\alpha$ on I_α for every $\alpha \in A$.

PROOF. The open interval I can be expressed as the union of a sequence of compact intervals $K_1 \subset K_2 \subset \cdots$. Now, according to the Heine-Borel principle, there exists for each n a finite system of open intervals $J_n^1, \ldots, J_n^{p_n}$, belonging to the given system (I_α) and *covering* K_n that is such that:

$$K_n \subset J_n, \text{ where } J_n = J_n^1 \cup \cdots \cup J_n^{p_n}.$$

Besides, these intervals may be chosen so that $J_n \subset J_{n+1}$, for any n. Then I is again the union of the increasing sequence of intervals J_n.

Let g_n^k be the distribution of the system (f_α), assigned to J_n^k, for every $n=1, 2, \ldots$, and $k=1, \ldots, p_n$. According to 2.8.3., there is one and only one distribution g_n on J_n such that $g_n = g_n^k$ on J_n^k for any n and k. On the other hand, by the hypothesis (iii), there exists necessarily for each n a function $G_n \in C(J_n)$, such that $g_n = D^\gamma G_n$. Since $g_{n+1} = g_n$ on J_n for every n, it is easily seen that we can choose the functions G_n so that $G_{n+1} = G_n$ on J_n for each n. But then it is obvious that there exists one and only one function $G \in C(I)$ such that $G = G_n$ on J_n for $n = 1, \ldots$. Consequently, if there exists a distribution f on I such that $f = f_\alpha$ *on* I_α for every $\alpha \in A$, then necessarily $f = g_n$ on J_n for $n = 1, 2, \ldots$ and therefore $f = D^\gamma G$.

Conversely, the distribution $g = D^\gamma G$ satisfies the condition $g = f_\alpha$ on I_α for every $\alpha \in A$. In fact, if we denote by g_α the restriction of g to I_α and if we put $J_{n_\alpha}^k = J_n^k \cap I_\alpha$, whenever this intersection is not empty, then I_α is the union of all $J_{n_\alpha}^k$ and we have $g_\alpha = f_\alpha$ on each interval $J_{n_\alpha}^k$. Hence, by the uniqueness property just proved, $g_\alpha = f_\alpha$; i.e. $g = f_\alpha$ on I_α (for every α). So the proof is concluded. ♦

Condition (iii) is obviously necessary in theorem 2.8.5. However, the preceding example (2.8.4.) suggests a generalization of the concept of distribution. *First of all we shall consider more generally open sets in /R instead of open intervals.* Remember that every open set Ω in /R, is the union of a finite or countable system of mutually disjoint open intervals (the so-called **components** of Ω). For the

same purpose, we could consider still more generally any set which results from any open set Ω by adding one or more *boundary points* to Ω. But this case reduces to the preceding one, as we shall see.

2.8.6. DEFINITION. Let Ω be any open set in $I\!R$ and let us suppose that to each compact interval $I \subset \Omega$ is assigned a distribution f_I on I in such a way that for any two compact intervals I_1 and I_2 contained in Ω, we have $f_{I_1} = f_{I_2}$ on $I_1 \cap I_2$, whenever $I_1 \cap I_2$ is not empty neither degenerate. The system (f_I) of distributions defined in this way is called a **global distribution** on Ω, and Ω is called the **domain** of f. The distributions f_I are the **components** of f.

We shall denote by $\overline{\mathscr{D}}(\Omega)$ the set of all global distributions on Ω.

Given two elements, $f = (f_I)$ and $g = (g_I)$, of $\overline{\mathscr{D}}(\Omega)$, and a complex number α, we shall define $f + g$ (**sum** of f and g), αf (**product** of α by f) and Df (**derivative** of f), by the formulas:

$$f + g = (f_I + g_I), \;\; \alpha f = (\alpha f_I), \;\; Df = (Df_I).$$

It is immediately seen that $\overline{\mathscr{D}}(\Omega)$ *is a complex vector space with respect to the first two operations* and that *D is a linear mapping of* $\overline{\mathscr{D}}(\Omega)$ *into itself.*

Observe now that to each continuous function f on Ω corresponds the global distribution (f_I), where f_I is the restriction of f to I and that this correspondence is a one-to-one linear mapping of $C(\Omega)$ onto a subspace $\overline{C}(\Omega)$ of $\overline{\mathscr{D}}(\Omega)$ such that if $f \in C^1(\Omega)$, then $D(f_I)$ corresponds to the derivative of f in the usual sense. *Then, we can identify every function $f \in C(\Omega)$ with the corresponding element (f_I) of $\overline{\mathscr{D}}(\Omega)$ so that $C(\Omega)$ becomes a subspace of $\overline{\mathscr{D}}(\Omega)$.*

In particular, Ω may be an interval. Then taking 2.8.5. into account, we see that the space $\mathscr{D}(\Omega)$ of all distributions on Ω, can be identified, in the same way, with a subspace of $\overline{\mathscr{D}}(\Omega)$.

In the general case, we shall call any element f of $\overline{\mathscr{D}}(\Omega)$ of the form $f = D^n F$, with $n \in I\!N_0$ and $F \in C(\Omega)$, a **distribution** on Ω. It is easily seen that the set of all distributions on Ω, which we shall denote by $\mathscr{D}(\Omega)$, is then a vector subspace of $\overline{\mathscr{D}}(\Omega)$.

Distributions may be called **global distributions of finite rank**. According, a global distribution which is not distribution is said to be of **infinite rank**.

The concept of restriction, as well as def. 2.8.1., can be extended, in a natural way, to global distributions. Then we can also extend to global distributions the *Collecting Principle* 2.8.5., considering more generally open sets instead of open intervals and suppressing condition (iii).

It should be observed that *there exists actually global distributions of infinite rank*. An example is suggested in 2.8.4.

2.9. Carrier of distribution

Global distributions of infinite rank have rather a theoretical interest mainly connected with the functional theory of L. Schwartz. So we shall hence forth confine our discussion to distributions.

We say that a distribution f on a open set Ω in $/R$ is **null on a open set** $O \subset \Omega$ iff f equals the zero function on O.

2.9.1. LEMMA. *The union of all open sets O where a distribution is null, is again an open set Ω_0 where f is null (hence the greatest open set where f is null).*

PROOF. Let I_0 be *any component* of Ω_0. Then I_0 is an open interval which is the union of a system (I_α) of open intervals where f is null. But the zero function is also null on all intervals of the system. Hence, by the collecting principle (2.8.5.), f is equal to the zero function on I_0, and since I_0 is any component of Ω_0, it follows that f is null on Ω_0. ♦

2.9.2. DEFINITION. Let f be a distribution on an open set Ω in $/R$ and let Ω_0, be the largest *open* set where f is null. Then the set $\Omega \backslash \Omega_0$ (complement of Ω_0 in Ω) is called the **carrier** of f.

According to this definition, the carrier of f is always *closed relatively to Ω*. In particular, if $\Omega = /R$, the carrier of f is a closed set.

Examples: I – If f is a continuous function on $/R$, the carrier of f is the *closure* of the set of all points x, such that $f(x) \neq 0$. Thus the function f, such that $f(x) = sin\,x$ when $sin\,x > 0$ and $f(x) = 0$ when $sin\,x \leq 0$, is a continuous function on $/R$ whose carrier is the set of all points x such that $sin\,x \geq 0$.

II – The carrier of the distribution $3\delta + \delta'(\hat{x}+1)$ reduces to the isolated points 0 and -1.

2.9.3. Proposition. *The carrier of a distribution f on $I\!R$ reduces to a single point a if and only if f is a linear combination of derivatives of $\delta(\hat{x}-a)$, $\sum_{j=0}^{m} c_j \delta^{(j)}(\hat{x}-a)$, where m is an arbitrary integer >0 and $c_0, \ldots, c_m$ are arbitrary complex constants which do not all vanish together.*

PROOF. This condition is obviously sufficient. Let us suppose, conversely, that f is a distribution on $I\!R$ whose carrier reduces to one point a. Then f is of the form $f = D^n F$ with $F \in C(I\!R)$ and, since $f = 0$ on the set of all points $x \neq a$, F is represented by two polynomials, P_1 and P_2, of degree $<n$ for $x<a$ and for $x>a$ respectively. Hence, putting $G = F - P_1$, $P = P_2 - P_1$, we have $f = D^n G$, with $G(x) = P(x)$ for $x>a$ and $G(x) = 0$ for $x<a$. *Since F is continuous on $I\!R$,* so is G and hence $G(a) = P(a) = 0$. Consequently, P must have the form:

$$P(x) = \alpha_1 (x-a) + \alpha_2 (x-a)^2 + \cdots + \alpha_{n-1}(x-a)^{n-1}.$$

Put now, for $k = 0, 1, \ldots$

$$x_+^k = \begin{cases} x^k, & \text{if } \; x \geq 0 \\ 0\;, & \text{if } \; x < 0 \end{cases}.$$

Then we have $G(x) = \sum_{k=1}^{n-1} \alpha_k (x-a)_+^k$ and $k!\,\delta(x) = D^{k+1} x_+^k$ for $k = 1, 2, \ldots$. Consequently, putting $n-2 = m$, $(k+1)!\,\alpha_{k+1} = c_{m-k}$, we obtain

$$f = D^n G = \sum_{k=0}^{m} c_k \delta^{(k)}(\hat{x}-a). \; \blacklozenge$$

Página em branco

CHAPTER III

SPECIAL TYPES OF DISTRIBUTIONS

3.1. Locally summable functions

A function f is said to be locally summable on an open set Ω in IR iff f is summable on every compact interval contained in Ω. For example, the function $x^{-1}(x-1)^{-1/3}$ is locally summable on the interval $]0, +\infty[$ or even on the set Ω of all points $x \neq 0$; but it is not locally summable on IR, for it is not summable on any interval containing 0. On the contrary, *$\log|x|$ is locally summable (thought not summable) on IR.*

Instead of an open set, we may consider more generally any set which results from an open set by adding to it one or more of its boundary points.

If f is a function locally summable on an interval I, we shall call a **primitive** of f any function F of the form:

$$F(x) = K + \int_c^x f(\xi)\, d\xi\,, \qquad \forall x \in I$$

where c is an arbitrary point of I and K an arbitrary complex number.

From the properties of the Lebesgue integral, the following theorem is deduced:

3.1.1. *If F is a primitive of a locally summable function f on I, then F is continuous on I and has a derivative a.e. (almost everywhere) in the ordinary sense such that: $F'(x)=f(x)$, almost everywhere on I.*

It should be observed that the converse of theorem 3.1.1. is not true. There are examples of continuous functions which have a derivative a.e. in the ordinary sense on an interval I *and which are primitives of no locally summable functions on I.*

The functions which are primitives of locally summable functions are said to be **absolutely continuous**. A direct characterization of such functions was given by Vitali.

Theorem 3.1.1. suggests calling the function f **a derivative** of its primitive F. But then F would have, of course, *infinitely many derivatives* (of 1st order).

3.1.2. *Two locally summable functions f_1 and f_2 on I have the same primitive F if and only if $f_1(x)=f_2(x)$ almost everywhere on I.*

In such a case the functions are said to be **equivalent** and it is written $f_1 \sim f_2$ (on I).

It is readily seen that this is actually an equivalence relation. The class of all functions which are equivalent, in this sense, to a given function f, locally summable on I, will be denoted by $[f]$. That being so, if F is any primitive of f, it will be natural to call the class $[f]$ the **derivative of** F and to write:

$$DF=[f].$$

So the derivative of F is uniquely defined as *one class* of functions instead of a single function. On the other hand it is natural to define the sum of two such classes $[f]$ and $[g]$ and the product of $[f]$ by a complex number α according to the formulas:

$$[f]+[g]=[f+g]\ ,\quad \alpha[f]=[\alpha f]$$

It is readily seen that with these definitions the set of all such classes $[f]$ becomes a complex vector space. Finally, if f is a continuous function on I, it is natural to identify $[f]$ with f, so that $C(I)$ becomes a subspace of the preceding vector space.

However, it will be troublesome to have to speak throughout of classes of functions. To avoid this we shall use a simple device.

Let f be any locally summable function on I and let us place:

$$\widetilde{f}(x) = \frac{d}{dx}\int_c^x f(\xi)\,d\xi \quad (\text{with } c \in I).$$

Then $\widetilde{f}$ is defined *only* at the points x of I for which the preceding derivative exists in the ordinary sense. On the other hand, we have, of course:

$$\widetilde{f} \sim f \quad \text{and} \quad \widetilde{\widetilde{f}} = \widetilde{f}.$$

We shall call the operation $f \to \widetilde{f}$, **standardization**, and the functions f such that $\widetilde{f} = f$, **standard functions**. For example, the Heaviside function:

$$H(x) = \begin{cases} 1, & \text{for } x \geq 0 \\ 0, & \text{for } x < 0 \end{cases}$$

is not a standard function. By standardization of H we obtain the standardized Heaviside function:

$$\widetilde{H}(x) = \begin{cases} 1, & \text{for } x > 0 \\ 0, & \text{for } x < 0 \end{cases}$$

which is not defined at $x = 0$.

In particular, all continuous functions are standard functions. It is natural to replace any equivalence class $[f]$ by the standard function $\widetilde{f}$ belonging to this class.

From now on, when we speak of locally summable functions, it will be understood that they are standard functions. We shall denote by $\mathring{L}(I)$, or simply $\mathring{L}$, the set of all (standard) locally summable functions on I. *According to the preceding remarks,* $\mathring{L}$ *is a complex vector space and* C *is a linear subspace of* $\mathring{L}$.

3.2. Locally summable functions as distributions

Let f be any (standard) locally summable function on I and let us denote by F one primitive of f:

$$F(x)=K+\int_a^x f(\xi)\,d\xi,\ (a\in I,\ \ K\in\mathbb{C}).$$

Since F is a continuous function on I, there exists a distribution which is the derivative of F. We shall denote by F' $(=f)$ the derivative of F in the functional sense and by DF the derivative of F in the distributional sense. Subsequently, we shall identify DF with F', this identification being based on the following theorem:

3.2.1. THEOREM. *By assigning to each function $f\in\mathring{L}(I)$ the distribution $f^*=DF$ where F is a primitive of f, there is defined a one-to-one linear mapping of $\mathring{L}(I)$ into $\mathscr{D}(I)$ such that:*

(i) if f is continous on I, then $f=f^*$;

(ii) if f if absolutely continuous on I, then Df^* corresponds to f'.

PROOF. First of all, it must be observed that the distribution DF assigned to each function $f\in\mathring{L}$ does not depend on the choice of the primitive F of f. In fact, if G is another primitive of f, then $F-G$ is a constant function, so that $DF=DG$.

Now consider two functions f, $g\in\mathring{L}$; we have to prove that if, to f and g corresponds the same distribution, then $f=g$. Let F, G be primitives of f, g respectively and suppose $DF=DG$. Then $F-G\in\mathscr{P}_1$ is a constant on I, and therefore, F and G have the same derivatives in the functional sense (as a standard function), that is $f=g$.

For the remaining parts of the theorem, the proof is quite trivial. ♦

This theorem shows that we can identify every distribution DF, where *F is an absolutely continuous function,* with the locally summable function f, which is the derivative of F in the functional sense 3.1.1. We then write:

$$DF=F'=f.$$

Since every locally summable function f is a distribution, f will have derivatives of all orders (in distributions sense). *Conversely,*

every distribution may be expressed in the form $D^n f$ where $n \in /N_0$ and $f \in \mathring{L}$. For simplicity of notation, even if a locally summable function f is not a standard function, we shall denote by $D^n f$ the distribution $D^n \widetilde{f}$.

For example, the δ distribution may be defined as the derivative of the (standardized) Heaviside function, and we may write in general:

$$\delta^{(n)} = D^{n+1} H, \text{ for } n = 0, 1, \dots .$$

3.3. Functions which are not distributions and pseudofunctions

Consider for example the function $\frac{1}{\hat{x}}$. Since this function is continuous on the set of all points $x \neq 0$, it is a distribution on this open set. But it is not a locally summable function *on /R,* and as we shall next see, it may not be interpreted as a distribution on /R. This function is the derivative *in the ordinary sense* (not defined at 0) of all functions f of the form:

$$f(x) = \begin{cases} log|x| + c_1, & \text{for } x > 0 \\ log|x| + c_2, & \text{for } x < 0 \end{cases}$$

or shortly:

$$f(x) = log|x| + aH(x) + b$$

with $a = c_1 - c_2$, $b = c_2$, where c_1 and c_2 are arbitrary complex numbers. Now, contrary to $\frac{1}{\hat{x}}$, *any function f of this form is locally summable on /R, hence it is a distribution on /R,* whose derivative is:

$$Df = D\,log|\hat{x}| + a\delta$$

where the symbol $D\,log|\hat{x}|$ denotes the derivative *of the locally summable function $log|\hat{x}|$ on /R, in distributions sense.*

Thus, there exist infinitely many *distinct* distributions on $I\!R$ which are the derivatives of the functions f. But for any f, we have:

$$Df = \frac{1}{\hat{x}}, \text{ on the set of all } x \neq 0.$$

Therefore the function $\frac{1}{\hat{x}}$ is a distribution on this set, but may not be interpreted as a distribution on $I\!R$.

The distribution $D\,log|\hat{x}|$ on $I\!R$ is called the **finite part** of $\frac{1}{\hat{x}}$ and denoted by $Pf\frac{1}{\hat{x}}$. But $Pf\frac{1}{\hat{x}}$ *is not* a *function, as* $\frac{1}{\hat{x}}$ *is not a distribution.*[4]

More generally the *finite part of* $\frac{1}{\hat{x}^n}$ is defined to be the distribution:

$$Pf\frac{1}{\hat{x}^n} = \frac{(-1)^{n-1}}{(n-1)!}D^n log|\hat{x}|, \quad n = 1, 2, \dots .$$

This belongs to an important class of distributions which are called **pseudofunctions** by L. Schwartz. We shall further see other examples of pseudofunctions.

3.4. Measures and functions of bounded variation

We have already discussed the concept of measure in chapter I. It is not dificult to see that an equivalent definition is the following:

A measure μ is defined on $I\!R$ iff to every *bounded interval* J in $I\!R$ is assigned a complex number, called the μ-**measure** of J and denoted by $\mu(J)$ or μJ, in such a way that:

---

(4) The expression "finite part" is connected with the concept of finite part of certain divergent integral which L. Schwartz used for defining this distribution. Note that there is no special reason to identify the function $1/x$ with $Pf1/x$ rather than with a distribution $Pf1/x + a\delta$, with $a \neq 0$.

M1. If J is expressed as the union of two disjoint intervals J_1 and J_2, then:

$$\mu(J)=\mu(J_1)+\mu(J_2).$$

M2. If J is the union of the intervals $J_1 \subset J_2 \subset \cdots$, then:

$$\mu(J)=\lim_{n\to\infty}\mu(J_n)$$

M3. For each *bounded* interval J, there exists a positive number $m(J)$ such that for every partition of J into a finite number of intervals $J_1, J_2, \ldots, J_n$, we have:

$$\sum_{k=1}^{n}|\mu(J_k)|\leq m(J).$$

Observe that the variable interval J which we are now considering may be a degenerate interval $[a, a]$.

We can define analogously the concept of measure on any open set $A \subset /R$ or even on a more extensive class. But then we must consider bounded intervals J, such that $\bar{J} \subset A$.

3.4.1. DEFINITION. If μ is a measure on the interval I on $/R$, a primitive of μ will be any function F defined on I by putting:

$$F(x)=\begin{cases} k+\mu[c, x]\ , & \text{if } x\geq c \\ k-\mu]x, c[\ , & \text{if } x<c \end{cases}$$

where c is an arbitrary point of I and k an arbitrary complex number.

From this definition and M1 follows:

3.4.2. *$F(b)-F(a)=\mu]a, b]$, whenever $a<b$.*

On the other hand, from M1 and M2 we have:

3.4.2. $F(a)-F(a^-)=\mu[a, a]$, for all $a\in I$.

To see this we consider the case $c<a$ and it suffices to express $[c, a[$ as the union of a sequence of intervals $[c, x_n]$ such that

$c < x_1 < \cdots < x_n < \cdots < a$ and $x_n \to a$. Then by M2:

$\mu[c, a[= \lim_{x_n \to a} \mu[c, x_n] = \lim_{x_n \to a} F(x_n) - k = F(a^-) - k$. Formula 3.4.3. is analogously proved in the case $a \leq c$.

Finally, from 3.4.2. and 3.4.3. follows:

$$F(b) - F(a^-) = \mu[a, b]$$

3.4.4.

$$F(b^-) - F(a) = \mu]a, b[$$

for every pair of points a, b in I such that $a < b$. Consequently:

3.4.5. *If μ is a measure on I and if F is any primitive of μ, then μ is uniquely determined by F according to formulas* 3.4.2., 3.4.3. *and* 3.4.4.

Now we need a characterization of the functions which are the primitives of the measures on I. Let F be such a function. From M1 and M2 it can be easily deduced (as in 3.4.3.) that $F(a) = F(a^+)$ for any $a \in I$; i.e. F is *continuous on the right* at every point of I. In addition, M3 implies that to each compact interval $J = [a, b]$, there exists a number $m(J)$ such that, for every partition of J by means of points $a = x_0 < x_1 < \cdots < x_n = b$, we have:

$$\sum_{k=1}^{n} |F(x_k) - F(x_{k-1})| \leq m(J).$$

But this means that F is a function of locally bounded variation on I.

Conversely, it is easily seen that these two properties are sufficient to characterize primitives of measures. Thus:

3.4.6. *A necessary and sufficient condition for a function F on I to be a primitive of a measure μ on I, is that F be of locally bounded variation on I and continuous on the right at every point of I. Moreover two such functions F_1 and F_2 are primitives of the same measure μ if and only if $F_1 - F_2$ is constant on I.*

We shall denote by $\mathfrak{M}(I)$, or simply $\mathfrak{M}$, the set of all measures on I. The sum $\mu+\upsilon$ of two measures and the product $\alpha\mu$ of a complex number α by μ are defined by the formulas:

$$(\mu+\upsilon)(J)=\mu(J)+\upsilon(J),$$
$$(\alpha\mu)(J)=\alpha(\mu J).$$

for each bounded interval J such that $\bar{J}\subset I$.

Then $\mathfrak{M}(I)$ becomes a complex vector space.

3.5. Measures as distributions. Order of a distribution

Let I be any (non-degenerate) interval on $/R$. Observe that to each function $f\in\mathring{L}(I)$ and each bounded interval J such that $\bar{J}\subset I$, there corresponds the number $\int_J f$ and this correspondence $J\to\int_J f$ is a measure, μ_f, whose primitives are just the primitives of the function f. Thus, every function $f\in\mathring{L}(I)$ determines one measure $\mu_f\in\mathfrak{M}(I)$, and if $\mu_f=\mu_g$, then $f=g$, since f and g have the same primitives. Besides it obvious that $\mu_{(f+g)}=\mu_f+\mu_g$ and $\mu_{\alpha f}=\alpha\mu_f$ for any $\alpha\in\mathbb{C}$. Thus it is natural to identify each $f\in\mathring{L}(I)$ with μ_f so that $\mathring{L}(I)$ becomes a vector subspace of $\mathfrak{M}(I)$.

Now, remember that every function of locally bounded variation on I is Riemann integrable on each compact subinterval $J\subset I$; hence locally summable on I. Then taking 3.4.6. into account, it is easily shown that:

3.5.1. *By assigning to each measure μ on I the distribution DF, where F is any primitive of μ, there is defined a one-to-one linear mapping of $\mathfrak{M}(I)$ into $\mathscr{D}(I)$ such that if μ is a locally summable function f on I, then μ corresponds to $DF=f$.*

The proof is quite similar to the one of 3.2.1. It should however be observed that the primitive F of a measure is not in general a standard function; but since $\widetilde{F}$ is defined *only* at the continuity points of

F, with the same values, it is readily seen that the correspondence $F \to \widetilde{F}$, is one-to-one.

Recording 3.5.1., it is natural to identify every measure μ on I with the distribution DF, where F is a primitive of μ and to write $\mu = DF$. So $\mathfrak{M}(I)$ becomes a vector subspace of $\mathscr{D}(I)$ and more precisely of $C_2(I)$:

$$C \subset \mathring{L} \subset \mathfrak{M} \subset C_2 \subset \mathscr{D}.$$

For every integer $n \geq 0$, we shall denote by $\mathfrak{M}_n(I)$ the **set of all distributions** f such that $f = D^n \mu$ with $\mu \in \mathfrak{M}(I)$. It is obvious that

$$\mathscr{D} = \bigcup_1^{\infty} \mathfrak{M}_n .$$

3.5.2. DEFINITION. Given a distribution f on I the least n such that $f \in \mathfrak{M}_n$ is called the **order** of f.

For example δ, which is a distribution of rank 2 (see 2.3.) is of order 0 (i.e. is a measure). In general $\delta^{(n)}$ is of order n.

3.6. Product of a continuous function by a measure and the Stieltjes integral

Consider $f \in C(I)$ and $\mu \in \mathfrak{M}(I)$. Let J be any bounded interval such that $J \subset I$ and let P be any partition of J into a finite number of (mutually disjoint) intervals $J_1, \ldots, J_n$. Denote by $N(P)$ the greatest lenght of the intervals $J_1, \ldots, J_n$. Let x_k be an arbitrary point in J_k and put:

$$S_p(J) = \sum_{k=1}^{n} f(x_k)\, \mu(J_k) .$$

Then it is a classical result that $S_p(J)$ tends to a finite limit, $S(J)$, as $N(P) \to 0$; that is, to every $\delta > 0$, corresponds an $\varepsilon > 0$, such that:

$$N(P) < \varepsilon \text{ implies } |S(J) - S_p(J)| < \delta.$$

Moreover, it can be shown that the correspondence $J \to S(J)$ is a measure on I.

This measure is called the **product** of f by μ and denoted by $f\mu$. Thus, by definition:

$$(f\mu)(J) = S(J).$$

Previously (1.3.1.) we have adopted the convention that the μ-measure of an interval J *should be called the integral of* μ *on J and denoted by* $\int_J \mu$. Thus:

$$S(J) = (f\mu)(J) = \int_J f\mu \,.$$

Remember that $\int_J f\mu$ is usually called the Stieltjes integral of f with respect to μ. The notation $\int_J f d\mu$ is commonly used instead of $\int_J f\mu$, but that notation in the theory of distributions may induce in error. If F is a primitive of μ, it is quite natural to denote the integral of f with respect to μ by:

$$\int_J f(x)\, dF(x),$$

and since $\mu = F'$ (in the distribution sense) we could also write:

$$\int_J f(x)\, dF(x) = \int_J f(x) F'(x)\, dx = \int_J f\mu \,.$$

But then we should have:

$$\int_J f(x)\, d\mu(x) = \int_J f\mu',$$

and this is the integral of f with respect to the distribution μ' that we shall define later on.

As an example, let us calculate $f\delta$, where $f \in C(/R)$. If we consider a bounded interval J such that $0 \in J$ and a partition P of J, into intervals $J_1, \ldots, J_n$, then one and only one of these will contain 0, say J_j. Thus for every choice of $x_k \in J_k$:

$$S_p(J)=\sum_{k=0}^{n} f(x_k)\,\delta(J_k)=f(x_j)\,.$$

Therefore:

$$\lim_{N(P)\to 0} S_p(J)=\lim_{x_j\to 0} f(x_j)=f(0)\,;$$

that is $(f\delta)(J)=f(0)$, if $0\in J$. It is easily seen that $(f\delta)(J)=0$, if $0\notin J$. Hence:

3.6.1. $$f\delta=f(0)\delta.$$

More generally: $f(\hat{x})\,\delta(\hat{x}-a)=f(a)\delta(\hat{x}-a)$.

3.7. Derivatives of peace-wise smooth functions

We begin with the following proposition:

3.7.1. THEOREM. *Let f be a function on an interval $I=\,]a,b[$. Suppose that f is absolutely continuous on two intervals $]a,c[$ and $]c,b[$ and tends to finite limits as $x\to c^-$ and as $x\to c^+$. Then, denoting by f' the derivative of f in the ordinary sense (not necessarily defined at c) and putting $s=f(c^+)-f(c^-)$, we have:*

$$Df=[f']+s\delta(\hat{x}-c)\,.$$

PROOF: Since f is absolutely continuous on $]a,c[$ and $]c,b[$ and has finite limits $f(c^-)$ and $f(c^+)$ it is easily seen that $[f']$ is locally summable on I. Hence if we put:

$$g(x)=f(c^-)+\int_c^x [f'](\xi)\,d\xi\,,\qquad \forall x\in I$$

the function g will be absolutely continuous on I and:

$$g'=f',\quad g(x)=\begin{cases} f(x) & ,\ \text{for } a<x<c \\ f(x)-s, & \text{for } c<x<b \end{cases}.$$

Hence $f(\hat{x})=g(\hat{x})+sH(\hat{x}-c)$ and $Df=[f']+s\delta(\hat{x}-c)$. ♦

Consider now a similar situation concerning a finite number of points $c_1, \ldots, c_p$ in I and put $s_k = f(c_k^+) - f(c_k^-)$. Then:

$$Df = [f'] + \sum_{k=1}^{p} s_k \, \delta(\hat{x} - c_k) .$$

Suppose more generally that:

i) *f has a derivative of order $n \geq 0$ in the ordinary sense except for a set of isolated points, $c_k (k = 0, \pm 1, \pm 2, \ldots)$;*

ii) *$f^{(n-1)}$ is absolutely continuous on each subinterval of I not containing any point c_k;*

iii) *for every $k = 0, \pm 1, \ldots$ and every $j = 0, \ldots, n-1$ there exists finite limits $f^{(j)}(c_k^-)$ and $f^{(j)}(c_k^+)$.*

Then it is easily deduced from 3.7.1.:

3.7.2.
$$D^n f = [f^{(n)}] + \sum_{-\infty}^{+\infty} \sum_{j=0}^{n-1} s_k^{(n-1)-j} \, \delta^{(j)}(\hat{x} - c_k) ,$$

where $s_k^j = f^{(j)}(c_k^+) - f^{(j)}(c_k^-)$. The last term in 3.7.2. (involving eventually a sum of infinitely many distributions) denotes the distribution whose restriction to each of the compact intervals $J \subset I$ is the sum of the distributions $\sum_{j=0}^{n-1} s_k^{(n-1)-j} \, \delta^{(j)}(\hat{x} - c_k)$ where $c_k \in J$ (in finite number).

For example, it is easily seen that:

$$D^2 |x| = 2\delta$$

$$D^3 |x^2 - 1| = 4\,(\delta_{(1)} - \delta_{(-1)} + \delta'_{(-1)} + \delta'_{(1)})$$

$$D^2\,(x^{4/3} + |x|) = \frac{4}{9} x^{-2/3} + 2\delta .$$

Remarks about notation: In the preceding considerations when it has been necessary to distinguish the derivatives of a function f in the ordinary sense from its derivatives in the distributional sense, we have used the notation f' in the first case and Df in the second. But whenever no confusion is possible we shall consider $f^{(n)}$ and $D^n f$ as perfectly equivalent.

Página em branco

CHAPTER IV

MULTIPLICATION AND CHANGE OF VARIABLES

4.1. Multiplication of a C^n function by a C_n distribution

As we have seen, the concept of distribution was introduced in order to render the operation D always possible, though in a formal generalized sense. But as in the case of number theory, any advantage we gain in this direction is counterbalanced by the loss of some good properties.

Multiplication of functions on the same interval is always feasible; in particular the product of two continuous functions is again a continuous function uniquely defined. But it is not possible to define the product of two completely *arbitrary* distributions, as to guarantee a minimum of properties giving some interest to such a definition.

We shall try to define the product of two distributions f and g on an interval I in $/R$, as to guarantee, *at least,* the three following conditions:

M1. *The product of two distributions f and g on I, when it exists, is again a distribution on I (which can be denoted by fg or $f \cdot g$).*
M2. *If $f, g \in C(I)$, then fg is the product of the functions f, g in the ordinary sense.*
M3. *If the product fg and $Df \cdot g$ exist, then $f \cdot Dg$ exists and $D(fg) =$ $= f \cdot Dg + Df \cdot g$.*

We shall see next that in the case when $f \in C^n$ and $g \in C_n$, the product is implicitly defined by these conditions. More precisely:

4.1.1. THEOREM. *For any integer* $n \geq 0$, *it is possible in one single way to assign to each couple* (f, g) *where* $f \in C^n(I)$ *and* $g \in C_n(I)$, *a distribution* $fg \in C_n(I)$ *not depending on n and satisfying the following conditions*:

(i) *If* $f, g \in C(I)$, *then* fg *is the product of the functions* f, g *in the ordinary sense,*

(ii) *If* $f \in C^{n+1}(I)$ *and* $g \in C_n(I)$ *then* $D(fg) = f \cdot Dg + Df \cdot g$.

By these conditions, if $f \in C^n(I)$ *and* $g = D^n G$ *with* $G \in C(I)$ *the distribution* fg *is, for each n, uniquely defined by:*

4.1.2.
$$fg = f \cdot D^n G = \sum_{k=0}^{n} (-1)^k \binom{n}{k} D^{n-k}(f^{(k)} G).$$

PROOF. a) We shall first prove, by induction on *n,* that in order for conditions (i) and (ii) to be satisfied, the product of $f \in C^n$ by $g \in C_n$ is *necessarily* given by 4.1.2. This statement is obviously true for $n = 0$. Suppose it is true for $n \geq 0$, we prove it is also true for $n+1$. Let $f \in C^{n+1}$, $g = D^{n+1}G$, with $G \in C$; then by condition (ii):

$$D(f \cdot D^n G) = f \cdot D^{n+1} G + Df \cdot D^n G.$$

Hence

4.1.3.
$$f \cdot D^{n+1} G = D(f \cdot D^n G) - f' D^n G.$$

Now, by the induction hypothesis, we have:

$$f \cdot D^n G = \sum_{k=0}^{n} (-1)^k \binom{n}{k} D^{n-k}(f^{(k)} G)$$

and

$$f' \cdot D^n G = \sum_{k=0}^{n} (-1)^k \binom{n}{k} D^{n-k}(f^{(k+1)} G).$$

Hence by substitution in 4.1.3. and by applying the well-know property $\binom{n}{k}+\binom{n}{k-1}=\binom{n+1}{k}$ we find:

$$f\cdot D^{n+1}G=\sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k}D^{n+1-k}(f^{(k)}G).$$

So the statement is true for $n+1$ and consequently for all $n\geq 0$.
b) We now prove that for each n the product fg with $f\in C^n$ and $g\in C_n$, is uniquely defined by 4.1.2. Suppose $g=D^nG=D^nG^*$; then $G-G^*$ is a polynomial function P of degree $<n$ and:

$$f\cdot D^nG-f\cdot D^nG^*=\sum_{k=0}^{n}(-1)^k\binom{n}{k}D^{n-k}(f^{(k)}P)=f\cdot D^nP=0.$$

c) Now we prove that fg does not depend on *n*. Suppose $g=D^nG=$ $=D^{n+1}\widetilde{G}$, with $D\widetilde{G}=G$ (in the ordinary sense). Then:

$$f\cdot D^nG=\sum_{k=0}^{n}(-1)^k\binom{n}{k}D^{n-k}(f^{(k)}G)$$

and since $f^{(k)}G=f^{(k)}D\widetilde{G}=D(f^{(k)}\widetilde{G})-f^{(k+1)}\widetilde{G}$, we find as we did in a):

$$f\cdot D^nG=\sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k}D^{n+1-k}(f^{(k)}\widetilde{G})=f\cdot D^{n+1}\widetilde{G}.$$

d) Condition *(i)* is obviously satisfied if we define fg by 4.1.2., $n=0$. As to condition *(ii)*, it is also implied by 4.1.2., as can easily be proved by applying the property of the binomial coefficients as we did in a). ♦

Therefore, in the case $f\in C^n(I)$ and $g\in C_n(I)$, the natural definition of the product is given by 4.1.2.

The conditions formulated in 4.1.1. (equivalent to M1, M2 and M3 in this particular case) were taken as a minimal request, in order to define implicity the product in this case. This product has most of the properties of the ordinary product except that it does not exist for all couples of distributions.

4.1.4. THEOREM. *Given any integer $n \geq 0$, any two functions $\varphi, \psi \in C^n(I)$ and any two distributions $f, g \in C_n(I)$ we have:*

(j) $$(\varphi+\psi)f = \varphi f + \psi f$$

(jj) $$\varphi(f+g) = \varphi f + \varphi g$$

(jjj) $$\varphi(\psi f) = (\varphi\psi) f.$$

PROOF. For (j) and (jj) the proof is immediate. As for (jjj) observe that, to each pair of functions $\varphi, \psi \in C^n$ and to each distribution $f \in C_n$ there is assigned the product $\varphi(\psi f) \in C_n$ is such a way that:

(i) If $f \in C$, then $\varphi(\psi f)$ is the product $(\varphi\psi) f$ in the ordinary sense.

(ii) If $\varphi, \psi \in C^{n+1}$ then:

$$D[\varphi(\psi f)] = (\varphi\psi) Df + D(\varphi\psi) \cdot f.$$

By theorem 4.1.4., this is possible only if $\varphi(\psi f) = (\varphi\psi) f$. So the proof is finished. ♦

Observe that, for any $n \geq 0$, the product $\varphi\psi$ is defined, for every couple $\varphi, \psi \in C^n$ and belongs again to C^n; moreover, this operation is associative and distributive (with respect to addition) and commutative. Thus for $n = 0, 1, \ldots$, C^n is a **commutative ring**. But C^n is also a **vector space** over the field $\mathbb{C}$, and multiplication of vectors $f \in C^n$ by scalars $\lambda \in \mathbb{C}$ is related with multiplication of two vectors $f, g \in C^n$ according to the rules:

$$\lambda(fg) = (\lambda f) g = f(\lambda g).$$

All these facts can be expressed by saying that the ring C^n is a **commutative algebra** over $\mathbb{C}$.

On the contrary, C_n, for $n > 0$, is not a ring, since the product fg does not exist for all couples of distributions $f, g \in C_n$. But C_n is a complex vector space and, on the other hand, *there exists one and only one product φf for each $\varphi \in C^n$ and $f \in C_n$, with properties (j), (jj) and (jjj).* The conjunction of all these facts can be expressed by saying:

4.1.5. *For each n, $C_n(I)$ is a module over the complex algebra $C^n(I)$.*

We denote by $C^\infty(I)$, or simply C^∞, the set of all infinitely differentiable functions on I. Then C^∞ is the intersection of the C^n and it *is again a complex algebra.*

On the other hand, we have adopted the symbol $C_\infty(I)$, or simply C_∞, as an alternative notation for the set $\mathscr{D}(I)$ of all distributions on I. So C_∞ is the *union* of all vector spaces C_n.

As it follows from 4.1.5.:

4.1.6. COROLLARY. $C_\infty(I)$ is a module over the complex algebra $C^\infty(I)$.

Observe now that multiplication by complex numbers can be interpreted as a particular case of multiplication by C^∞ functions. In fact, to each $\lambda \in \mathbb{C}$ corresponds a constant function $\tilde{\lambda} \in C^\infty$, defined on *any interval I* by:

$$\tilde{\lambda}(x) = \lambda \text{ for all } x \in I.$$

It is obvious that the correspondence $\lambda \to \tilde{\lambda}$ is a one-to-one mapping of $\mathbb{C}$ onto a subset $\tilde{C}$ of C^∞ such that if $\lambda, \mu, \nu \in \mathbb{C}$, then:

$$\lambda = \mu + \nu \Leftrightarrow \tilde{\lambda} = \tilde{\mu} + \tilde{\nu}$$
$$\lambda = \mu\nu \Leftrightarrow \tilde{\lambda} = \tilde{\mu}\tilde{\nu}.$$

Moreover $\lambda f = \tilde{\lambda} f$ for every $f \in \mathscr{D}$.

Thus we can identify each number $\lambda \in \mathbb{C}$ with the corresponding function $\tilde{\lambda} \in C^\infty$ so that the field $\mathbb{C}$ becomes a subalgebra of C^∞. The number 1 *is identified with the constant function* 1 *which is the unity element of C^∞.*

4.2. Extensions of the preceding concept of product. Examples

We have seen previously (3.6) how the product of a continuous function by a measure is defined. We have defined the vector space $\mathfrak{M}_n(I)$ of *all distributions of order $\leq n$ on I.*

Then we can replace condition (i) of the theorem 4.1.1. by the stronger one:
(i′) If $f \in C(I)$, $g \in \mathfrak{M}(I)$. then fg is the product of the continuous function f by the measure g as previously defined.

That being so, it is readily seen that theorems 4.1.1. and 4.1.4. can be immediately extended, by replacing $C_n(I)$ by $\mathfrak{M}_n(I)$, $C(I)$ by $\mathfrak{M}(I)$ and (i) by (i'). Thus:

4.2.1. *For each n, $\mathfrak{M}_n(I)$ is a module over the complex algebra $C^n(I)$.*

We can analogously define the product of a function f such that $f^{(n)} \in \mathfrak{M}$, by a distribution $g \in C_n$. Then $fg \in \mathfrak{M}_n$, but property (jjj) in 4.1.4. requires, in the present case, the additional assumption that $(\varphi\psi)^{(n)} \in \mathfrak{M}$ in addition to the hypothesis $\varphi^{(n)}$, $\psi^{(n)} \in \mathfrak{M}$ (which replaces the hypothesis $\varphi, \psi \in C^n$).

Another similar possibility concerns the product of a function f such that $f^{(n)} \in L^2$ by a distribution $g = D^n G$, with $G \in L^2$. Then fg is of the form $D^n \Phi$, with $\Phi \in L^1$.

Other variations can be imagined in a similar way. We have considered the product of a function by a distribution as though it were not commutative, but it is obvious that we did so only for the sake of convenience.

4.2.2. *In the preceding definitions of products, the order does not matter; i.e., the product is commutative.*

We reach another natural extension of the concept of product by trying to satisfy the following supplementary condition, which is of course satisfied in the preceding cases:

M4. *If f and g are two distributions on an interval I and if fg exists, then the product of their restrictions to every subinterval J of I exists and:*

$$\rho_J(fg) = (\rho_J f)(\rho_J g).$$

Suppose I is represented as the union of a system (I_α) of open intervals. Denote by f_α and g_α the restrictions of f and g respectively

to I_α, and suppose that $f_\alpha\, g_\alpha$ exists *according to one of the preceding definitions,* (regardless of order). Then, placing $f_\alpha\, g_\alpha = h_\alpha$, one easily sees that:

1) $h_\alpha = h_\beta$ on $I_\alpha \cap I_\beta$;

2) there exists an integer υ such that the rank of h_α is less than υ for all α. Therefore by the Collecting principle (2.8.) there exists one (and only one) distribution h such that $h = h_\alpha$ on each I_α. It is natural to place $h = fg$ and it is readily seen that this new concept of product satisfies M4 as well as the preceding properties of the product holding for each interval I_α.

We can, of course, consider any open subset Ω of $/R$ instead of the interval I, and even two global distributions instead of simple distributions.

We have already seen (3.6.1.) that $f\delta = f(0)\delta$ for any continuous function f on $/R$. Suppose now that $f \in C^n(/R)$. Then by formula 4.1.2. (with $G \in \mathfrak{M}$) and 3.6.1., we obtain:

4.2.3. $$f\delta^{(n)} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} f^{(k)}(0)\,\delta^{(n-k)}.$$

More generally, for any $a \in /R$:

4.2.4. $$f\delta^{(n)}(\hat{x}-a) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} f^{(k)}(a)\,\delta^{(n-k)}(\hat{x}-a).$$

Observe now that condition $f \in C^n(/R)$ is not necessary for the existence of $f\delta^{(n)}(\hat{x}-a)$; for that is sufficient, according to the last extension, that the *restriction of f to some neighborhood of* 0 *be a C^n function.* (This condition can even be enlarged by using the concept of value of a distribution at a point defined later on). In particular:

$$\delta^{(m)}(\hat{x}-a)\,\delta^{(n)}(\hat{x}-b) = 0 \text{ for } a \neq b,$$

and for $a=b$, this expression has no meaning according to the preceding definitions. However physicists frequently consider such products as $\delta\delta$, $\delta\delta'$, etc.

4.3. Impossibility of defining an associative multiplication for arbitrary distributions

We are going to show that:

4.3.1. *It is impossible to assign to every couple* (f, g) *of distributions a distribution* fg *as to satisfy the following:*

(i) *If* $f, g \in C(I)$ *then* fg *is the ordinary product,*

(ii) $D(fg) = Df \cdot g + f \cdot Dg\,,\ \ \forall f, g \in \mathscr{D}'(I)$,

(iii) $(fg)h = f(gh)\,,\ \ \forall f, g, h \in \mathscr{D}'(I)$,

PROOF. Suppose we have a multiplication satisfying *(i)* and *(ii)*, and consider the distribution $Pf\dfrac{1}{x} = D\,log|x|$ (cf. 3.3.). Then

$$\left(Pf\frac{1}{x}\right)\cdot x = (D\,log|x|)x = D(x\,log|x|) - log|x| =$$
$$= D(x\,log|x|) - D(x\,log|x| - x) = 1.$$

Hence:

$$\left[\left(Pf\frac{1}{x}\right)\cdot x\right]\delta = 1\cdot\delta = \delta.$$

On the other hand, we have $x\cdot\delta = 0\cdot\delta = 0$ and therefore $\left(Pf\dfrac{1}{x}\right)(x\delta) = 0.$

Consequently, conditions *(i)* and *(ii)* imply $\left[\left(Pf\dfrac{1}{x}\right)\cdot x\right]\delta \neq \left(Pf\dfrac{1}{x}\right)(x\delta)$, which contradicts *(iii)*.

This argument works for any interval I containing 0, if we consider the restrictions of $Pf\dfrac{1}{x}$ and δ to *I, and can be extended to any interval in IR by a suitable translation of* $Pf\dfrac{1}{x}$ and δ. So 4.3.1. is proved. ♦

It is clear that 4.3.1. continues to be true if we consider only the space of all distributions of the form $f=DF$, with $F\in L(I)$ instead of the space $\mathscr{D}(I)$.

It should be observed that difficulties connected with the concept of product are already found in the space $L(I)$ since the product of two locally summable functions would not be locally summable. For example, the square of the locally summable function $\frac{1}{\sqrt{x}}$ is $\frac{1}{x}$, *which corresponds to infinitely many distributions on $/R$.*

H. KÖNIG proved that it is possible to construct in infinitely many ways an extension $\widetilde{\mathscr{D}}(I)$ of $\overline{\mathscr{D}}(I)$, with the linear operator D, so that to each pair (f, g) of elements of $\overline{\mathscr{D}}(I)$, is assigned an element fg of $\widetilde{\mathscr{D}}(I)$ as to satisfy some conditions like M2, M3 and M4. However it must be observed that:

(I) *in such an extension the product of two elements of $\overline{\mathscr{D}}(I)$ is not necessarily in $\overline{\mathscr{D}}(I)$,*

(II) *the product of two elements of $\widetilde{\mathscr{D}}(I)$ does not in general exist,*

(III) *multiplication is not associative,*

(IV) *it is not possible to find, among all extensions, one that is "minimal" up to an isomorphism.*

Hence there is an irreducible indetermination in defining the product of two distributions.

4.4. Linear differential operators

Let $\alpha_0, \alpha_1, \ldots, \alpha_n$ be $n+1$ functions on O, an open set in $/R$. The linear differential operator $\sum_{j=0}^{n} \alpha_j D^j$ (of order n) is usually defined by the formula:

4.4.1. $$\left(\sum_{j=0}^{n} \alpha_j D^j\right) f(x) = \sum_{j=0}^{n} \alpha_j(x) f^{(j)}(x),$$

where f is any function having a derivative on O, in the ordinary sense, of order n. It is obvious that the same operator can be extended to any distribution f on O for which all terms exist. In particular, whenever the $\alpha_j \in C^\infty$; Then, the sum and product of two operators A and B of this form, defined by

$$(A+B)f = Af + Bf, \quad (AB)f = A(Bf), \quad \forall f \in \mathscr{D}(O),$$

are again operators of the same form. Moreover, it is easily seen that:

4.4.2. *The set Ω, of all operators of the form $\sum_{j=0}^{n} \alpha_j D^j$ with $\alpha_j \in C^\infty(O)$, is a complex non-commutative algebra.*

The n^th power, A^n, of an operator $A \in \Omega$ is defined by:

$$A^0 = 1 \text{ (identity)}, \; A^n = A \cdot A^{n-1}, \text{ for } n = 1, 2, \dots .$$

As a commutative sub-algebra of Ω, it should be mentioned the algebra Ω^* of all linear differential operators of finite order with constant coefficients, i.e., the algebra of all operators of the form

$$\sum_{j=0}^{n} \alpha_j D^j, \text{ where } \alpha_j \in \mathbb{C}.$$

As we did for functions (4.1.6.), we can identify each $\lambda \in \mathbb{C}$ with the operator λD^0, where D^0 is the identity operator, so that $\mathbb{C}$ becomes a sub algebra of Ω^*. For example:

$$(D^2+3)f = D^2 f + 3f, \quad \forall f \in \mathscr{D}$$
$$D^2+3 = (D+i\sqrt{3})(D-i\sqrt{3}) = (D-i\sqrt{3})(D+i\sqrt{3}), \text{ etc.}$$

If the coefficients α_j of an operator $A = \sum_{j=0}^{n} \alpha_j D^j$ are not all in $C^\infty(O)$, then A is not defined (*in any sense defined until now*) on every distribution on O. In this case, instead of the space of distributions, we can conceive other extensions of the space of continuous functions by introducing new entities (which we shall call generically **para-distributions**), as to render the operator A always defined.

This method is quite similar to the one for distributions in 2.1.

Finally, we can consider linear differential operators whose coefficients are either distributions, like δ and its derivatives, or functions which are not distributions, such as $\frac{1}{x}$, $e^{\frac{1}{x}}$, etc..

4.5. Change of variable in distributions

Let us consider a *complex*-valued function f defined on an open set $O \subset IR$ and a *real*-valued function h, which maps $O^* \subset IR$ into O. Then h assigns to each point $t \in O^*$, a point $x=h(t)$ in O; in turn, f assigns to each x the complex number $f(x)=f(h(t))$. The correspondence $t \to f(h(t))$ is a complex-valued function defined on O^* which is called the **composition** of f and h and denoted by $f \circ h$. Thus $(f \circ h)(t)=f(h(t))$.

The final operation $f \to f \circ h$ is said to be the **change of variable** (or the **substitution**) defined by h. In particular, this operation is feasible for all continuous functions f on O.

Moreover if f and h are both C^1 functions, we can apply the chain rule:

$$\frac{d}{dt} f(h(t))=f'(h(t))h'(t)$$

or

4.5.1. $$(f \circ h)'=h'(f' \circ h).$$

It should be observed that in this formula, two derivation operators are involved; one operating on functions $f \in C^1(O)$ and the other on functions $h \in C^1(O^*)$. For the sake of convenience, we denote the first by D_x and the second by D_t. Thus we can write 4.5.1. as follows:

$$D_t(f \circ h)=h'(D_x f \circ h)$$

whence, supposing $h'(t) \neq 0$ for all $t \in O^*$,

4.5.2. $$(D_x f)\circ h = \frac{1}{h'} D_t (f\circ h).$$

This formula can be expressed by saying:

4.5.3. *The change of variable defined by h transforms D_x into the differential operator* $\frac{1}{h'} D_t$.

More generally, let $f = D^n F$, $F \in C(O)$, be a distribution. By 4.5.2., we are induced to write formally:

$$(D_x^n F\circ h) = \left(\frac{1}{h'} D_t\right)^n (F\circ h).$$

Justification of this is given by the following

4.5.4. THEOREM. *If $F \in C(O)$ and $h \in C$ maps O^* into O in such a way that $\frac{1}{h'} \in C^n(O^*)$, then the expression $\left(\frac{1}{h'} D_t\right)^n (F\circ h)$ denotes a distribution in $C_n(O^*)$. Moreover, if $D_x^n F = D_x^m G$, $G \in C(O)$ we have again $\left(\frac{1}{h'} D_t\right)^n (F\circ h) = \left(\frac{1}{h'} D_t\right)^m (G\circ h)$.*

PROOF. The first part is proved by induction. The statement is obviously true for $n = 0$. Suppose it is true for $n \geq 0$ and assume $\frac{1}{h'} \in C^{n+1}(O^*)$.

Then, since

$$\left(\frac{1}{h'} D_t\right)^{n+1} (F\circ h) = \frac{1}{h'} D_t \left[\left(\frac{1}{h'} D_t\right)^n (F\circ h)\right],$$

and the right side is the product of the function $\frac{1}{h'} \in C^{n+1}(O^*)$ by a distribution in $C_{n+1}(O^*)$, the statement is also true for $n+1$. Hence it is true for every integer $n \geq 0$.

The second part we can reduce to the case where O is an interval. Suppose $D_x^n F = D_x^m G$, with $m \geq n$ and place $\Phi = \Im^{m-n} F$; then $D^m \Phi = D^m G$ and $\Phi - G = P$ is in $\mathcal{P}_m$. By 4.5.2., we find

$$\left(\frac{1}{h'} D_t\right)^m (\Phi \circ h) = \left(\frac{1}{h'} D_t\right)^n [(D_x^{m-n} \Phi) \circ h] = \left(\frac{1}{h'} D_t\right)^n (F \circ h).$$

On the other hand:

$$\left(\frac{1}{h'} D_t\right)^m (\Phi \circ h) - \left(\frac{1}{h'} D_t\right)^m (G \circ h) = \left(\frac{1}{h'} D_t\right)^m (P \circ h) = (D_x^m P) \circ h = 0.$$

Hence

$$\left(\frac{1}{h'} D_t\right)^n (F \circ h) = \left(\frac{1}{h'} D_t\right)^m (G \circ h). \blacklozenge$$

4.5.5. DEFINITION. Under the hypothesis of theorem 4.5.4., we shall write:

$$f \circ h = \left(\frac{1}{h'} D_t\right)^n (F \circ h)$$

and we say that the distribution $f \circ h$, defined by this formula, is the **composition** of the distribution f with the function h. We sometimes use the notation $f(h(\hat{t}))$.

Now the following are easy to verify:

(i) *If $f \in C(O)$ and $h \in C(O^*)$, then $f \circ h$ is the compound of f with h in the ordinary sense.*

(ii) *If $f \in C_n(O)$ and $\dfrac{1}{h'} \in C^{n+1}(O^*)$, then $D_t(f \circ h) = h'(D_x f \circ h)$. (Chain rule).*

(iii) *If $f, g \in C_n(O)$, $\lambda \in \mathbb{C}$ and $\dfrac{1}{h'} \in C^n(O^*)$, then*

$$(f+g) \circ h = f \circ h + g \circ h \text{ and } (\lambda f) \circ h = \lambda (f \circ h).$$

(iv) *If* $f \in C_n(O)$ *and* h, k *are* C^1 *mappings of* O^* *into* O *and of* O^{**} *into* O^* *respectively, such that* $\frac{1}{h'} \in C^n(O^*)$ *and* $\frac{1}{k'} \in C^n(O^{**})$, *then* $(f \circ h) \circ k = f \circ (h \circ k)$. (Associative law).

Examples I. Translation operators are the most simple examples of change of variable. In fact the translation $\tau_a f$ of a distribution f is the distribution $f(\hat{t} - \mathrm{a})$, composition of $f(\hat{x})$ with the function $x = t - a$.

II. For any real $k \neq 0$, and $n \geq 0$, we have:

4.5.6.
$$\delta^{(n)}(kt) = \frac{1}{|k|k^n}\, \delta^{(n)}(t).$$

Indeed, putting $x = kt$, we find:

$$\delta^{(n)}(kt) = D_x^{n+1} H(x) = \frac{1}{k^{n+1}}\, D_t^{n+1} H(kt) = \frac{1}{|k|k^n}\, \delta^{(n)}(t),$$

since $H(kt) = H(t)$ or $H(kt) = 1 - H(t)$, depending on whether $k > 0$ or $k < 0$.

III. Let us see whether the change of variable defined by $x = t^2 - c^2$, with $c > 0$, is feasible on $\delta(x)$. The function $h(t) = t^2 - c^2$ maps $/R$ into $[-c^2, +\infty[$. Also, $h'(t) = 2t$, and $\frac{1}{h'(t)} = \frac{1}{2t}$ is a C^∞ function *on the open set* O *of all* $t \neq 0$ *in* $/R$. Hence the distribution $\delta(t^2 - c^2)$ of t is defined on O and:

$$\delta(t^2 - c^2) = D_x H(t^2 - c^2) = \frac{1}{2t}\, D_t H(t^2 - c^2).$$

Now it is easily seen that:

$$H(t^2 - c^2) = H(t - c) + H(-t - c) = \begin{cases} 0, & \text{if } -c < t < c \\ 1, & \text{if } t < -c \text{ or } t > c \end{cases}$$

Hence:

$$\delta(t^2-c^2)=\frac{1}{2t}[D_t H(t-c)+D_t H(-t-c)]=\frac{1}{2t}[D_x H(t-c)-D_x H(-t-c)]=$$

$$=\frac{1}{2t}[\delta(t-c)-\delta(-t-c)].$$

But $\delta(-t-c)=\delta(t+c)$ (cf. 4.5.6.) and by 3.6.1.:

$$\frac{1}{2t}\delta(t-c)=\frac{1}{2c}\delta(t-c),$$

$$\frac{1}{2t}\delta(t+c)=-\frac{1}{2c}\delta(t+c).$$

Consequentely:

4.5.7. $$\delta(t^2-c^2)=\frac{1}{2c}[\delta(t-c)+\delta(t+c)] \text{ for } t\neq 0.$$

REFERENCES

[1] H. KÖNIG. *Multiplication und Variablentransformation in der Theorie der Distributionen*. Arch. Math. 6 (1955).

[2] H. KÖNIG. *Multiplication von Distributionen* I. Math. Annalen 128 (1955) 420-452. (Maths. Reviews 19-935).

Página em branco

CHAPTER V

TOPOLOGIES ON SPACES OF DISTRIBUTIONS

5.1. Limit of a sequence of continuous functions

Let I be a compact interval in $I\!R$. Then, if f is any continuous function on I, the supremum of $|f(x)|$ on I is a non-negative real number, called the **norm** of $f \in C(I)$ and denoted by $\|f\|$.

$$\|f\| = \sup_{x \in I} |f(x)| = \max_{x \in I} |f(x)|.$$

With this definition, the vector space $C(I)$ becomes a normed space.

5.1.1. A sequence $f_0, \ldots, f_n, \ldots$ of vectors of $C(I)$ **converges** in norm to a vector g of $C(I)$ iff $\|f_n - g\| \to 0$ as $n \to \infty$. Then we write $f_n \to g$ **in norm** on I.

It is well-known that this notion of convergence is equivalent to that of "uniform convergence" on I. Remember that the sequence f_n is said to **converge uniformly** to g on I iff for every $\delta > 0$, there exists an integer N (depending on δ, but not on x), such that:

$$|f_n(x) - g(x)| < \delta, \text{ for all } n > N \text{ and all } x \in I.$$

Remember also, that uniform convergence is a sufficient condition for the limit operation to be interchangeable with the integral operator.

5.1.2. *If the sequence of continuous functions f_n converges uniformly on I to a (continuous) function g and if c is any point of I, then the sequence of (continuous) functions*

$$F_n(x)=\int_c^x f_n(\xi)\,d\xi \quad (n=1,2,\ldots;\, x\in I)$$

converges uniformly on I to the function $G(x)=\int_c^x g(\xi)\,d\xi$.

To see that, it is sufficient to apply the hypothesis observing that:

$$|F_n(x)-G(x)|\le\int_c^x|f_n(\xi)-g(\xi)|\,d\xi\le|I|\sup_{\xi\in I}|f_n(\xi)-g(\xi)|\le|I|\,\|f_n-g\|<|I|\frac{\delta}{|I|},$$

if n is such that $\|f_n-g\|<\dfrac{\delta}{|I|}$, where $|I|$ denotes the lenght of I.

Thus, if we put $\Im f(x)=\int_c^x f(\xi)\,d\xi$, for any $f\in C(I)$, we can express 5.1.2. by the formula:

5.1.3. $lim(\Im f_n)=\Im(lim f_n)$, whenever $f_n\to g$ in norm (or by saying: *the operator* $\Im$ *is continuous on the normed space* $C(I)$).

On the contrary, the operator D is not continuous on the subspace $C^1(I)$ of $C(I)$, *with the same norm*. For example, consider

$f_n(x)=\dfrac{1}{n}\,sin(nx)$, for $n=1,2,\ldots$, then $sup|f_n(x)|=\dfrac{1}{n}$ on $/R$ for any n,

so that f_n converges to 0 uniformly on every compact subinterval I (even on $/R$); but the sequence of derivatives $f_n'(x)=cos\ nx$ does not converge uniformly (or even point-wise) on any compact interval I.

The space $\mathscr{D}(I)$ was constructed (2.2.) in order to make the operation D^n always feasible, on continuous functions. Our next purpose is to define a suitable topology on $\mathscr{D}(I)$ *so as to render the same operation continuous.*

For that purpose, we begin with the concept of convergence of distributions.

Notation. In all subjects about limits, we shall use the symbol "$\to$" as an abbreviation of "converges to", "tends to" or "approaches".

5.2. Limits of sequences of distributions

Let us consider first a *compact* interval I on IR. The concept of convergence for sequences of vectors in the space $\mathscr{D}(I)$, is defined so as to guarantee *at least* the two following properties:

L1. *If a sequence of elements $f_0, \ldots, f_n, \ldots \in \mathscr{D}(I)$ converges to an element g of $\mathscr{D}(I)$ then the sequence of derivatives $Df_0, \ldots, Df_n, \ldots$ converges to Dg; that is*:

$$f_n \to g \text{ implies } Df_n \to Dg.$$

L2. *If a sequence of functions $f_n \in C(I)$ converges uniformly on I to a function $g \in C(I)$ then $f_n \to g$ in the distributional sense (i.e. according to the new concept which we will define).*

From L1 and L2 it follows immediately that if a sequence of functions $f_n \in C(I)$ converges uniformly on I to g then $D^p f_n \to D^p g$, for any integer p.

5.2.1. DEFINITION. We say that a sequence of distributions f_n on I **converges** (or tends) to $g \in \mathscr{D}(I)$ iff there are a fixed integer p, a sequence of functions $F_n \in C(I)$ and a function $G \in C(I)$, such that:

(i) $f_n = D^p F_n$, for all n;

(ii) $g = D^p G$;

(iii) F_n converges uniformly on I to G.

It is obvious that this definition satisfies L1 and L2. Therefore, we shall see that:

5.2.2. *If $f_n \to g$ and $f_n \to g^*$, then $g = g^*$.*

In fact suppose: $\exists p, q \in I\!N_0$, sequences $F_n, F_n^* \in C(I)$ and $G, G^* \in C(I)$ such that:

(j) $f_n = D^p F_n = D^q F_n^*$, for all n;

(jj) $g = D^p G$ and $g^* = D^q G^*$;

(jjj) F_n and F_n^* converge uniformly on I, to G and G^* respectively.

Assume for example $p \geq q$ and set

$$P_n = F_n - \Im^{p-q} F_n^* \text{ for } n = 0, 1, \dots .$$

Then by (j), every $P_n \in \mathcal{P}_p$ and according to 5.1.2. and (jjj) P_n converges uniformly on I to the function $Q = G - \Im^{p-q} G^*$. Hence $Q \in \mathcal{P}_p$ and according to (jj), $g = g^*$. ♦

Remark. We have here used the well-known property: "If a sequence of polynomial functions π_n of degree $< r$ is convergent at, at least r distinct points $x_1, \dots, x_r$, then π_n converges to a polynomial function π of degree $< r$ at every point x (even uniformly on every compact interval)". This property can be proved with the aid of the **Lagrange interpolation formula**; remember that:

$$\pi_n(x) = \sum_{k=1}^{r} \varphi_k(x)\, \pi_n(x_k), \text{ where}$$

$$\varphi_k(x) = \frac{(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_r)}{(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_r)} .$$

Then by putting $c_k = \lim_n \pi_n(x_k)$ for $k=1, \ldots, r$, we have $\lim_n \pi_n(x) = \sum_{k=1}^{r} c_k \varphi_k(x)$ for every $x \in \mathbb{R}$, and therefore the limit is the polynomial $\pi(x) = \sum_{k=1}^{r} c_k \varphi_k(x)$, of degree $< r$. On the other hand, if we put for every compact interval I, $M(I) = \max_{x,\, k} |\varphi_k(x)|$ in I, we find:

$$|\pi_n(x) - \pi(x)| \leq \sum_{k=1}^{r} M(I)\, |\pi_n(x_k) - c_k|\,, \quad \forall x \in I\,, \quad n = 1, 2, \ldots$$

and this shows that $\pi_n \to \pi$ uniformly on I.

This property can be expressed by saying: *for all integer r, the set $\mathcal{P}_r$ is closed in the normed space $C(I)$.*

5.2.3. DEFINITION. If $f_n \to g$ in $\mathscr{D}(I)$, we say that g is the **limit** of the sequence f_n and we write $g = \lim f_n$.

Observe that the uniqueness of the limit is assured by 5.2.2., which in turn is a direct consequence of definition 5.2.1.

Now it is readily seen that:

5.2.4. *If $f_n \to f^*$ and $g_n \to g^*$ in $\mathscr{D}(I)$ and if $\alpha, \beta \in \mathbb{C}$, then $\alpha f_n + \beta g_n \to \alpha f^* + \beta g^*$; hence $\lim(\alpha f_n + \beta g_n) = \alpha \lim f_n + \beta \lim g_n$. More generally, this is also true if α and β are C^∞ functions on I.*

Let us now consider any *open* set Ω in $\mathbb{R}$.

5.2.5. DEFINITION. We say that a sequence f_n in $\mathscr{D}(\Omega)$ **converges** to g in $\mathscr{D}(\Omega)$ iff for every compact interval $I \in \Omega$, the sequence of distributions $\rho_I f_n$ converges to $\rho_I g$ in $\mathscr{D}(I)$ according to definition 5.2.1.

It is a simple matter to extend properties L1, L2, 5.2.2. and 5.2.4. to this new concept. For 5.2.2. apply 2.8.5. The uniqueness property enables us to write $g = \lim f_n$ iff $f_n \to g$ also in this case.

Observe that definition 5.2.5. *extends immediately with the same properties, to the space* $\overline{\mathscr{D}}(\Omega)$ *of global distributions on* Ω (cf. 2.8). For example, we can represent by

$$\sum_{n=0}^{\infty} \delta^{(n)}(\hat{x}-n) = \lim_{N\to+\infty} \sum_{n=0}^{N} \delta^{(n)}(\hat{x}-n)$$

the global distribution considered in 2.8.4.

5.3. Convergence in the mean and convergence in distributional sense. Examples.

Let us consider again a *compact interval I*. In the vector space $L(I)$, of all summable functions on *I*, the norm $\|f\|_1$ of a vector f is usually defined by:

$$\|f\|_1 = \int_I |f(x)|\,dx.$$

A sequence of vectors f_n in $L(I)$ is said to **converge in the mean** to a vector g in $L(I)$ iff $\|f_n - g\|_1 \to 0$.

5.3.1. *If a sequence of functions* f_n *in* $L(I)$ *converges almost everywhere (a.e.) in I to g and if there exists a number M such that* $|f_n(x)| \le M$ *for all* $n=0, 1, \ldots$ *and all x in I, then* $g \in L(I)$ *and* f_n *converges to g in the mean.*

This is an immediate consequence of the Lebesgue theorem.

5.3.2. *If* f_n *converges in the mean to g, then* f_n *converges to g in distributional sense.*

PROOF. Suppose that $\int_I |f_n - g| \to 0$ and put $F_n(x) = \int_c^x f_n(\xi)\,d\xi$, $G(x) = \int_c^x g(\xi)\,d\xi$, where $c \in I$. Since

$$|F_n(x)-G(x)| \leq \int_c^x |f_n(\xi)-g(\xi)|\, d\xi \leq \int_I |f_n-g|, \quad \forall x \in I$$

it is seen that $F_n \to G$ uniformly on I. As $f_n = DF_n$ and $g = DG$, it follows that $f_n \to g$ in distributional sense. ♦

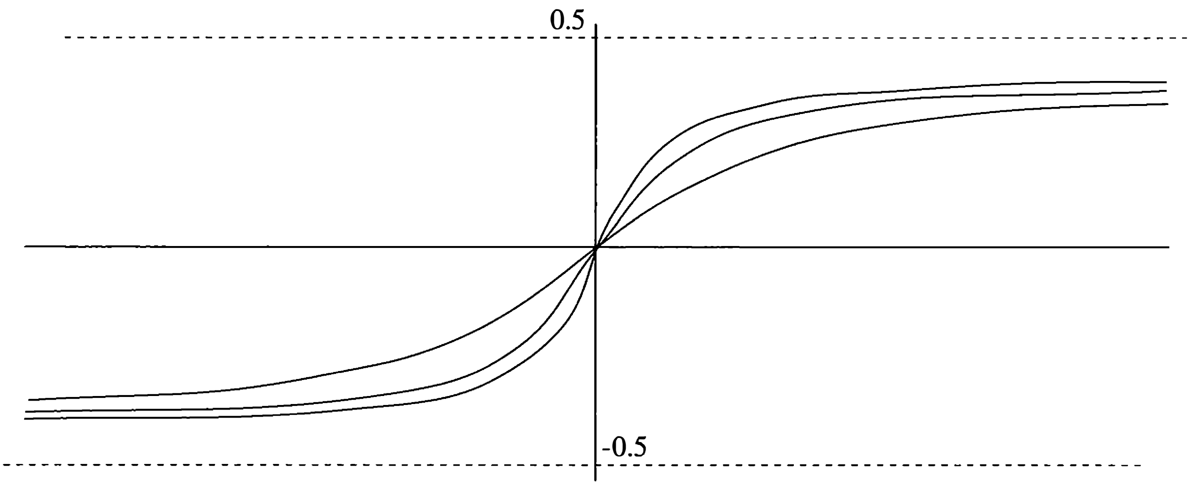
Example. Consider the sequence of functions:

$$\varphi_n(x) = \frac{1}{\pi} \frac{n}{1+(nx)^2}, \quad (n=0,\ 1, \ldots).$$

Then, if we put $\Phi_n(x) = \int_0^x \varphi_n(\xi)\, d\xi$ for $n=0, 1, \ldots$ and all x in $/R$, we have $\Phi_n(x) = \left(\frac{1}{\pi}\right) arctan(nx)$ and hence:

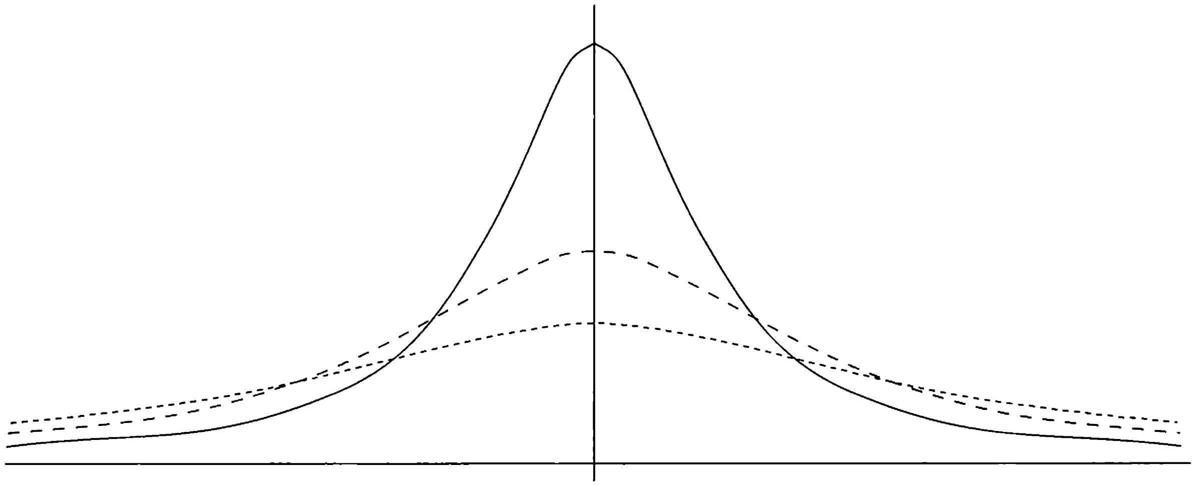
$$|\Phi_n(x)| \leq \frac{1}{2}, \text{ for } n=0,\ 1,\ \ldots \text{ and any } x \in /R,$$

$$\lim_{n\to\infty} \Phi_n(x) = \begin{cases} 1/2, & \text{if } x>0 \\ 0, & \text{if } x=0 \\ -1/2, & \text{if } x<0. \end{cases}$$



So $\Phi_n(x)$ converges at every point x of $/R$, except zero, to the function $H(x) - \frac{1}{2}$. *Therefore*, according to 5.3.1. and 5.3.2., Φ_n

converges in the distributional sense to $H(x)-\frac{1}{2}$ *on every compact interval in /R,* and hence on /R, according to definition 5.2.5. But $\varphi_n = D\Phi_n$ for every n. Hence $\varphi_n \to D\left(H-\frac{1}{2}\right)=\delta$, that is $\delta = \lim \varphi_n$.



More generally, since φ_n is a C^∞ function on /R for each n, we have:

$$\delta^{(k)} = \lim \varphi_n^{(k)}, \quad \text{for } n=0, 1, \ldots$$
$$\varphi_n \to \delta \Rightarrow \varphi_n' \to \delta' \Rightarrow \varphi_n'' \to \delta'' \Rightarrow \cdots \Rightarrow \varphi_n^{(k)} \to \delta^{(k)}.$$

So the distribution $\delta^{(k)}$ *is expressed as the limit of a sequence of* C^∞ *functions,* $\varphi_0^{(k)}, \ldots, \varphi_n^{(k)}, \ldots$.

5.4. Inductive limits; (*LN**)-spaces

Till now, we have only defined a concept of convergence for sequences of distributions. In order to define in the preceding spaces of distributions a suitable topology leading to that concept of convergence (which is however sufficient for the following chapters) we need some special notions and results concerning the theory of locally convex spaces.

Consider a vector space E, a family $(E_\alpha)_{\alpha\in A}$ of vector spaces over the same field ($\mathbb{R}$ or $\mathbb{C}$) and let φ_α be, for every $\alpha\in A$, a linear mapping of E_α into E. Suppose that, on each space E_α is defined a locally convex topology τ_α (not necessarily Hausdorff). Then it is easily seen that among all locally convex topologies for which the mappings φ_α are continuous for all α, there exists one τ^* which is stronger than all the others. (A fundamental system of neighborhoods of 0 for τ^* may be the family of all circled convex and absorbing subsets $\mathcal{V}$ of E such that for every α, $\varphi_\alpha^{-1}(\mathcal{V})$ is a neighborhood of 0 for τ_α). That being so, τ^* is said to be the **inductive limit** of the topologies τ_α.

In particular, E may be the union of all E_α and φ_α the injection (identity mapping) $E_\alpha\to E$. Then $E(\tau^*)$ is said to be the inductive limit of the spaces $E_\alpha(\tau_\alpha)$. In this particular case τ^* is the strongest locally convex topology on E, inducing on each E_α, a topology weaker than τ_α.

5.4.1. DEFINITION. A locally convex space E is called a ***(LN*)* space** iff E can be represented as the inductive limit of a sequence $E_1,\ldots, E_n,\ldots$ of normed spaces such that:
(1) $E_n\subset E_{n+1}$, for all n,
(2) the injection $E_n\to E_{n+1}$ is, for all n, compact (which means that all bounded sets in E_n are relatively compact with respect to the norm of E_{n+1}).
Such a sequence E_n is said to be **regular**.

5.4.2. LEMMA. *For every regular sequence (E_n) or normed spaces, there exists an increasing sequence (F_n) of Banach spaces such that:*
(i) every bounded closed set in F_n is compact in F_{n+1};
(ii) for all n, $E_n\subset F_n\subset E_{n+1}$ and the injections $E_n\to F_n$, $F_n\to E_{n+1}$, are continuous.

PROOF. Let E be the inductive limit of (E_n). For every n, denote by $\|\cdot\|_n$ the norm in E_n and set

$$B_n=\{x:\|x\|_n\leq 1\}\,;\ \widetilde{B}_n=\text{closure of } B_n \text{ in } E_{n+1}\,;\ \widetilde{E}_n=\bigcup_{k=1}^{\infty} k\widetilde{B}_n\,.$$

As the injection $E_n \to E_{n+1}$ is continuous for all n, we can suppose that the norms $\|\cdot\|_n$ have been chosen so that $B_n \subset B_{n+1}$; i.e. $\|x\|_n \geq \|x\|_{n+1}$ for all n. Then $B_n \subset \widetilde{B}_n \subset B_{n+1}$. Since B_n is circled and convex so is $\widetilde{B}_n$, and therefore $\widetilde{E}_n$ is a vector subspace of E_{n+1}. Besides, if we place

$$g_n(x) = inf\{\rho > 0 : x \in \rho \widetilde{B}_n\}, \ \ \forall x \in \widetilde{E}_n,$$

g_n will be a *norm* defined on $\widetilde{E}_n$ (since $\widetilde{B}_n \subset B_{n+1}$). Thus $F_n = \widetilde{E}_n$ becomes a normed space and (ii) follows immediately from the double inclusion $B_n \subset \widetilde{B}_n \subset B_{n+1}$ for all n. Moreover, every bounded closed set H in F_n will be compact in F_{n+1}, since there exists $\rho > 0$ such that $H \subset \rho \widetilde{B}_n$ and $\widetilde{B}_n$ is compact in E_{n+1} hence in F_{n+1} (which induces in E_{n+1} a weaker topology). It can be also proved that the spaces F_n are complete, but that is not required for the following applications. ♦

Observe that according to condition (ii), *the sequences* (E_n) *and* (F_n) *have the same inductive limit.*

In the following propositions, (E_n) denotes a regular sequence, E the inductive limit of (E_n), hence a (LN^*) space; $\|\cdot\|_n$ denotes the norm of E_n, $B_n = \{x : \|x\| \leq 1\}$. τ_n is the topology on E_n given by $\|\cdot\|_n$ and τ_∞ the topology of E (inductive limit of the topologies τ_n). *We can assume without loss of generality that* $B_n \subset B_{n+1}$ *for all n and that* B_n *is compact in* E_{n+1} (according to the lemma). That being so

5.4.3. THEOREM. *A set H is closed in E iff for every n,* $H \cap E_n$ *is* τ_n *closed.*

PROOF. a) Suppose H is closed in E. Since τ_n induces in each E_n a topology weaker than τ_∞, $H \cap E_n$ must be τ_n-closed.

b) Suppose $H \cap E_n$ is closed in E_n for all n and H is non-empty (if $H = \varnothing$, the statement is obvious). Let x_0 be any point of E such that $x_0 \notin H$. We must prove that there exists a τ_∞ neighborhood of x_0 whose

intersection with H is empty. Since $x_0 \in E = \bigcup_n E_n$ and $H \neq \varnothing$, there exists at least one p such that $x_0 \in E_p$ and $H \cap E_p \neq \varnothing$. We may assume without loss of generality, that $p=1$. Now it suffices to show that there exists an increasing sequence of circled, convex sets $U_1, U_2, \ldots$ such that for all n:

(i) U_n is neighborhood of 0 in E_n,

(ii) $x_0 + U_n$ does not intersect H.

In fact, if such a sequence (U_p) exists, then $x_0 + \bigcup_1^\infty U_n$ will be a x_0-neighborhood of 0 (by the definition of inductive limit) which does not intersect H by virtue of (ii).

Since $H \cap E_1$ is closed in E_1, and $x_0 \in E_1$ we can choose a ball H_1 of center 0 in E_1, such that $x_0 + U_1$ does not intersect $H \cap E_1$. Suppose now that we have already chosen n sets $U_1, \ldots, U_n$, satisfying the preceding conditions. Then $x_0 + U_n$ is compact in E_{n+1}, and *does not intersect H*. On the other hand, $H \cap E_{n+1}$ *is closed in* E_{n+1}. Therefore the distance δ_n between $x_0 + U_n$ and $H \cap E_{n+1}$, in the normed space E_{n+1} *must be* >0. We set $V_{n+1} = \left\{ x : \|x\|_{n+1} < \frac{\delta_n}{2} \right\}$ and U_{n+1} the circled convex hull of $U_n \cup V_{n+1}$. Then $U_n \subset U_{n+1} \subset U_n \cup V_{n+1}$, so that $x_0 + U_{n+1}$ cannot intersect H, the distance between $U_n \cup V_{n+1}$ and $H \cap E_{n+1}$ being $\geq \frac{\delta_n}{2}$ in E_{n+1}. On the other hand, since U_n and V_{n+1} are bounded in E_{n+1}, so is their circled convex hull, that is U_{n+1}. Finally, as V_{n+1} is a neighborhood of 0 in E_{n+1}, so is $U_{n+1} \supset V_{n+1}$. Thus all the preceding conditions are satisfied by the sets $U_1, \ldots, U_{n+1}$ and the theorem is proved by induction on n. ♦

5.4.4. COROLLARY. *Every* (LN*) *space is a Hausdorff space.*

In fact, the theorem implies that every set reducing to a point is τ_∞-closed.

5.4.5. COROLLARY. *Let F be any topological space. Then a mapping* $\varphi: E \to F$ *is* τ_∞ *continuous iff its restriction to each* E_n *is a* τ_n*-continuous mapping of* E_n *into F.*

PROOF. Suppose φ is τ_∞ continuous and let M be any closed subset of F. Then $\varphi^{-1}(M)$ is τ_∞ closed and hence $\varphi^{-1}(M) \cap E_n$ is τ_n closed. The converse is analogously proved. ♦

Remark. This corollary is equivalent to the theorem itself. What the theorem means is that among *all topologies* in *E,* (not necessarily locally convex), inducing on each E_n a topology weaker than τ_n, τ_∞ is the strongest one.

5.4.6. THEOREM. *A set H is bounded in E iff there exists an integer p, such that H is contained in* E_p *and bounded in this normed space.*

PROOF. a) Suppose H is contained and bounded in E_p, and let V be any τ_∞ neighborhood of 0 in E. Then $V \cap E_p$ contains a τ_p-neighborhood of 0; i.e. there exists an $\varepsilon > 0$ such that $\varepsilon B_p \subset V \cap E_p$. On the other hand, H being bounded in E_p implies that there exists a $\rho > 0$ such that $H \subset \rho(\varepsilon B_p)$. Hence $H \subset \rho V$; i.e. H is absorbed by any τ_∞-neighborhood V of 0 and therefore is bounded in E.

b) Suppose now H is bounded in E, and put

$$C_n = \{x : \|x\|_n < n\} = n\mathring{B}_n .$$

We are going to show that H is contained in one of the open balls C_n. Suppose this is not true. Then it will be possible to take in H a sequence of points $x_1, \ldots, x_n, \ldots$ such that $x_n \notin C_n$ for all n. Now by a technique similar to the one used for theorem 5.4.3. we are going to prove the existence of a sequence $U_1 \subset U_2 \subset \cdots$ of circled convex sets such that: (i) U_k is a bounded and closed neighborhood of 0 in E_k contained in $\mathring{B}_k$ for all k; (ii) $\frac{1}{n} x_n \notin U_k$ for all k and n. For example, take $U_1 = \frac{1}{2} B_1$; since $x_n \notin C_n = n\mathring{B}_n$ and $B_1 \subset B_n$ for all n, then $\frac{1}{n} x_n \notin U_1$ for all n. Suppose that we have already chosen k sets

$U_1, \ldots, U_k$ satisfying the preceding conditions and place:

$$M_k = \left\{x_1, \frac{1}{2}x_2, \ldots, \frac{1}{k}x_k\right\} \cup \left(E_{k+1} \setminus \mathring{B}_{k+1}\right).$$

Then M_k is a *closed set* in E_{k+1} which contains all points $\frac{1}{n}x_n$ and *does not intersect* U_k, according to (i) and (ii). On the other hand, U_k is compact in E_{k+1}. Hence, if we set $\delta_k = dist(U_k, M_k)$, $V_k = \left\{x : \|x\|_k < \frac{\delta_k}{2}\right\}$ and U_{k+1} = circled convex hull of $U_k \cup V_k$, we can prove, as in theorem 5.4.3., that U_{k+1} is a circled convex hull compact in E_{k+2} contained in $\mathring{B}_{k+1}$, and it is obvious that $\frac{1}{n}x_n \notin U_{k+1}$ for all n. Thus the existence of a sequence satisfying (i) and (ii) is now proved.

Now it is readily seen that the set $V = \bigcup_{1}^{\infty} U_n$ is a τ_∞-neighborhood of 0 such that $\frac{1}{n}x_n \notin V$. But then, for every $\rho > 0$, we should have $x_n \notin \rho U$ for all $n > \rho$ and this is impossible, the set H being bounded in E. Consequently, there exists at least one p such that $H \subset C_p$, which implies that H is contained and bounded in E_p. ♦

5.4.7. COROLLARY. *A sequence (x_n) of points of E converges to a point x of E if and only if there exists a p such that all the points x_n and x belong to E_p and $x_n \to x$ in E_p.*

PROOF. a) Suppose there exists p such that $x_n \to x$ in E_p. Then, since τ_∞ induces on E_p a topology weaker than τ_p, $x_n \to x$ in E.

b) Suppose $x_n \to x$ in E. Then the set $X = \{x_n\}_{n \in \mathbb{N}} \cup \{x\}$ is bounded in E and so there exists an r such that X is bounded in E_r. Hence the aderence $\overline{X}$ of X in E_{r+1} is compact and, according to a general theorem of Topology, τ_∞ *induces* on $\overline{X}$ the same topology as does τ_{r+1}. So, as $x_n \to x$ in E, we can conclude that $x_n \to x$ in E_{r+1}. ♦

5.4.8. COROLLARY. *If E is infinite dimensional, then E is not metrisable.*

PROOF. Suppose E is metrisable. Then there exists a fundamental system $\mathcal{V}$ of neighborhoods of 0 in E which is countable: $\mathcal{V} = \{V_1, \ldots, V_n, \ldots\}$. Now at least one of these neighborhoods must be bounded in E; otherwise there would exist, for each n, an $x_n \in V_n$ such that $x_n \notin C_n = n\overset{\circ}{B}$ (by the theorem) and thus (x_n) would be a sequence *converging to* 0 *and unbounded*, which is impossible. Let V_p be a neighborhood of the system $\mathcal{V}$ which is bounded in E, then V_p is bounded in E_m for some m and hence relatively compact in E_{m+1}. As V_p is also a neighborhood of 0 in E_{m+1}, it follows that E_{m+1} is finite dimensional and $E_{m+1} = E_{m+2} = \cdots = E$ since V_p is a neighborhood of 0 in E which implies $E = \bigcup_1^\infty kV_p$. ♦

5.4.9. COROLLARY. *Every* (LN*) *space E is an* (M)*-space (i.e. a Montel space where every bounded set is relatively compact).*

In fact if M is a bounded set in E, then M is bounded in some E_p and hence relatively compact in E_{p+1}. Since the Hausdorff topology τ_∞ induces on E_{p+1} a topology weaker than τ_{p+1}, it follows that M is also relatively compact in E.

5.4.10. COROLLARY. *Every* (LN*) *space E is reflexive (i.e. the strong bidual E'' of E is topologically isomorphic to E).*

In fact, E being the inductive limit of a family of normed spaces is a *barreled* space, and this along with 5.4.9., implies that E is reflexive.

5.4.11. COROLLARY. *Every* (LN*) *space is complete.*

PROOF. By the theorem there exists a sequence (C_n) of bounded sets in E such that every bounded set H in E is contained in one of the C_n. This implies that the polar sets $\overset{\circ}{C}_1, \overset{\circ}{C}_2, \ldots$ form a fundamental sequence of neighborhoods of 0 in E' which is *countable*. Hence

E' is metrisable and as E is isomorphic to E'', it follows that E is complete. ♦

Remark. The (LN^*) spaces turn out to be "Schwartz spaces", according to the terminology of Grothendieck. They can be characterized as the *strong duals of the Schwartz metrisable spaces*. Observe, however that for a direct definition of (LN^*) spaces, as well as for their application to define *directly* the topology in spaces of distributions, the preceding theorems are needed, the results of Grothendieck being insufficient. For further information, see the references at the end of the chapter.

5.5. Topology of $\mathscr{D}(I)$, when I is a compact interval

In order to arrive at our goal, we still need a criterium concerning a particular case of the concept of inductive limit introduced at the beginning of 5.4. Let E be a normed space, F a vector space (over the same field $\mathbb{R}$ or $\mathbb{C}$), and φ a linear mapping of E onto F. It is easily seen that the strongest topology on F making φ continuous (the so called **image top** of the topology of E by φ) can be defined by the semi-norm corresponding to the set $\varphi(B)$ where B is the unit ball in E. Then, *F becomes a seminormed space, which is a normed space iff the kernel $N(=\varphi^{-1}(0))$ of φ is closed in E (which is a necessary and sufficient condition for the set $\{0\}$ to be closed in F).*

This being so, we let I be a compact interval on $\mathbb{R}$. Then $C(I)$ is a normed space according to the usual definition of norm recalled in 5.1. On the other hand, we have seen that the operator D^n for $n=1, \ldots$ defines a linear mapping of $C(I)$ *onto* the space $C_n(I)$ of distributions of rank $\leq n$. In these circumstances it is natural to consider the space $C_n(I)$ provided with the image topology τ_n of the topology τ of $C(I)$ by means of D^n. Now, the kernel of D^n (i.e. the set of all functions φ such that $D^n\varphi=0$ in C_n) is the set $\mathscr{P}_n$, which as we have seen *is closed* in $C(I)$ (cf. remark to theorem 5.2.2.). Consequently:

5.5.1. *The vector space $C_n(I)$ with the topology τ_n is a normed space.*

Observe that now we have, both topologically and algebraically:

$$C_n(I) \cong C(I)/\mathcal{P}_n .$$

Besides, it is easily seen that:

5.5.2. *A sequence* (f_k) *of distributions on* I *converges to a distribution* g *on* I *in the* τ_n *topology iff exists a sequence of functions* F_k *in* $C(I)$ *and a function* $G \in C(I)$ *such that* $f_k = D^n F_k$ *for all* k, $g = D^n G$ *and* $F_k \to G$ *uniformly on* I.

Now, we are going to prove that:

5.5.3. *The normed spaces* $C_n(I)$, $n=1, 2, \ldots$ *form a regular sequence according to definition* 5.5.1.

PROOF. Remember that the unit ball B_n in $C_n(I)$, is the image of the unit ball B in $C(I)$; that is

$$B_n = D^n B = \{f : f = D^n F,\ F \in C(I),\ \|F\| \leq 1\}.$$

Now set

$$B' = \left\{\Phi : \Phi(x) \equiv \int_c^x F(\xi)\, d\xi,\ F \in B\right\} \quad (c \in I).$$

Then for all $\Phi \in B'$ and all x, $x+h \in I$,

$$|\Phi(x+h) - \Phi(x)| = \left| \int_x^{x+h} F(\xi)\, d\xi \right| \leq |h|.$$

This shows that B' is *equicontinuous* on the (compact) interval l; B' is also bounded, of course. Hence, according to Ascoli's theorem, B' is *relatively compact* in $C(I)$. But D^{n+1} defines a continuous mapping of $C(I)$ onto $C_{n+1}(I)$. Consequentely, the set $B_n = D^n B = D^{n+1} B'$ is *relatively compact in* $C_{n+1}(I)$, for all n, and this implies that the sequence $(C_n(I))$ is regular. ♦

Remember now that:

$$\mathscr{D}(I) = C_\infty(I) = \bigcup_{n=1}^{\infty} C_n(I);$$

thus it is natural to consider the vector space $\mathscr{D}(I)$ provided with the topology C_∞, which is the inductive limit of the topologies of the normed spaces $C_n(I)$.
Then, according to 5.5.3.:

5.5.4. *$\mathscr{D}(I)$ is a (LN^*) space.*

In particular from 5.4.7. and 5.5.2.:

5.5.5. *The concept of convergence for sequences in the locally convex spaces $\mathscr{D}(I)$ is the same as the introduced directly in* 5.2.1.

5.6. Topology of $\mathscr{D}(\Omega)$, where Ω is an open set

For this case, we need the concept of **projective limit**.

Consider a vector space E, a family $(F_\alpha)_{\alpha \in A}$ of vector spaces over the same field ($/R$ or $\mathbb{C}$) and let φ_α be, for each $\alpha \in A$, a linear mapping of E onto F_α. Suppose that, on each F_α there is defined a locally convex topology τ_α. Then it is easily seen that among all locally convex topologies on E for which each φ_α is continuous, there is one τ^* weaker than all the others: this is called the **projective limit** of the topologies τ_α in E, with respect to the mapping φ_α. To define τ^* directly it is sufficient to observe the following:

5.6.1. *A filter $\mathfrak{F}$ converges to* 0 *in $E(\tau^*)$ if and only if the filter $\varphi_\alpha(\mathfrak{F})$ converges to* 0 *in $F_\alpha(\tau_\alpha)$ for each $\alpha \in A$.*

Now let Ω be any open set in $/R$. For any compact interval $I \subset \Omega$, the restriction operator P_I is a linear mapping of $\mathscr{D}(\Omega)$ onto $\mathscr{D}(I)$. But $\mathscr{D}(I)$ has been defined as a locally convex space, as a (LN^*)-space.

Hence, *it is natural to consider the vector space* $\mathscr{D}(\Omega)$ *provided with the projective limit of the topologies of the spaces* $\mathscr{D}(I)$ *with respect to the operators* P_I.

Then according to 5.6.1.:

5.6.2. *The concept of convergence for sequences in locally convex spaces* $\mathscr{D}(\Omega)$ *is the same as introduced in* 5.2.5.

It must be observed however that the locally convex space $\mathscr{D}(\Omega)$ is not complete. For example, the sequence of distributions

$f_n = \sum_1^n \delta^{(k)}(\hat{x} - k)$ on $/R$ is a Cauchy sequence on each compact interval

I, hence on $/R$, but it does not converge to a distribution on $/R$.

Obviously, we can define on the space $\overline{\mathscr{D}}(\Omega)$ of all global distributions on Ω a topology as we did for $\mathscr{D}(\Omega)$. Then it is readily seen that $\mathscr{D}(\Omega)$ *is a locally convex subspace of* $\overline{\mathscr{D}}(\Omega)$*, which is dense in* $\overline{\mathscr{D}}(\Omega)$. Moreover, remembering that every space $\mathscr{D}(I)$ (I a compact interval) is complete being a (LN*)-space, is easily shown that:

5.6.3. $\overline{\mathscr{D}}(\Omega)$ *is complete.*

So $\overline{\mathscr{D}}(\Omega)$ can also be obtained by the completion of $\mathscr{D}(\Omega)$.

Finally, it can be proved that the preceding topologies on $\mathscr{D}(I)$ and $\mathscr{D}(\Omega)$ coincide with the strong topologies introduced by L. Schwartz in these spaces, considered as the duals of certain spaces of C^∞ functions.

Remark. The (LN*)-spaces turn out to be a special category of Schwartz spaces; they can be characterized as the strong duals of Schwartz metrisable spaces. This category of spaces has been pointed out by the author in 1952, first for the study of spaces of analytic functions (see references below). But its application to spaces of distributions required some developments introduced in 1954. The

(*LN**)-spaces occur in a great number of situations in functional analysis, as far as distributions and analytic functions are concerned. The specific properties of these spaces are not implied in the paper by Grothendieck, on (*F*) and (*DF*) spaces, where the Schwartz spaces were introduced. Some research on (*LN**)-spaces has been made by Yoshinaga and a generalization of this class of spaces has been presented by Kaikov.

REFERENCES

[1] A. GROTHENDIECK. *Sur les Espaces (F) et (DF).* Summa Brasiliensis Math. 1953.

[2] G. KÖTHE. *Topologishe Lineare Räume.* Springer 1960.

[3] D. A. KAIKOV. *Inductive and Projective Limits with Completely Continuous Mappings*. Dokl. Atad. Nank. SSSR (N.S.) 136 (1951) 984-986. (Math. Reviews 19-754).

[4] J. S. e SILVA. *Sui fondamenti della teoria dei funzionali analitici.* Portugaliae Math. 1953.

[5] J. S. e SILVA. *Su certe classi di spazi localmente convessi importanti per le applicazioni.* Rendiconti di Mat. e delle sue applicazioni, Roma (5). IA (1955), 355-410.

[6] R. YOSHINAGA. *On a Locally Convex Space introduced by J. S. e Silva.* Journ. Science Hiroshima Univ. (A), 21, n.° 2.

Página em branco

CHAPTER VI

LIMITS AND INTEGRALS OF DISTRIBUTIONS

6.1. Limits of a distribution as $x \to +\infty$

Let I be an open interval unbounded on the right; i.e. of the form $I =]a, +\infty[$ with $a \in I\!R \cup \{-\infty\}$. The following two definitions are well known in classical analysis.

6.1.1. DEFINITIONS. Let f and φ be two functions on I. The function f is said to be **of order less than φ** iff $\exists\, x_0 \in I\!R$ and a function f_0 such that:

$$f = \varphi f_0 \quad \text{for} \quad x > x_0 \quad \text{and} \quad f_0(x) \to 0 \quad \text{as} \quad x \to +\infty.$$

On the other hand, f is said to be **at most of the order of φ** as $x \to +\infty$ iff $\exists\, x_0 \in I\!R$ and a function f_0 bounded for $x > x_0$, such that $f = \varphi f_0$.

In the first case we shall write:

$$f \in o(\varphi) \text{ as } x \to +\infty \text{ (or, on the right)}$$

and in the second case:

$$f \in O(\varphi) \text{ as } x \to +\infty \text{ (or, on the right).}$$

These relations replace the classical $f=o(\varphi)$ and $f=O(\varphi)$ which are not logically correct and may produce confusion in functional analysis.

Observe that:

6.1.2. *If there exists x_0 such that $\varphi(x)\neq 0$ for $x>x_0$, then*

$$f\in o(\varphi) \text{ as } x\to+\infty \Leftrightarrow \frac{f(x)}{\varphi(x)}\to 0, \text{ as } x\to+\infty$$

$$f\in O(\varphi) \text{ as } x\to+\infty \Leftrightarrow \frac{f(x)}{\varphi(x)} \text{ is bounded on the right.}$$

In order to extend "o" to distributions, we first consider the case where $\varphi=\hat{x}^{\alpha}$, with $\alpha>-1$ (for simplicity the sign "^" will be omitted).

Let be $\Im$ the Lebesgue integral operator defined by $\int_c^x f(\xi)d\xi$ with c in I.

6.1.3. LEMMA. *If α is a real number >-1 and f a continuous function such that $f\in o(x^{\alpha})$ as $x\to+\infty$, then $\Im f\in o(x^{\alpha+1})$ as $x\to+\infty$.*

PROOF. Suppose $f\in o(x^{\alpha})$ as $x\to+\infty$. This means that there exists x_0 and f_0 such that $f=x^{\alpha}f_0$ for $x>x_0$ and $f_0\to 0$ as $x\to+\infty$. Let $\delta>0$ be given; then $\exists x_1\in I\!R$ such that $|f_0(x)|<\delta$ for all $x>x_1$. We can assume $x_1>x_0>0$. Now, for every $x>x_1$

$$\Im f(x)=K+\int_{x_1}^{x}\xi^{\alpha}f_0(\xi)d\xi \quad \text{where} \quad K=\int_c^{x_1}f.$$

Since $|f_0(x)|<\delta$ and $\xi>0$, for $\xi>x_1$, we have:

$$\left|\frac{\Im f(x)}{x^{\alpha+1}}\right|\leq\frac{|K|}{x^{\alpha+1}}+\frac{x^{\alpha+1}-x_1^{\;\alpha+1}}{(\alpha+1)x^{\alpha+1}}\,\delta, \quad \forall x>x_1,$$

and therefore, since $\alpha>-1$:

$$\lim_{x \to \infty} \left| \frac{\Im f(x)}{x^{\alpha+1}} \right| \leq \frac{\delta}{\alpha+1} .$$

As δ is arbitrary, this implies that

$$\frac{\Im f(x)}{x^{\alpha+1}} \to 0 \text{ as } x \to +\infty \text{; i.e. } \Im f \in o(x^{\alpha+1}). \ \blacklozenge$$

6.1.4. Remark. This lemma obviously extends to locally summable functions and even to measures, as we shall see.

The lemma suggests the following:

6.1.5. DEFINITION. Let α be a real number >-1 and f a distribution on I. We write $f \in o(x^{\alpha})$ as $x \to +\infty$ iff there exists an integer $p \geq 0$ and a continuous function F on I, such that:

$$f = D^p F \text{ and } \frac{F(x)}{x^{\alpha+p}} \to 0 \text{ as } x \to +\infty .$$

6.1.6. Remark. The lemma implies that if there exists $p \in I\!N_0$ and $F \in C(I)$ satisfying the preceding conditions, *then every integer $m \geq p$ and every function G such that $G = \Im^{m-p} F + P$ where $P \in \mathcal{P}_m$, satisfies the same conditions* (observe that if $P \in \mathcal{P}_m$, then $\dfrac{P(x)}{x^{\alpha+m}} \to 0$ as $x \to +\infty$).

6.1.7. LINEARITY PROPERTY. *If $f \in o(x^{\alpha})$ and $g \in o(x^{\alpha})$ as $x \to +\infty$, with $\alpha > -1$, then:*

$$\lambda f + \mu g \in o(x^{\alpha}) \text{ as } x \to +\infty, \ \forall \lambda, \ \mu \in \mathbb{C}.$$

For the proof, it is sufficient to represent f and g as derivatives of the same order of continuous functions, taking into account 6.1.6.

In particular, α *may be equal to* 0. Then $x^0=1$ and if $f\in o(1)$ as $x\to+\infty$, it is natural to say that $f\to 0$ as $x\to+\infty$. More generally, let λ be any complex number and $f\in\mathscr{D}(I)$; then:

6.1.8. DEFINITION. We say that f **converges to** λ as $x\to+\infty$ if and only if $f-\lambda\in o(1)$ as $x\to+\infty$. A distribution f is said to be **convergent** as $x\to+\infty$ if and only if $\exists\lambda\in\mathbb{C}$ such that $f\to\lambda$, as $x\to+\infty$.

Taking definition 6.1.5. into account and observing that $\lambda=D^p\left(\dfrac{\lambda x^p}{p!}\right)$ for every $p\in IN_0$, we can define the preceding concept as follows:

6.1.9. DEFINITION. We say that $f\to\lambda$ as $x\to+\infty$ if and only if there exists $p\in IN_0$ and $F\in C(I)$ such that:

$$f=D^pF \text{ and } \frac{F(x)}{x^p}\to\frac{\lambda}{p!} \text{ as } x\to+\infty \text{ (in the ordinary sense).}$$

Remark. Instead of "f tends to λ as $x\to+\infty$", we shall sometimes write "$f(x)\to\lambda$, as $x\to+\infty$", but it should be remembered that in these cases x is a dummy variable.

6.1.10. *If $f\to\lambda$ as $x\to+\infty$ and $f\to\mu$ as $x\to+\infty$ then $\lambda=\mu$.*

In fact, if $f-\lambda\to 0$ and $f-\mu\to 0$ as $x\to+\infty$, then, by 6.1.7. $(f-\lambda)-(f-\mu)=\mu-\lambda\to 0$ as $x\to+\infty$. But, for every integer $p\geq 0$ and every continuous function F such that $\mu-\lambda=D^pF$, we have necessarily $F=(\mu-\lambda)\dfrac{x^p}{p!}+P$ where $P\in\mathscr{P}_p$.

Hence by definition 6.1.9., $\mu-\lambda$ cannot tend to 0 *unless* $\lambda=\mu$.

This makes legitimate the definition complementary to 6.1.8.

6.1.11 DEFINITION. We say that λ is the **limit** of f as $x \to +\infty$, iff $f \to \lambda$ as $x \to +\infty$. In this case, we shall write $\lambda = \lim\limits_{x \to +\infty} f(x)$ or $\lambda = f(+\infty)$.

The uniqueness of the limit is guaranteed in 6.1.10, and from 6.1.7., follows:

6.1.12. LINEARITY PROPERTY. *If f and g are convergent as $x \to +\infty$, then:*

$$\lim_{x \to +\infty} (\alpha f + \beta g) = \alpha \lim_{x \to +\infty} f + \beta \lim_{x \to +\infty} g, \quad \forall \alpha, \beta \in \mathbb{C}.$$

In turn, from 6.1.3. and the preceding definitions, it follows:

6.1.13. *If f is a continuous function such that $\lim\limits_{x \to +\infty} f(x) = \lambda$ in the ordinary sense, then the same fact holds in the distributional sense; i.e., in the sense of definitions* 6.1.11. *and* 6.1.9..

Observe, that according to 6.1.5., this theorem extends to locally summable functions (and even to measures). However, it must be observed that the converse of this theorem is not true.

6.1.14. Example. As is well-known, the function $\cos x$ is not convergent in the ordinary sense as $x \to +\infty$. But we have:

$$\lim_{x \to +\infty} \cos x = 0, \text{ in the distributional sense.}$$

To see that, it is enough to apply definition 6.1.5. observing that $\cos x = D \sin x$ and $\dfrac{\sin x}{x} \to 0$, as $x \to +\infty$.

6.1.15. General remark. All preceding definitions may be extended and all propositions remain true, if we replace throughout $+\infty$ by $-\infty$ and "on the right" by "on the left". In particular, we must then consider an interval I, unbounded on the left, $I =]-\infty, a[$, instead of an interval unbounded on the right.

6.1.16. DEFINITION. We say that f **tends to** λ as $x \to \infty$ and we write $\lim_{x\to\infty} f(x)=\lambda$ if and only if $\lim_{x\to+\infty} f(x)=\lim_{x\to-\infty} f(x)=\lambda$.

For example, it is easily seen that (cf. 6.1.14): $\lim_{x\to\infty} \cos x = 0$ (in the distributional sense).

6.2. Limits and value of a distribution at a point of *IR*

Let now I be any open interval $]a, b[$, *bounded on the left*. Then, definitions 6.1.1. and 6.1.2. are readily extended to this case, replacing throughout "$x \to +\infty$" by "$x \to a^+$" and "on the right" by "on the left."

If we place $\Im_a f(x) = \int_a^x f(\xi)\,d\xi$, we prove, as for 6.1.3. (the proof is even simpler):

6.2.1. LEMMA. *If f is a continuous function on I, such that $f \in o[(x-a)^\beta]$ as $x \to a^+$ where $\beta > -1$,* then *$\Im_a^n f \in o[(x-a)^{\beta+n}]$ as $x \to a^+$,* for $n=0, 1, \dots$.

This lemma justifies the following

6.2.2. DEFINITION. If $f \in \mathscr{D}'(I)$ and $\beta > -1$, we write $f \in o[(x-a)^\beta]$ as $x \to a^+$ iff there exists $p \in IN_0$ and $F \in C(I)$ such that $f = D^p F$ and

$$\frac{F(x)}{(x-a)^{\beta+p}} \to 0 \text{ as } x \to a^+.$$

6.2.3. Remark. The lemma implies that if there exist p and F satisfying these conditions, *then every integer $m \geq p$, along with the function $\Im_a^{m-p} F$, satisfies the same conditions.* (But it must be observed that for each integer $m \geq p$, there is no function different from $\Im_a^{m-p} F$ satisfying the same conditions).

Now we are able to extend definitions 6.1.8. and 6.1.11., as well as propositions 6.1.7., 6.1.10., 6.1.12. and 6.1.13., replacing $+\infty$ by a^+. In particular, the convergence as $x \to a^+$ can be defined directly as follows:

6.2.4. DEFINITION. A distribution f on $I =]a, b[$ **tends to** λ as $x \to a^+$ iff there exists $p \in /N_0$ and $F \in C(I)$, such that:

$$f = D^p F \text{ and } \frac{F(x)}{(x-a)^p} \to \frac{\lambda}{p!} \text{ as } x \to a^+ \text{ (in the ordinary sense).}$$

Besides, the concepts of convergence corresponding to the cases $x \to +\infty$ and $x \to a^+$ are related to each other according to the following rule:

6.2.5. *Suppose* $I =]a, +\infty[$, $\beta > 0$ *and* $f \in \mathscr{D}(I)$. *Then, if*

$$g(t) = f\left(a + \beta \frac{1}{t}\right), \text{ we have: } \lim_{t \to +\infty} g(t) = \lambda \Leftrightarrow \lim_{x \to a^+} f(x) = \lambda.$$

PROOF. This obviously reduces to the case $a = 0$ and $\lambda = 0$ with $\beta = 1$. Suppose $f(x) \to 0$ as $x \to 0^+$. Then, there exists $p \in /N_0$ and $F \in C(I)$ such that $f = D^p F$ and $\frac{F(x)}{x^p} \to 0$ as $x \to 0^+$. Moreover (cf. 4.5), we have $g(t) = (-t^2 D_t)^p F\left(\frac{1}{t}\right)$ and it is easily shown by induction on p that there exists $p+1$ numbers a_k (whose expression are not needed here) such that

6.2.6. $$g(t) = \sum_{k=0}^{p} a_k D_t^k \left[t^{p+k} F\left(\frac{1}{t}\right) \right].$$

Now, since $\frac{F(x)}{x^p} \to 0$ as $x \to 0^+$, $t^p F\left(\frac{1}{t}\right) \to 0$ as $t \to +\infty$. Hence

$$\lim_{t\to+\infty} \frac{t^{p+k}F\left(\frac{1}{t}\right)}{t^k} = 0, \quad \text{for } k=0,\ldots,p,$$

which according to definition 6.1.9. means that all terms on the right side of 6.2.6. $\to 0$ as $t\to+\infty$. In a similar way, we prove that, if $g(t)\to 0$ as $t\to+\infty$, then $f(x)\to 0$ as $x\to 0^+$. ♦

We can obviously *define the concept*:

$$f(x)\to\lambda \text{ as } x\to b^-$$

as we did for the case $x\to a^+$ considering now an interval $]a, b[$, *bounded on the right*. It is readily seen that all preceding propositions and remarks can be extended to this case.

Let I, be now any open interval in $/R$, $I=]a, b[$, and let c be any point of I, that is $a<c<b$. Then if $f\in\mathscr{D}(I)$, we define the concepts:

$$\text{“}f(x)\to\lambda \text{ as } x\to c^+\text{”}$$

$$\text{“}f(x)\to\lambda \text{ as } x\to c^-\text{”}$$

by considering, instead of f, *its restrictions* to the intervals $]a, c[$ and $]c, b[$. As in classical analysis, we shall put

$$f(a^+) = \lim_{x\to a^+} f(x) \text{ (right-hand limit of } f \text{ at } a)$$

$$f(a^-) = \lim_{x\to a^-} f(x) \text{ (left-hand limit of } f \text{ at } a)$$

whenever the limit in question exists.

6.2.7. DEFINITION. We say that f **tends to** λ as $x\to c$ iff $f(x)\to\lambda$ as $x\to c^+$ and $f(x)\to\lambda$ as $x\to c^-$. In this case, we write $\lambda=\lim\limits_{x\to c} f(x)$.

According to preceding definitions and remarks, we can also define directly this concept:

6.2.8. DEFINITION. The distribution f **tends to** λ as $x \to c$ iff there exists an integer $p \geq 0$ and a function F continuous at every point x of I distinct from c, such that:

$$f = D^p F \text{ and } \lim_{x \to c} \frac{F(x)}{(x-c)^p} = \frac{\lambda}{p!} \text{ in the ordinary sense.}$$

6.2.9. Remark. Suppose, more generally, that I is any non-degenerate interval in $I\!R$ and that c is in the closure of I. Then, definition 6.2.8. applies, *even* if c is a extremity of the domain I of f; for example, if c is a left extremity of I, we have by definition:

$$\lim_{x \to c} f(x) = \lim_{x \to c^+} f(x).$$

With respect to the general hypothesis considered above, we have:

6.2.10. DEFINITION. A distribution f on I is said to be **continuous at a point** c iff there exists $p \in I\!N_0$ and $F \in C(I)$ such that $f = D^p F$ and $\dfrac{F(x)}{(x-c)^p}$ is convergent in the ordinary sense as $x \to c$. Then, we write:

$$f(c) = \lim_{x \to c} f(x) = p! \lim_{x \to c} \frac{F(x)}{(x-c)^p}$$

and the number $f(c)$ is said to be the **value of the distribution** f at the point c (or, for $x = c$).

From the linearity property of limits follows:

6.2.11. *If f and g are continuous at c, so is $\alpha f + \beta g$ for $\alpha, \beta \in \mathbb{C}$ and $(\alpha f + \beta g)(c) = \alpha f(c) + \beta g(c)$.*

Examples. 1 – Consider $f(x) = cos\dfrac{1}{x}$. Then f is a locally summable function on $I\!R$ and since:

$$cos\frac{1}{x}=2x\,sin\frac{1}{x}-D\left(x^2\,sin\frac{1}{x}\right),$$

$$\lim_{x\to 0}\left(x\,sin\frac{1}{x}\right)=0,$$

it is easily seen that f is continuous at the point 0 with the value 0 (in distributional sense, and not in ordinary sense!).

2 – It can be seen that $\lim\limits_{x\to 0}\delta^{(k)}=0$, and yet $\delta^{(k)}$ is not continuous at 0 for any $k=0, 1,\ldots$.

3 – It can be proved, as an exercise, that: *If f is a distribution on an interval I minus a point c of $\bar{I}$, and if f is convergent as $x\to c$, then there exists one and only one distribution $\widetilde{f}$ on $I\cup\{c\}$, which is continuous at c and such that $\widetilde{f}=f$ on I.*

Remark. The previous concepts of limits and value of a distribution at a point of $/R$ have been introduced by Lojasiewicz. As for the concepts of limit as $x\to+\infty$ or as $x\to-\infty$, the definitions given by Mikusinski and Sikorski seem to be to restrictive as they are not invariant for very simple substitutions such as $x=1/t$ and do not allow the justification of certain integral formulas occuring in applications. The definitions that we are using here do not present these inconveniences.

6.3. Primitives and integrals of distributions

If f is a distribution with domain in $/R$, we call **primitive** of f any distribution φ such that $D\varphi=f$. From this definition follows:

6.3.1. THEOREM. *Every distribution f has infinitely many primitives, and, if the domain of f is an interval then any two primitives of f differ by a constant.*

PROOF. In the general case, the domain of f will be the union of a system of mutually disjoint intervals (cf. 2.5); so we can reduce this to the case of a single interval. Let f be a distribution on I. Then f is of the form $f = D^n F$, with $F \in C(I)$, and every distribution φ of the form $\varphi = D^n \mathfrak{I} F + \mathrm{K}$, where $\mathfrak{I}$ is an integration operator and $K \in \mathbb{C}$, is obviously a primitive of f. Suppose now that $D\varphi_1 = D\varphi_2 = f$; then if $\varphi_1 = D^n \Phi_1$ and $\varphi_2 = D^n \Phi_2$, with Φ_1 and Φ_2 in $C(I)$, we have $D^{n+1}\Phi_1 = D^{n+1}\Phi_2$, which implies, by axiom 4 (cf. 2.2), that $\Phi_1 - \Phi_2$ is a polynomial P of degree $< n+1$. Thus $\varphi_1 - \varphi_2 = D^n P =$ constant. ♦

From 6.2.2. and 6.2.10. follows immediately:

6.3.2. COROLLARY. *If there exists a primitive of f which is continuous at a point a, then every primitive of f is continuous at a. If, in addition, the domain of f is an interval I, then for every complex number K, there exists one and only one primitive φ of f such that $\varphi(a) = K$.*

It will be natural to denote by the symbol

$$\int_a^{\hat{x}} f(\xi)\, d\xi \quad \text{or shortly by} \quad \int_a^{\hat{x}} f$$

the primitive of f assuming the value 0 at a. (Remember that the sign ^ indicating that x is a dummy variable may be omitted whenever no confusion is possible). Thus according to 6.3.2., *if there exists at least one primitive of f which is continuous at a,* the differential equation $D\varphi = f$ will have a single solution satisfying the initial condition $\varphi(a) = K$, and such a solution is:

$$\varphi(x) = K + \int_a^x f(\xi)\, d\xi.$$

As we have observed, it is understood that here x is only a dummy variable; the distribution φ need not actually have a value $\varphi(x)$ at *every* point x of I. But, obviously, if φ has a value at *some* point b of I, this value is naturally denoted by:

$$\varphi(b)=K+\int_a^b f(\xi)\,d\xi.$$

Thus the integral $\int_a^b f(\xi)\,d\xi$ $\left(\text{in short } \int_a^b f\right)$ is *defined* by the generalized Barrow Formula:

$$\int_a^b f(x)\,dx=\varphi(b)-\varphi(a).$$

Corollary 6.3.2. can be extended as follows:

6.3.3. COROLLARY. *If there exists a primitive of f having a limit as $x\to a^+$ [resp. as $x\to a^-$], then every primitive of f has a limit as $x\to a^+$ [resp. as $x\to a^-$]. If, in addition, the domain of f is an interval I, then for every complex number K, there exists one and only one primitive φ of f such that $\varphi(a^+)=K$ [resp. $\varphi(a^-)=K$].*

Remember that the existence of both $\varphi(a^+)$ and $\varphi(a^-)$ does not imply the existence of $\varphi(a)$.

All preceding remarks and conventions may now be extended to the newly considered cases. For example, we shall denote by

$$\int_{a^-}^x f(\xi)\,d\xi \left(\text{in short } \int_{a^-}^x f\right)$$

the primitive of f on I which tends to zero as $x\to a^-$; accordingly, if such a limit exists, the differential equation $D\varphi=f$ along with the initial condition $\varphi(a^-)=K$ will have the only solution

$$\varphi(x)=K+\int_{a^-}^x f(\xi)\,d\xi.$$

So, we have by definition

$$\int_{a^-}^{b^-} f(x)\,dx=\varphi(b^-)-\varphi(a^-)\,,\quad \int_{a^-}^{b^+} f(x)\,dx=\varphi(b^+)-\varphi(a^-).$$

If $a<b$, these are, respectively, *the integral of the distribution* f *on the intervals* $[a, b[$ *and* $[a, b]$. The integrals of f on $]a, b]$ and $]a, b[$ are analogously defined. Naturally such an integral is said to *exist* or to be *convergent* iff the two corresponding limits exist. If $b \le a$, we have of course:

$$\int_{a^-}^{b^+} f = -\int_{b^+}^{a^-} f\,, \quad \int_{a^+}^{b^+} f = -\int_{b^+}^{a^+} f\,, \text{ etc.}$$

Finally, all preceding definitions may be extended to *infinite intervals.* For example, we have by definition

$$\int_{a^-}^{+\infty} f(x)\,dx = \varphi(+\infty) - \varphi(a^-),$$

if φ is a primitive of f such that the limits on the right-hand side exist; and $\int_{a^-}^{+\infty} f(x)\,dx$ is called the integral of f on the interval $[a, +\infty[$. For other kinds of infinite intervals the definitions are quite analogous.

In the general case, a distribution f is said to be **integrable over an interval** I, iff the integral of f on I exists. This integral may be denoted by $\int_I f(x)\,dx$ or simply by $\int_I f$.

From the linearity property of limits follows immediately the corresponding property for integrals:

6.3.4. LINEARITY PROPERTY. *If two distributions f and g are integrable over I, so is $\alpha f + \beta g$ for any $\alpha, \beta \in \mathbb{C}$ and*

$$\int_I (\alpha f + \beta g) = \alpha \int_I f + \beta \int_I g\,.$$

On the other hand it should be observed that:

6.3.5. *If f is a function summable on I, then the integral of f over I, in the distributional sense, exists and equals the Lebesgue integral over I. More generally, if f is a locally summable function on I such that $\int_I f$ is convergent in the classical sense (even simply convergent), then $\int_I f$ exists, in the distributional sense, and has the same value.*

However, the converse of this proposition is not true, as we shall presently see:

Examples. 1 – Consider the integral $\int_I f(x)\delta^{(n)}(x-a)$ where I is any interval in $/R$, n an integer ≥ 0 and $f \in C^n$ a function on I. Then:

$$f\delta^{(n)}(\hat{x}-a)=\sum_{k=0}^{n}(-1)^k\binom{n}{k}D^{n-k}[f^{(k)}(a)\delta(\hat{x}-a)].$$

Now, for every $k<n$, a primitive of $D^{n-k}[f^{(k)}(a)\delta(\hat{x}-a)]$ is the distribution $f^{(k)}(a)\delta^{(n-k-1)}(\hat{x}-a)$ which tends to zero as x tends to any point x_0 in $/R$. Hence:

$$\int_I f(x)\delta^{(n)}(x-a)\,dx=(-1)^n f^{(n)}(a)\int_I \delta(x-a)\,dx=\begin{cases}(-1)^n f^{(n)}(a)\ , & \text{if } a\in I\\ 0 & , \text{if } a\notin I.\end{cases}$$

For example:

$$\int_{a^-}^{a^+} f(x)\delta''(x-a)\,dx = f''(a)\int_{a^-}^{a^+}\delta(x-a)\,dx = f''(a).$$

2 – Consider the integral $\int_{/R} e^{i\omega t}dt$, where ω is a real parameter. This integral is obviously divergent, in the classical sense, for every value of ω. However, for $\omega\neq 0$, one primitive of $e^{i\omega t}$ is $\dfrac{e^{i\omega t}}{i\omega}$ and

$$\frac{e^{i\omega t}}{i\omega} = \frac{1}{(i\omega)^2} De^{i\omega t}\ , \quad \lim_{t\to\infty} \frac{e^{i\omega t}}{t} = 0.$$

Hence, we have, in the distributional sense, for every $\omega \neq 0$:

$$\int_{-\infty}^{+\infty} e^{i\omega t} dt = \frac{1}{i\omega}\left(\lim_{t\to+\infty} e^{i\omega t} - \lim_{t\to-\infty} e^{i\omega t}\right) = 0.$$

For $\omega = 0$, this integral is divergent, even in the distributional sense. This result agree with the intuition of physicists, which have, long since, adopted the formula:

$$\int_{/R} e^{i\omega t} dt = 2\pi\delta(\omega).$$

However, a complete justification of this formula cannot be achieved, without a suitable definition of parametric integral, which will be given in chapter VIII.

The case considered in example 1 is included in the following proposition:

6.3.6. *Every distribution with a bounded carrier on /R is integrable on /R.*

PROOF. Let f be a distribution of bounded carrier on $/R$. This means that there exists a bounded interval $I = [a, b]$ such that $f = 0$ outside I. Hence, if φ is a primitive of f, $D\varphi = 0$ outside I and φ reduces to constants c_1 and c_2, respectively, on $]-\infty, a[$ and on $]b, +\infty[$. Thus $\varphi(-\infty) = \varphi(a^-) = c_1$ and $\varphi(b^+) = \varphi(+\infty) = c_2$. Hence, f is integrable on $/R$ and

$$\int_{/R} f = \int_I f = \int_{a^-}^{b^+} f = c_2 - c_1 . \ \blacklozenge$$

A complementary proposition to 6.3.6., which can be proved in a similar way is the following:

6.3.7. *Whenever f is integrable on $/R$, we have $\int_{/R} f = \int_I f$, for every interval containing the carrier of f.*

For example, if f is integrable on $/R$ and zero for $x<a$, then $\int_{/R} f = \int_{a^-}^{+\infty} f$.

In order to obtain more powerful tests for the convergence of integrals, we are going to develop the concept of order of growth for distributions.

6.4. Orders of growth for distributions

For brevity, we shall confine ourselves to the typical case where $x \to +\infty$, since the considerations in the other cases are analogous. Let I be any interval *unbounded on the right* and $\Im f(x) = \int_c^x f$, with $c \in I$, for $f \in C(I)$. The extension of the symbol "O" to distributions is based on the following lemma, whose proof is similar to the one of 6.1.3. and even more simple:

6.4.1. LEMMA. *If f is a continuous function on I such that $f \in O(x^\alpha)$ as $x \to +\infty$, with $\alpha > -1$, then $\Im f \in O(x^{\alpha+1})$ as $x \to +\infty$.*

6.4.2. DEFINITION. If $f \in \mathscr{D}(I)$ and $\alpha > -1$, then we write $f \in O(x^\alpha)$ as $x \to +\infty$ iff there exist $n \in /N_0$ and $F \in C(I)$, such that $f = D^p F$ and $\dfrac{F(x)}{x^{n+\alpha}}$ is bounded on the right.

The lemma guarantees the linearity property for this case. In particular:

6.4.3. DEFINITION. A distribution f on I is said to be **bounded on the right** iff $f \in O(1)$ as $x \to +\infty$, that is iff there exist $n \in \mathbb{N}_0$ and $F \in C(I)$ such that $f = D^n F$ and $\dfrac{F(x)}{x^n}$ is bounded on the right.

That being so, we are able to define the meaning of the expression "$f \in o(\varphi)$" and "$f \in O(\varphi)$" in the more general case when $f \in \mathscr{D}(I)$ and $\varphi \in C^\infty(I)$. For all that purpose, we can take as a model the classical definition 6.1.1.:

6.4.4. DEFINITION. We shall write $f \in o(\varphi)$ as $x \to +\infty$ iff there exists a real x_0 and a distribution f_0 such that:

$$f = \varphi f_0 \text{ for } x > x_0 \text{ and } f_0 \to 0 \text{ as } x \to +\infty.$$

We shall write $f \in O(\varphi)$ as $x \to +\infty$ iff there exists a real x_0 and a distribution f_0 such that $f = \varphi f_0$ for $x > x_0$ and f_0 is bounded on the right.

The first thing to do is to see whether these definitions are equivalent to the preceding ones in the particular case, when φ is of the form x^α, with $\alpha > -1$. This equivalence is easily proved by means of the formulas:

$$x^\alpha D^n F_0 = \sum_{k=0}^{n} (-1)^k \binom{n}{k} D^{n-k}(F_0 D_x^k x^\alpha)$$

$$D^n(x^\alpha G_0) = \sum_{k=0}^{n} \binom{n}{k} (D_x^k x^\alpha) D^{n-k} G_0$$

taking into account the linear property.

On the other hand, this same property can be now immediately extended to the general case. Moreover definition 6.4.4. introduce a remarkable new property which is a counterpart of the preceding lemmas.

6.4.5. DIFFERENTIATION PROPERTY. *If $f \in O(x^{\alpha})$ on the right, then $Df \in O(x^{\alpha-1})$ on the right, for every $\alpha \in /R$.*

We shall begin the proof in the case $\alpha=0$:

6.4.6. *If f is bounded on the right, then $Df \in O(x^{-1})$ as $x \to +\infty$.*

Suppose f bounded on the right. Then, there exists $p \in /N_0$, $F \in C(I)$ and c such that $f = D^p F$ for $x > c$ and $\dfrac{F(x)}{x^p}$ is bounded on the right. We may choose $c > 0$; then we have:

$$Df = x^{-1}(xD^{p+1}F) = x^{-1}[D^{p+1}(xF) - (p+1)D^p F] \text{ for } x > c$$

and it is readily seen that $D^{p+1}(xF)$ is bounded on the right, as well as $D^p F$. Hence $Df \in O(x^{-1})$ as $x \to +\infty$.

Suppose now $f \in O(x^{\alpha})$ $x \to +\infty$, where $\alpha \in /R$. Then there exist x_0 and f_0 such that $f = x^{\alpha} f_0$ for $x > x_0$ and $f_0 \in O(1)$ on the right. It follows that $Df = \alpha x^{\alpha-1} f_0 + x^{\alpha} Df_0$ and it is readily seen, applying 6.4.6., that $Df \in O(x^{\alpha-1})$ as $x \to +\infty$. ♦

By an identical argument, it is shown that the *differantiation property extends to the "o" symbol.*

Furthermore it is a simple matter to prove the following properties where the expression "on the right" or "as $x \to +\infty$" is omitted for simplicity.

6.4.7. *If f is convergent, then f is bounded.*

6.4.8. *If $f \in o(\varphi)$ then $f \in O(\varphi)$.*

6.4.9. *If $f \in O(x^{\alpha})$ and $\alpha < \beta$, then $f \in o(x^{\beta})$.*

Obviously we have chosen the case when $x \to +\infty$ as a model; the concepts and properties are quite analogous in cases such as $x \to -\infty$, $x \to c^+$, etc..

6.4.10. Convention. If a distribution f has the same growth property as $x \to +\infty$ and as $x \to -\infty$, we shall say that f has this property as $x \to \infty$. If $f \in \mathscr{D}'(I)$ is bounded on the right and on the left (respectively as x tends to the right extremity and to the left extremity of I), we shall say that f is **bounded on I** or simply **bounded**.

Remark. The concept of bounded distribution that we have just introduced is more general than the concept of bounded distribution according to Schwartz and necessary for the integral theory as we shall next see.

6.5. Convergence tests for integrals

Let us consider, at first, the case of integrals on $I\!R$. We have the following test, which is not true in classical analysis:

6.5.1. (A NECESSARY CONDITION FOR CONVERGENCE). *If a distribution f is integrable on $I\!R$, then $f \in O(x^{-1})$ as $x \to \infty$.*

PROOF. Suppose there exists a primitive φ of f such that φ is convergent as $x \to +\infty$ and as $x \to -\infty$[6]. Then by 6.4.7., φ is bounded on $I\!R$ and, by 6.4.6. (and its analog for the case $x \to -\infty$) we have $D\varphi \in O(x^{-1})$ *as* $x \to \infty$. ♦

The following theorem extends to distributions a well known classical test.

6.5.2. (A SUFFICIENT CONDITION FOR CONVERGENCE). *If there exists a number $\alpha < -1$ such that $f \in O(x^{\alpha})$ as $x \to \infty$, then f is integrable on $I\!R$.*

(6) – This does not mean that φ is convergent as $x \to \infty$, for the limits are in general different.

PROOF. Suppose $f \in O(x^\alpha)$ as $x \to \infty$ with $\alpha < -1$. Then there exists a number $c > 0$, an integer $n \geq 0$ and a continuous function F such that:

$$f = x^\alpha D^n F \text{ for } |x| > c, \text{ with } \frac{F(x)}{x^n} \text{ bounded for } |x| > c.$$

Set: $$F_1(x) = \begin{cases} F(x), & \text{for } |x| > c \\ 0, & \text{for } |x| < c \end{cases}, \quad f_1 = x^\alpha D^n F_1, \quad f_2 = f - f_1.$$

Then f_2 is a distribution with carrier contained in $[-c, +c]$; hence integrable on $/R$ (cf. 6.3.6.). So we have only to prove that f_1 is integrable on $/R$, for then we have:

$$\int_{/R} f = \int_{/R} f_1 + \int_{/R} f_2.$$

We shall put $f_1 = f$ and $F_1 = F$. Then:

$$f = x^\alpha D^n F = \sum_{k=0}^{n} (-1)^k c_k D^{n-k} (x^{\alpha-k} F)$$

where $c_k = \alpha(\alpha-1) \cdots (\alpha-k+1) \binom{n}{k}$. From here we deduce the following primitive of f:

6.5.3. $$\varphi = \sum_{k=0}^{n-1} (-1)^k c_k D^{n-k-1} (x^{\alpha-k} F) + (-1)^n c_n \int_0^x \xi^{\alpha-n} F(\xi)\, d\xi.$$

But since $F \in O(x^n)$ as $x \to \infty$ in the ordinary sense, we have $\xi^{\alpha-n} F \in O(\xi^\alpha)$ as $x \to \infty$ with $\alpha < -1$ and, according to the classical test, this implies that the primitive $\xi^{\alpha-n} F$ is summable on $/R$. Hence the last term in 6.5.3. is convergent as $x \to +\infty$ and as $x \to -\infty$.

As to the other terms, observe that the functions

$$\frac{x^{\alpha-k}F(x)}{x^{n-k-1}}=x^{\alpha+1}\frac{F(x)}{x^n} \text{ for } k=0,\dots,n-1,$$

tend to zero as $x\to\infty$ since $\alpha+1<0$ and $\dfrac{F(x)}{x^n}$ is bounded (in the ordinary sense). Hence, by definition 6.1.9.

$$D^{n-k-1}(x^{\alpha-k}F)\to 0 \text{ as } x\to\infty,$$

so that $\varphi(+\infty)=(-1)^n c_n\int_0^{+\infty}x^{\alpha-n}F$, $\qquad \varphi(-\infty)=(-1)^n c_n\int_0^{-\infty}x^{\alpha-n}F$.

Therefore f is integrable on $/R$ and

$$\int_{/R}f=\int_{/R}x^\alpha D^n F=(-1)^n c_n\int_{/R}x^{\alpha-n}F=(-1)^n\int_{/R}FD^n x^\alpha. \ \blacklozenge$$

We can deduce similar tests for integrals on intervals distinct from $/R$. For example, consider an interval $I=]a,+\infty[$ and $f\in\mathscr{D}(I)$. Then it is easily seen that *if f is integrable in I, then $f\in O(x^{-1})$ as $x\to+\infty$ and $f\in O((x-a)^{-1})$ as $x\to a^+$. If there exists $\alpha<-1$ and $\beta>-1$ such that $f\in O(x^\alpha)$ as $x\to+\infty$ and $f\in O((x-a)^\beta)$ as $x\to a^+$, then f is integrable on I.*

6.6. Multiplication and change of variables in connection with limits and integrals

It is a simple matter to prove the following propositions:

6.6.1. *If $f\in\mathscr{D}(I)$ is convergent as $x\to c^+$ with $c\in I$ and if $g\in C^\infty(I)$ then fg is convergent as $x\to c^+$ and:*

$$\lim_{x\to c^+}(fg)=\left(\lim_{x\to c^+}f\right)\left(\lim_{x\to c^+}g\right).$$

6.6.2. *If $f \in \mathscr{D}(I)$ is convergent as $x \to c^+$ with $c \in I$ and if h is a C^∞ mapping of an interval I^* into I, such that $h'(t) > 0$ in I^*, then $f(h(t))$ is convergent as $t \to \gamma^+$ with $h(\gamma) = c$ and:*

$$\lim_{t \to \gamma^+} f(h(t)) = \lim_{t \to c^+} f(x).$$

Obviously, these two propositions can be extended to the case when f is convergent as $x \to c^-$. Then the second one enable the usual substitution property to be extended to the integrals of distributions on bounded intervals. For example, assuming $f \in \mathscr{D}(I)$, $a, b \in I$ and h is an increasing C^∞ mapping of I^* into I such that $a = h(\alpha)$, $b = h(\beta)$, we have

$$\int_{a^+}^{b^-} f(x)\,dx = \int_{\alpha^+}^{\beta^-} f(h(t))h'(t)\,dt$$

whenever the first integral exist.

However these criterions are not sufficient in certain cases which occur in practice. Our next purpose is to introduce a stronger criterium than 6.6.2.. For simplicity, we shall reduce our discussion to the case where $x \to +\infty$ and $h(+\infty) = +\infty$, which can be taken as a model for other cases.

6.6.3. THEOREM. *Let $f \in \mathscr{D}(I)$, I unbounded on the right, and let h be a C^∞ mapping of an interval I^* into I such that $h'(t) \neq 0$ on I^* and $h(t) \to +\infty$ as $t \to +\infty$. Suppose that:*

(i) f is convergent as $x \to +\infty$
(ii) *h' tends to a number $c \neq 0$ as $t \to +\infty$ (in the ordinary sense)*
(iii) *$h^{(k)} \in o(t^{-k+1})$ as $t \to +\infty$ (in the ordinary sense), for $k > 1$.*

Then we have: $\lim\limits_{t \to +\infty} f(h(t)) = \lim\limits_{x \to +\infty} f(x)$

PROOF. Suppose $f \to \lambda$ as $x \to +\infty$. Then there exist $n \in /N_0$ and $F \in C(I)$ such that $f = D^n F$ and $\dfrac{F(x)}{x^n} \to \dfrac{\lambda}{n!}$ as $x \to +\infty$. Now

$f \circ h = \left(\frac{1}{h'} D_t\right)^n (F \circ h)$ and according to the hypothesis:

$$\lim_{t \to +\infty} \frac{h(t)}{t} = \lim_{t \to +\infty} h'(t) = c.$$

Hence:

6.6.3'.
$$\frac{F(h(t))}{t^n} = \frac{F(h(t))}{(h(t))^n} \left(\frac{h(t)}{t}\right)^n \to \frac{\lambda c^n}{n!}.$$

On the other hand it is easily seen that:

$$\left(\frac{1}{h'} D_t\right)^n (F \circ h) = \sum_{k=0}^{n} D_t^{n-k} \left(\alpha_k(t) F(h(t))\right)$$

where $\alpha_0 = \left(\frac{1}{h'}\right)^n$ and $\alpha_k \in o(t^{-k})$ *as* $t \to +\infty$, for $k=1, 2, \ldots, n$. Thus all terms in the last sum tend to zero as $x \to +\infty$, except $D_t^n(\alpha_0(F \circ h))$, which, by 6.6.3', tends to λ. ♦

This criterium and the corresponding ones for the cases when $x \to -\infty$, $t \to -\infty$, etc., lead to the following substitution rule for integrals.

6.6.4. COROLLARY. *Let f be a distribution integrable on IR and h a C^∞ mapping of IR onto IR such that:*

(j) $h'(t)$ is $\neq 0$ on IR and tends to numbers $\neq 0$ as $t \to +\infty$ and as $t \to -\infty$ (in the ordinary sense)

(jj) $h^{(k)} \in o(t^{-k+1})$ as $t \to \infty$ (in the ordinary sense) for all $k=2, 3, \ldots$.

Then $f(h(t))$ is integrable on IR and:

$$\int_{IR} f(x)\,dx = \int_{IR} f(h(t))\,|h'(t)|\,dt.$$

This rule is an immediate consequence of theorem 6.6.3. and its

corresponding theorems applied to a primitive φ of f. Observe that, in the case $h'(t)<0$

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{+\infty}^{-\infty} f(h(t))h'(t)\,dt = -\int_{-\infty}^{+\infty} f(h(t))h'(t)\,dt\,.$$

In particular 6.6.4. applies in the elementary cases when $x=t+a$ or $x=ct$, with $a\in IR$ and $c\in\mathbb{C}$. Then we have:

6.6.5. $$\int_{IR} f(x)\,dx = |c|\int_{IR} f(cx)\,dx\,.$$

6.6.6. $$\int_{IR} f(x+a)\,dx = \int_{IR} f(x)\,dx\,.$$

The last formula can be expressed by saying that the integral is invariant under translations.

More refined criterions can be obtained by using the concept of measure as we did for multiplication in chapter IV.

Remember that if μ is a measure on an open interval I, the **total variation** of μ in a bounded interval J such that $\bar{J}\subset I$ is defined to be the supremum of the sums $S_p = \sum_1^p |\mu(J_i)|$, for all finite partitions P of J into intervals J_1, ..., J_p. We shall denote by $|\mu|(J)$ the total variation of μ in J; as is well known, $|\mu|$ is again a measure on IR (the **modulus** of μ) such that:

(i) if $\mu\in\mathring{L}$, then $|\mu|$ is the modulus of the function μ in the ordinary sense;

(ii) $|\varphi\mu| = |\varphi||\mu|$ for all $\varphi\in C(I)$.

On the other hand, if μ and ν are two measures on I, we write $\mu\le\nu$ iff $\mu(J)\le\nu(J)$ for all bounded intervals J such that $\bar{J}\subset I$.

Suppose I is unbounded on the right. A measure μ on I is said to be bounded on the right if and only if there exist two numbers x_0 and k such that $|\mu|\le k$ for $x>x_0$; i.e. $|\mu|(J)\le k|J|$ for all bounded intervals $J\subset[x_0, +\infty[$. On the other hand, we say that μ **converges to**

a number c as $x \to +\infty$, iff for every $\varepsilon > 0$, there exists a real x_0 such that $|\mu - c| < \varepsilon$ for $x > x_0$. It is readily seen that these concepts coincide with the classical ones if μ is a function. Besides, the preceding lemma for the "o" and "O" symbols keep true if f is a measure.

These remarks suggest the following refinement of the concept of convergence for distributions:

6.6.7. DEFINITION. Let $f \in \mathscr{D}'(I)$, $n \in I\!N_0$ and $\lambda \in \mathbb{C}$. We write $f \xrightarrow[n]{} \lambda$ as $x \to +\infty$ if and only if there exist a real x_0 and a measure F such that $f = D^n F$ and $\dfrac{F(x)}{x^n} \to \dfrac{\lambda}{n!}$ (in measure sense) as $x \to +\infty$. On the other hand, if $\varphi \in C^\infty(I)$, we shall write $f \in o_n(\varphi)$ as $x \to +\infty$ iff there exists x_0 and f_0 such that $f = \varphi f_0$ for $x > x_0$ and $f_0 \xrightarrow[n]{} 0$ as $x \to +\infty$.

The expression "$f \in O_n(\varphi)$" can be analogously defined and the "dual" concepts of the preceding ones can be introduced as follows:

6.6.8. DEFINITION. Let $n \in I\!N_0$, $f \in C^n(I)$ and $\lambda \in \mathbb{C}$. We shall write $f \xrightarrow{n} \lambda$ as $x \to +\infty$ iff f tends to λ and $f^{(k)} \in o(x^{-k})$ as $x \to +\infty$, for $k = 1, \ldots, n$ (in the ordinary sense). We write $f \in o^n(\varphi)$ as $x \to +\infty$ iff there exists x_0 and f_0 such that $f = \varphi f_0$ for $x > x_0$ and $f_0 \xrightarrow{n} 0$ as $x \to +\infty$.

Thus, it is readily seen that:

6.6.9. *If $f \xrightarrow[n]{} \lambda$ as $x \to +\infty$ and $g \xrightarrow{n} \mu$ as $x \to +\infty$, then $fg \xrightarrow[n]{} \lambda\mu$ as $x \to +\infty$.*

6.6.10. *If $f \xrightarrow[n]{} \lambda$ as $x \to +\infty$ and if h is a C^n mapping of an interval I^* into I such that $h \to +\infty$ as $t \to +\infty$ and $h' \xrightarrow{n} c$, with $c \neq 0$ and $c \neq \infty$, then $f \circ h \xrightarrow[n]{} \lambda$ as $t \to +\infty$.*

6.6.11. *If $f \in o_n(\varphi)$ and $g \in o^n(\psi)$ on the right, then $fg \in o_n(\varphi\psi)$ on the right, and analogously for the "O" symbol.*

6.7. Scalar products. Definition of distributions according to Sobolev-Schwartz

We shall say the two distributions f and g are multipliable if and only if the product fg exists in some of the senses considered in chapter IV. That being so:

6.7.1. DEFINITION. If two distributions f and g on an interval I are multipliable and fg is integrable on I, then $\int_I fg$ will be called the **symmetrical scalar product** or simply the **scalar product** of f by g and is denoted by $\langle f, g\rangle$:

$$\langle f, g\rangle = \int_I fg.$$

Obviously, the scalar product is in fact *symmetrical* (or *commutative*). Moreover it is *bilinear*: for all $\lambda, \mu \in \mathbb{C}$, we have

$$\langle \lambda f_1 + \mu f_2, g\rangle = \lambda\langle f_1, g\rangle + \mu\langle f_2, g\rangle$$

whenever $\langle f_1, g\rangle$ and $\langle f_2, g\rangle$ exist and analogously on the right.

Observe that every distribution f on I can be written in the form $f = u + iv$, where u, v are *real-valued distributions* (i.e., of the form $u = D^n U$, $v = D^n V$, where U and V are real-valued continuous functions on I). Then we put $\bar{f} = u - iv$ (*conjugate of* f). It is readily seen that if $\langle f, g\rangle$ exists, then $\langle f, \bar{g}\rangle$ exists also.

We call $\int_I f\bar{g}$ the **hermitic scalar product** or simply the **hermitic product** of f by g, and it is denoted by $\langle f | g\rangle$:

$$\langle f | g\rangle = \int_I f\bar{g}.$$

The hermitic product is not commutative:

$$\langle f|g\rangle=\overline{\langle g|f\rangle};$$

but it is of course, linear on the left.

Remember that any function $f \in L^2(I)$ (square summable function on I) is locally summable, hence a distribution. It is well known that if $f, g \in L^2(I)$, then $fg \in L(I)$ so that $\langle f|g\rangle$ is a hermitic form on $L^2(I)$, which makes $L^2(I)$ a Hilbert space. The following is a classical theorem in functional analysis.

6.7.2. *If E is a Hilbert space, there is a one-to-one correspondence between the continuous linear functionals on E and the elements of E. The functional U corresponding to an element u of E is given by the formula:*

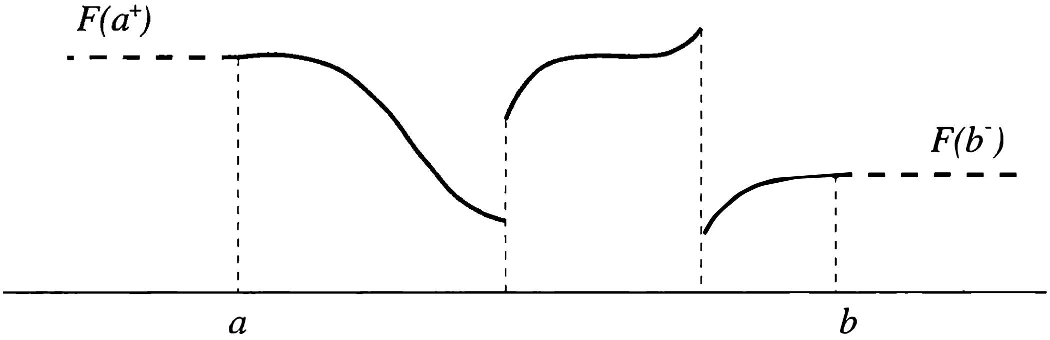
$$U(x)=\langle x|u\rangle \textit{ for all } x \in E.$$

Moreover, this correspondence is a vector isomorphism between E and E'. But the elements of E' (**covariant** vectors) do not behave like the elements of E (**contravariant** vectors) by change of bases; thus *it is not convenient in most cases to identify E' with E.*

For developing the study of scalar distributions, a remark about terminology is necessary. When I is a compact interval, the expression "measure on I" is commonly used with a meaning equivalent to that of "measure of an interval contained in I". For example, in this sense, δ may be considered as a measure on $I=[0, 1]$; *but the restriction of the δ distribution to* $[0, 1]$ *is* $D(\rho_I H)=0$. To avoid confusion, we shall say "measure in I" instead of "measure on I" for a distribution f of the form $f=DF$, when F is a standardized function of bounded variation on I. On the other hand, we shall denote by $M^*(I)$ the vector space of all measures on $/R$ which vanish outside I and by $M(I)$ the set of all measures on I. Observe that:

6.7.3. *If $I=[a, b]$, every measure μ in I can be uniquely extended as a measure $\widetilde{\mu} \in M^*(I)$ such that $\widetilde{\mu}[a, a]=\widetilde{\mu}[b, b]=0$.*

In fact, if $\mu \in M(I)$, then $\mu = DF$ where F is a function of bounded variation on I. Now F can be uniquely extended to a function $\widetilde{F}$ of bounded variation on $/R$ such that $\widetilde{F}(x) = F(a^+)$ for $x < a$ and $\widetilde{F}(x) = F(b^-)$ for $x > b$.



Hence, if we put $\widetilde{\mu} = D\widetilde{F}$, we have $\widetilde{\mu} \in M^*(I)$, $\widetilde{\mu} = [a, a] = \widetilde{\mu}[b, b] = 0$, and it is readily seen that $\widetilde{\mu}$ is uniquely determined by μ. We call $\widetilde{\mu}$ the **minimal extension** of μ to $/R$.

Another classical theorem in functional analysis is the following:

6.7.4. F. RIESZ THEOREM. *There is a one-to-one correspondence between the measures $f \in M^*(I)$ and the continuous linear functionals u on $C(I)$. This correspondence $f \leftrightarrow u$ is given by:*

$$u(\varphi) = \langle f, \varphi \rangle = \int_I f\varphi, \quad \forall \varphi \in C(I).$$

We are going to deduce some important consequences from this formula. We shall denote by $M_n^*(I)$ the set of all distributions of order $\leq n$ on $/R$ vanishing outside I. Suppose $I = [a, b]$; then

6.7.5. *Every distribution $f \in M_n^*(I)$ can be written in the form:*

$$f = D^n F_0 + \sum_{k=0}^{n-1} c_k \delta^{(k)}(x-a)$$

where $F_0 \in M^*(I)$ and $c_0, \ldots, c_{n-1} \in \mathbb{C}$.

PROOF. Consider $f \in M_n^*(I)$. Then f is of the form $f = D^n F$ where $F \in M(/R)$. On the other hand, since $f = 0$ outside I, F reduces

to polynomials P and P^* of degree $<n$, respectively on the left and on the right of I. Put $\widetilde{F}=F-P^*$; then $\widetilde{F}=0$ for $x>b$ and $f=D^n\widetilde{F}$. We can suppose that $\widetilde{F}=F=0$ for $x>b$. Set:

$$P_0=\begin{cases}P & \text{for } x<a\\ 0 & \text{for } x\geq a\end{cases}\qquad F_0=\begin{cases}0 & \text{for } x<a\\ F & \text{for } x\geq a.\end{cases}$$

Then $F=F_0+P_0$ and $F_0\in M^*(I)$. On the other hand, P_0 is a polynomial of degree $<n$ for $x<0$ and zero for $x>0$, so that D^nP_0 is of the form:

$$\sum_{k=0}^{n-1} c_k\delta^{(k)}(x-a).$$

Consequently

$$f=D^nF_0+\sum_{k=0}^{n-1} c_k\delta^{(k)}(x-a). \ \blacklozenge$$

Let now φ be any C^n function on $/R$ and $f\in M_n^*(I)$. Then, of course, $\varphi f\in M_n^*(I)$; so that φf is integrable on $/R$. Put

$$f=D^n\mu+\sum_{k=0}^{n-1} c_k\delta^{(k)}(x-a) \text{ with } \mu\in M^*(I).$$

A primitive of $\varphi D^n\mu$ will then be:

$$\Phi=\sum_{k=0}^{n-1}(-1)^k\binom{n}{k}D^{n-k-1}(\varphi^{(k)}\mu)+(-1)^n\int_{a^-}^{x^+}\varphi^{(n)}\mu$$

and since $\mu=0$ outside I, we have:

$$\int_I \varphi D^n\mu=\Phi(b^+)-\Phi(a^-)=(-1)^n\int_{a^-}^{b^+}\varphi^{(n)}\mu.$$

On the other hand (cf. 6.3.6. and example 1 in 6.3.)

$$\langle \varphi, \delta^{(k)}(x-a)\rangle=(-1)^k\varphi^{(k)}(a).$$

Hence:

6.7.6. $$\langle f, \varphi\rangle = \sum_{k=0}^{n} (-1)^k c_k \varphi^{(k)}(a) + (-1)^n \int_I \varphi^{(n)} \mu .$$

Observe that in this formula the values of $\varphi(x)$ for $x \notin I$ do not matter. So we may extend this formula by definition to all functions $\varphi \in C^n(I)$. In $C^n(I)$ it is common to define the norm by:

6.7.7. $$\|\varphi\|^n = \sup_{x \in I} \{|\varphi(x)|, |\varphi'(x)|, \ldots, |\varphi^{(n)}(x)|\}.$$

Then $C^n(I)$ becomes a Banach space and the convergence of a sequence (φ_p) to 0 in this norm means the convergence of the n sequences (φ_p), $(\varphi_p^{(1)}), \ldots, (\varphi_p^{(n)})$ to 0 uniformly on I. Now we have the following consequence of the Riesz theorem:

6.7.8. THEOREM. *There is an isomorphism $f \leftrightarrow u$ between the vector spaces $M_n^*(I)$ and $C^n(I)'$ defined by $u(\varphi) = \langle f, \varphi\rangle$ $\forall \varphi \in C^n(I)$.*

PROOF. a) Take $f \in M_n^*(I)$. Then f is of the form:

$$f = D^n \mu + \sum_{k=0}^{n-1} c_k \delta^{(k)}(x-a)$$

with $\mu \in M^*(I)$ and:

$$u(\varphi) = \sum_{k=0}^{n} (-1)^k \varphi^{(k)}(a) + (-1)^n \int_I \varphi^{(n)} \mu .$$

This defines clearly a linear functional u on $C^n(I)$. On the other hand, we have:

$$|u(\varphi)| \leq \sum_{0}^{n-1} |c_k| \, \|\varphi\|^n + \|\varphi\|^n |\mu|(I)$$

which shows that $u(\varphi) \to 0$ as $\varphi \to 0$.

b) Take $u \in C^n(I)'$ and set, for all $\psi \in C(I)$, $v(\psi) = (-1)^n u(\Im^n \psi)$, with $\Im \psi(x) = \int_a^x \psi(\xi)\, d\xi$. Then v is a continuous linear functional on

$C(I)$ and by 6.7.3., there exists $\mu \in M^*(I)$ such that $v(\psi)=\langle \mu, \psi\rangle$. Besides, every function $\varphi \in C^n(I)$ can be written in the form:

$$\varphi(x)=\sum_{k=0}^{n-1} \frac{\varphi^{(k)}(a)}{k!}(x-a)^k+\mathfrak{I}^n\varphi^{(n)}, \quad \text{with } \varphi^{(n)}\in C(I).$$

Hence, if we set $c_k=(-1)^k u\left(\frac{(\hat{x}-a)^k}{k!}\right)$ we find:

$$u(\varphi)=\sum_{k=0}^{n}(-1)^k c_k \varphi^{(k)}(a)+(-1)^n\int_I \varphi^{(n)}\mu$$

since $u(\mathfrak{I}^n\varphi^{(n)})=(-1)^n v(\varphi^{(n)})=(-1)^n\langle u, \varphi^{(n)}\rangle$.

So if we put $f=\sum_0^{n-1} c_k\delta^{(k)}(x-a)+D^n\mu$, we have $u(\varphi)=\langle f, \varphi\rangle$ for all $\varphi\in C^n(I)$. ♦

We shall denote by $C_*^n(I)$ the set of all C^n functions φ on I such that $\varphi^{(k)}(a)=\varphi^{(k)}(b)=0$ for $k=0,\dots, n$. It is obvious that $C_*^n(I)$ is a vector subspace of $C^n(I)$. Also, every $\varphi\in C_*^n(I)$ can be uniquely extended as a C^n function on $/R$ vanishing outside I, so that $C_*^n(I)$ can also be identified with a subspace of $C^n(/R)$. We shall consider $C_*^n(I)$ provided with the norm $\|\cdot\|^n$ defined by 6.7.7. On the other hand $M_n(I)$ is the vector space of all distributions f on I of the form $f=D^n\mu$ with $\mu\in M(I)$. Now from 6.7.8., follows:

6.7.9. THEOREM. *There is an isomorphism $f \leftrightarrow g$ between $M_n(I)$ and $C_*^n(I)'$ defined by $u(\varphi)=\langle f, \varphi\rangle$ for all $\varphi\in C_*^n(I)$. Besides $\langle f, \varphi\rangle=(-1)^n\langle \mu, \varphi^{(n)}\rangle$ if $f=D^n\mu$.*

PROOF. a) Take $f=D^n\mu$ where $\mu\in M(I)$. Then if we set $\widetilde{f}=D^n\widetilde{\mu}$ where $\widetilde{\mu}$ is the minimal extension of μ to $/R$ (cf. 6.7.3.), f defines a functional $\widetilde{\mu}\in C^n(I)'$ whose restriction to $C_*^n(I)$ is obviously an element u of $C_*^n(I)'$ such that

$$u(\varphi)=(-1)^n\int_{a^-}^{b^+}\varphi^{(n)}\widetilde{\mu} \quad \text{for all} \quad \varphi\in C_*^n(I).$$

But as $\widetilde{\mu}[a, a] = \widetilde{\mu}[b, b] = 0$, it is easily seen that

$$\int_{a^-}^{b^+} \varphi^{(n)} \widetilde{\mu} = \int_{a^+}^{b^-} \varphi^{(n)} \widetilde{\mu} \, .$$

So, we can write $u(\varphi) = \langle f, \varphi \rangle = (-1)^n \langle u, \varphi^{(n)} \rangle$.

b) Take now $u \in C_*^n(I)'$. Observe that for every function $\varphi \in C^n(I)$ there is one and only one function $\varphi_0 \in C_*^n(I)$ such that:

$$\varphi_0(x) = \varphi(x) - \sum_{k=0}^{n} a_k (x-a)^k - (x-a)^n \sum_{k=0}^{n} b_k (x-b)^k,$$

where the coefficients a_k and b_k can be obtained as linear combinations of the values $\varphi^{(k)}(a)$, $\varphi^{(k)}(b)$ for $k=0, 1, \ldots, n$. We shall denote by π the mapping $\varphi \to \varphi_0$ of $C^n(I)$ onto $C_*^n(I)$. Since the a_k, b_k are linear combinations of the $\varphi^{(k)}(a)$, $\varphi^{(k)}(b)$, it is readily seen that π is a projection, i.e. a linear mapping such that $\pi\varphi_0 = \varphi_0$ for all φ_0 in $C_*^n(I)$ and continuous. So if we set $\widetilde{u}(\varphi) = u(\pi\varphi)$, u will be a continuous linear functional on $C^n(I)$ extending u; hence there exists a distribution $\widetilde{f} \in M_n^*(I)$ such that $\widetilde{u}(\varphi) = \langle \widetilde{f}, \varphi \rangle$ and therefore, if we put $f = \rho_I \widetilde{f}$, it follows that $u(\varphi) \equiv \langle f, g \rangle$. Finally suppose $\langle f, \varphi \rangle = \langle g, \varphi \rangle$ for all $\varphi \in C_*^n(I)$, with $f = D^n \mu$, $g = D^n \upsilon$, $\upsilon, \mu \in M(I)$. Then if we put $\widetilde{f} = D^n \widetilde{\mu}$, $\widetilde{g} = D^n \widetilde{\upsilon}$, where $\widetilde{\mu}, \widetilde{\upsilon}$ are the minimal extensions of μ, υ, it follows that $\langle \widetilde{f}, \varphi \rangle = \langle \widetilde{g}, \varphi \rangle$ for all $\varphi \in C^n(I)$, so that $\widetilde{f} = \widetilde{g}$ (by theorem 6.7.8.) and therefore $f = g$. ♦

We shall now denote by $C_*^\infty(I)$ the space of all C^∞ functions φ on $I = [a, b]$ such that $\varphi^{(n)}(a) = \varphi^{(n)}(b) = 0$, for all $n \in /N_0$. Such functions can be identified with the C^∞ functions on $/R$ with support contained in I. In that space there is defined a topology making C^∞ an (F)-space by means of the sequence of norms $\|\cdot\|^n$. This being so:

6.7.10. THEOREM. *There exists a vector isomorphism $f \leftrightarrow u$ between $\mathscr{D}(I)$ and $C_*^\infty(I)'$ which is given by the formula $u(\varphi) \equiv \langle f, \varphi \rangle$. Besides*

6.7.11. $$\langle Df, \varphi\rangle = -u(\varphi'), \quad \forall \varphi \in C_*^\infty(I)$$

and, if J is a compact interval contained in I, then

6.7.12. $$u_J(\varphi) = \langle f_J, \varphi\rangle, \quad \forall \varphi \in C_*^\infty(I)$$

where u_J and f_J are the restrictions of u and f respectively to $C_*^\infty(J)$ and J.

PROOF. a) Take $f \in \mathscr{D}(I)$. Then there exists an integer n such that $f = D^n F$ with $F \in M(I)$. So if we set $u(\varphi) = \langle f, \varphi\rangle = (-1)^n \langle F, \varphi\rangle$ for all $\varphi \in C_*^\infty(I)$, u is clearly a continuous linear functional on $C_*^\infty(I)$.

b) Take $u \in C_*^\infty(I)'$. Now, a fundamental system of neighborhoods of 0 in $C_*^\infty(I)$ is given by the sets:

$$\varepsilon B_n = \{x : \|x\|^n < \varepsilon\}, \quad \varepsilon > 0, \quad n = 0, 1, \ldots .$$

Hence, for any $\delta > 0$, there exists an $\varepsilon > 0$ and n such that $\varepsilon u(B_n) < \delta$. But this means that u is continuous with respect to the norm $\|\cdot\|^n$ on $C_*^\infty(I)$ and we shall see in the next paragraph that $C_*^\infty(I)$ is dense in the normed space $C_*^n(I)$. So u can be uniquely extended as a functional $\widetilde{u} \in C_*^n(I)'$; i.e. there exists one and only one distribution $f \in M(I)$ such that $u(\varphi) = \langle f, \varphi\rangle$.

Finally 6.7.11. and 6.7.12. are easy consequences of the preceding results. ♦

This theorem shows that the dual space of $C_*^\infty(I)$ affords a model of the axiom system in 2.2. if we identify every function $f \in C(I)$ with the functional u such that $u(\varphi) = \int_I f\varphi$ and if we define Du by $Du(\varphi) = -u(\varphi')$.

As a matter of fact, Sobolev had first (in 1936) the idea of taking such functionals as *generalized functions* (of real variables). This method was developed in a systematic way by L. Schwartz in

1944-45. So, according to Schwartz the elements of $C_*^\infty(I)'$ are called distributions on I. The sum of two distributions u and v is the sum of the functionals u and v in the usual sense, the restriction of u to an interval $J \subset I$ is the restriction of u to $C_*^\infty(J)$, and so forth.

Till now, we have been concerned only with a compact interval J. Let us consider now an open interval Ω in $/R$ and let us denote by $C_*^\infty(\Omega)$ the set of all C^∞ functions on Ω with bounded carrier, contained in Ω. According to Schwartz, $C_*^\infty(\Omega)$ is provided with the topology obtained as the inductive limit of the topologies of the (F)-spaces $C_*^\infty(I)$. Then, a linear functional u on $C_*^\infty(\Omega)$ is continuous if and only if the restriction u_I is continuous. This being so, Schwartz called the elements of $C_*^\infty(\Omega)'$ **distributions** on Ω. But now theorem 6.7.10. leads directly to the following:

6.7.13. COROLLARY. *There is a vector isomorphism $f \leftrightarrow u$ between $\overline{\mathscr{D}}(\Omega)$ and $C_*^\infty(\Omega)'$.*

The result can obviously be extended to any open set Ω in $/R$. It must be observed however that Schwartz denotes by $\mathscr{D}(\Omega)$ the space $C_*^\infty(\Omega)$ and by $\mathscr{D}'(\Omega)$ the space of global distributions on Ω. On the other hand, Schwartz defines the topology on $\overline{\mathscr{D}}(\Omega)$ as the strong topology of $C_*^\infty(\Omega)'$. But it can be proved without difficulty that this topology is the same one that we have defined directly in chapter V, i.e. the isomorphism in 6.7.10. is a topological isomorphism.

The functional theory of distributions requires some warning in order to avoid misunderstandings. This begins already with measures. Observe, for example, that if f is a locally summable function on $/R$ and μ the corresponding measure, we have $\mu(I) = \int_I f(x)\,dx$ for every bounded interval I; but if we consider a one-to-one C^1 mapping h of $/R$ on to $/R$, the transformed μ^* of μ by h is given by

$$\mu^*(I) = \int_I f(h(t))h'(t)\,dt$$

so that μ^* is defined by $(f \circ h)h'$, instead of by $f \circ h$. *Hence functions and measures behave differently by change of variables, so that the identification of functions with measures works only as far as a substitution $x = h(t)$ with $h' \neq 1$ is concerned.*

The same difference arises between global distributions (as we have defined them) and distributions according to Schwartz: the first behave like functions and the second like measures, by change of variable. *In such a situation, functions cannot be identified with linear functionals, since the first are contravariant vectors and the second covariant vectors.*

As a last example, let us consider the space $\mathcal{H}^n(\Omega)$ of all functions $f \in C^n(\Omega)$ such that $f^{(k)}$ is square-summable on Ω for $k = 0, \ldots, n$ ($n \in \mathbb{N}_0$), provided with the following definition of hermitic product:

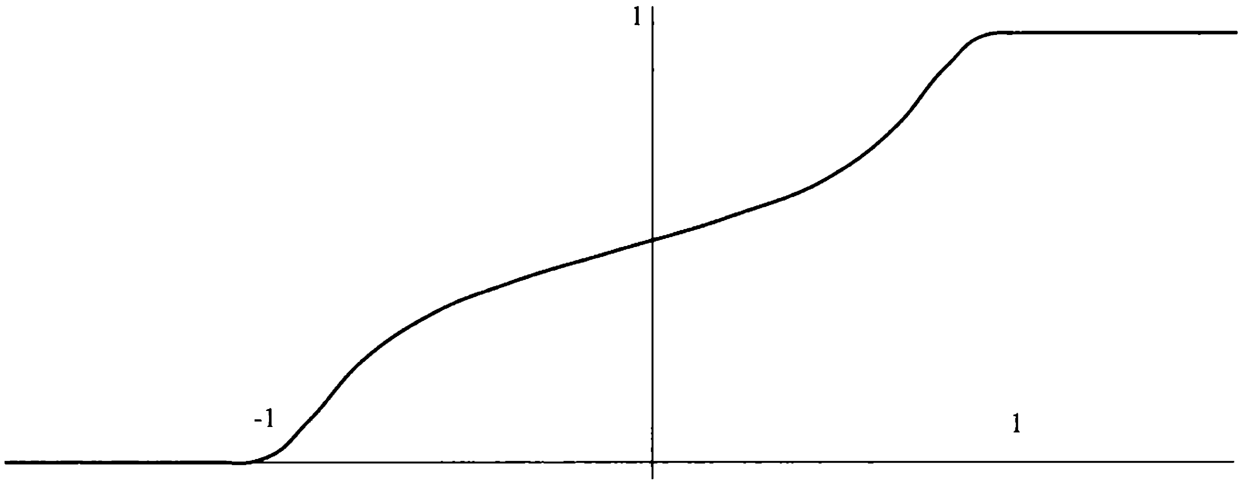
$$\langle f, g \rangle = \sum_{k=0}^{n} \int_{\Omega} f^{(k)} \bar{g}^{(k)}.$$

Then $\mathcal{H}^n(\Omega)$ is a Hilbert space, whose dual is just isomorphic to $\mathcal{H}^n(\Omega)$. But according to Schwartz, $\mathcal{H}^n(\Omega)'$ is identified with the space $\mathcal{H}_n(\Omega)$ of all distributions f of the form $f = \sum_{\nu=1}^{p} D^{k_\nu} f_\nu$ where p is an arbitrary integer ≥ 0 and $0 \leq k_\nu \leq n$, $f_\nu \in L^2(\Omega)$ for $\nu = 1, 2, \ldots, r$. Obviously this identification requires special care.

6.8. The approach of functions or distributions by means of C^∞ functions. Distributions according to Mikusinski

Consider the function y defined as follows:

$$y(x) = \begin{cases} \left(1 + exp \dfrac{x}{x^2 - 1}\right)^{-1} & \text{for } -1 < x < 1 \\ 0 & \text{for } \quad x \leq -1 \\ 1 & \text{for } \quad x \geq 1 \end{cases}$$



It can be seen by elementary calculations that y is a C^∞ function on $/R$ increasing from 0 to 1. Set

6.8.1. $H_n(x)=y(nx)$ and $\delta_n=H_n'$ for $n=1,2,\dots$.

Then $\delta_n\in C^\infty$, $\delta_n(x)=0$ if $|x|\geq\frac{1}{n}$, $\delta_n(x)>0$ if $|x|<\frac{1}{n}$ and $\int_{-1}^{1}\delta_n(x)dx=1$ for $n=1,2,\dots$. It follows that $\delta_n\to\delta$ (cf. 5.3.).

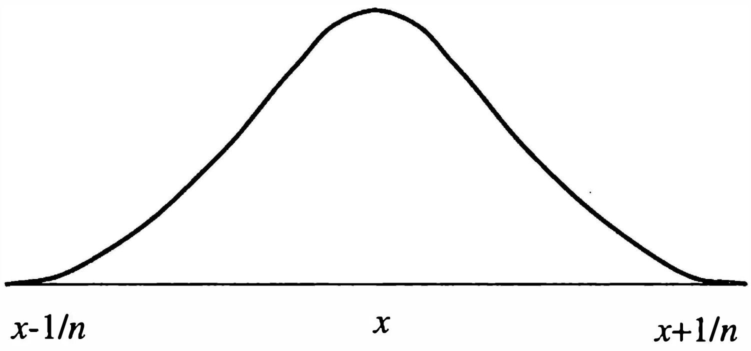
Moreover:

6.8.2. LEMMA. *If* $f\in C(/R)$ and

$$\varphi_n(x)=\int_{-\infty}^{+\infty}\delta_n(x-t)f(t)dt \quad for \quad x\in /R,\ n=1,2,\dots ,$$

then $\varphi_n\in C^\infty(/R)$ *for all n and* $\varphi_n\to f$ *uniformly on each compact interval.*

PROOF. Since $\delta_n(x-t)=0$ for $|x-t|>\frac{1}{n}$,



we have

$$\varphi_n(x)=\int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \delta_n(x-t)\, f(t)\, dt\ , \quad \forall x\in /R,\ n=1,2,\dots .$$

Consider now a compact interval $J=[a,\ b]$ and put $K=[a-1,\ b+1]$. Then:

$$\varphi_n(x)=\int_K \delta_n(x-t)\, f(t)\, dt\ , \quad \forall x\in J,\ n=1,2,\dots$$

and since $\delta_n\in C^\infty(/R)$ for all n, it is readily seen that $\varphi_n\in C^\infty(/R)$ for all n. On the other hand, by the *mean value theorem* there exists for each $x\in J$ and each $n=1,2,\dots$ a real ξ such that $|x-\xi|<\dfrac{1}{n}$ and

$$\varphi_n(x)=f(\xi)\int_{x-\frac{1}{n}}^{x+\frac{1}{n}} \delta_n(x-t)\, dt=f(\xi)\,.$$

But f is *uniformly continuous* on the compact interval K. So, for every $\varepsilon>0$ there exists an integer r such that $|f(x)-f(x')|<\varepsilon$, whenever $x,\ x'\in K$ and $|x-x'|<\dfrac{1}{r}$. Hence $|f(x)-\varphi_n(x)|=|f(x)-f(\xi)|<\varepsilon$ for all $x\in J$ and $n>r$, which proves the lemma. ♦

6.8.3. THEOREM. *Let I be any interval, p an integer ≥ 0 and $f\in C^p(I)$. Then there exists a sequence of functions $\varphi_n\in C^\infty(/R)$ such that $\varphi_n^{(k)}\to f^{(k)}$ uniformly on each compact interval $J\subset I$, for $k=0,\dots,p$.*

PROOF. a) Suppose $I = /R$. Consider the sequence φ_n defined as in the lemma. Observing that

$$D_x\delta_n(x-t) = -D_t\delta_n(x-t) \text{ and } \delta_n^{(k)}(x-t) = 0 \text{ for } |x-t| > \frac{1}{n},\ k = 0, 1, \dots,$$

it is easily seen that

$$\varphi_n^{(k)}(x) = \int_{/R} \delta_n^{(k)}(x-t)\, f(t)\, dt$$

$$= (-1)^k \int_{/R} [D_t^{(k)}\,\delta_n(x-t)]\, f(t)\, dt$$

$$= \int_{/R} \delta_n(x-t)\, f^{(k)}(t)\, dt\,, \quad \forall x \in /R,\ k=0, 1, \dots, p,\ n=1, 2, \dots .$$

Now, applying the lemma to the functions $f^{(k)}$, it is readily seen that $\varphi_n^{(k)} \to f^{(k)}$ uniformly on each compact interval.

b) Suppose I is closed. Then it is possible to extend f to a function $\widetilde{f} \in C^p(/R)$ (for example, if I is bounded, it is possible to make f equal to two polynomials outside I). Now, if we set

$\widetilde{\varphi}_n = \int_{/R} \delta_n(x-t)\, \widetilde{f}(t)\, dt$ and $\varphi_n = \rho_I\, \widetilde{\varphi}_n$ the theorem is proved in this case.

c) Suppose I is open. Then, there exists a one-to-one C^∞ mapping h of $/R$ onto I. So if we put $g = f \circ h$, $\psi_n = \int_{/R} \delta_n(x-t)\, g(t)\, dt$ and $\varphi_n = \psi_n \circ h^{-1}$ the theorem is proved in this case.

d) Suppose finally that I is half-open. Then it is possible to extend f as a continuous function $\widetilde{f}$ on an open interval $\widetilde{I} \supset I$, which reduces the proof to the previous case. ♦

We are going to establish a similar theorem for the space $C_*^p(I)$ but, for that purpose, it is convenient to prove, first, a lemma. It is sufficient to consider the case when I is compact, $I=[a,\ b]$.

6.8.4. LEMMA. *Let f be a C^∞ function on I such that $f^{(k)}(a)=f^{(k)}(b)=0$* for $k=0,\dots,p$. *Then there exists a sequence of functions $\varphi_n \in C_*^\infty(I)$ such that $\varphi_n^{(k)} \to f^{(k)}$ uniformly on I, for $k=0,\dots,p$.*

PROOF. Set $\alpha_n(x)=H_n\left(x-a-\frac{1}{n}\right)-H_n\left(x-b+\frac{1}{n}\right)$, $\forall x \in I\!R$, $n=1,2,\dots$, where H_n is given by 6.8.1. It is easily seen that for $n>\frac{4}{b-a}$, we have $\alpha_n(x)=0$ outside I, $\alpha_n(x)=1$ on $\left[a+\frac{2}{n},\ b-\frac{2}{n}\right]$ and $|\alpha_n(x)|\leq 1$ for all x. On the other hand, since $H_n(x)\equiv y(nx)$, we have $H_n^{(k)}(x)=n^k y^{(k)}(nx)$ for $k=0,\dots,n$. Then if we put $I_n=\left[a+\frac{2}{n},\ b-\frac{2}{n}\right]$ and $M=\max\limits_{-1\leq x\leq 1}(|y(x)|,\dots,|y^{(p)}(x)|)$, it is readily seen that for all $n>\frac{4}{b-a}$ and $k=0,1,\dots,p$,

6.8.5. $$|\alpha_n^{(k)}(x)|\leq 2Mn^k \text{ on } I\backslash I_n.$$

Thus, set $\varphi_n=\alpha_n f$ for $n=1,\ 2,\ \dots$. Then $\varphi_n \in C_*^\infty$ and $\varphi_n=f$ on I_n for $n>\frac{4}{b-a}$. On the other hand:

6.8.6. $$\varphi_n^{(k)}-f^{(k)}=\sum_{\nu=0}^{k}\binom{k}{\nu}(\alpha_n-1)^\nu f^{(k-\nu)}, \text{ for } k=0,\ 1,\dots .$$

But since $f^{(k)}(a)=f^{(k)}(b)=0$ for $k=0,\ 1,\dots,\ p$, it follows that:

$$\lim_{x\to a^+}\frac{f^{(k)}(x)}{(x-a)^{p-k}}=\lim_{x\to b^-}\frac{f^{(k)}(x)}{(x-b)^{p-k}}=0 \text{ for } k=0,1,\dots,p$$

and therefore there exists a sequence of numbers ε_n such that $\varepsilon_n\to 0$ and:

$$|f^{(k)}(x)|\leq\frac{\varepsilon_n}{n^{p-k}} \text{ on } I\setminus I_n, \text{ for } k=0,\dots,p.$$

From here, from 6.8.5. and from 6.8.6., follows:

$$|\varphi_n^{(k)}-f^{(k)}|\leq 2^k(2M+1)\frac{\varepsilon_n}{n^{p-k}}, \text{ for } k=0,\dots,p. \ \blacklozenge$$

6.8.7. THEOREM. *For every $f\in C_*^p(I)$, there exists a sequence of functions $\varphi_n\in C_*^\infty(I)$ such that φ_n converges to f in the norm $\|\cdot\|^p$.*

PROOF. Consider $f\in C_*^p(I)$. By 6.8.3., there exists a sequence of functions $\psi_n\in C^\infty(I)$ such that $\psi_n\to f$ in the normed space $C^p(I)$. We have also seen in the proof of 6.7.9. that there exists a continuous projection π of $C^p(I)$ onto $C_*^p(I)$ such that $\pi(\varphi)-\varphi$ is a polynomial for every $\varphi\in C^p(I)$. Set $\chi_n=\pi\psi_n$; then $\chi_n\in C^\infty(I)$ for all n and $\|\chi_n-f\|^p\to 0$. Finally, by the lemma, there exists for every n a sequence of functions $\chi_{n_1}, \chi_{n_2}, \dots$, belonging to $C_*^\infty(I)$ and converging to χ_n in $\|\cdot\|^p$. Then, from the double sequence χ_{n_k}, we can select a sequence of functions $\varphi_n\in C_*^\infty(I)$ converging to f in the norm $\|\cdot\|^p$. ♦

Consider now a distribution f on $/R$ and set:

$$\varphi_n(x)=\int_{/R}\delta_n(x-t)f(t)\,dt$$

where δ_n is given by 6.8.1. Then if $f=D^nF$ with $F\in C(/R)$, it is readily seen that:

$$\varphi_n(x)=(-1)^n\int_{/R}\delta_n(x-t)F(t)\,dt$$

and from 6.8.2., it is concluded that $\varphi_n\to f$ in the distributional sense. This result can be extended to every $f\in\overline{\mathscr{D}}(/R)$ observing that on

every compact interval I, f reduces to a distribution. Finally, if I is any interval in $\mathbb{R}$, we can see by the technique used in the proof of 6.8.3., that:

6.8.8. THEOREM. *For every $f \in \overline{\mathscr{D}}(I)$ there exists a sequence of functions $\varphi_n \in C^\infty(I)$ such that $\varphi_n \to f$.*

Remember that the space $\overline{\mathscr{D}}(I)$ is complete. Theorem 6.8.8. has suggested to Mikusinski a construction of the space $\overline{\mathscr{D}}(I)$ by completion of $C^\infty(I)$ with respect to the distributional topology. According to this approach, a *fundamental sequence* is a sequence of functions $\varphi_n \in C^\infty(I)$ such that, for every compact interval $J \subset I$, there exists an integer p and a sequence of functions $\Phi_n \in C^\infty(I)$ (dependent upon J) such that $\varphi_n = D^n \Phi_n$ on J and Φ_n is uniformly convergent on I. Two fundamental sequences (φ_n) and (ψ_n) are said to be equivalent if and only if for every compact interval $J \subset I$, there exists an integer p and two sequences of function Φ_n, ψ_n in $C^\infty(I)$ such that $\varphi_n = D^p \Phi_n$, $\psi_n = D^p \Psi_n$ and $\Phi_n - \Psi_n \to 0$ uniformly on J. This turns out to be actually an equivalence relation. Hence the corresponding equivalence classes are called **distributions** (i.e. global distributions according to our terminology).

REFERENCES

[1] S. LOJASIEWICZ. *Sur la valeur et la limite d'une distribution en un point.* Studia Math. 16 (1957) pág. 1-36.

[2] J. MIKUSINSKI-R. SIKORSKI. *The Elementary Theory of Distributions.* Panstrowe Wydunictwo Nankowe, Varsaw, I (1957), II (1961).

[3] J. SEBASTIÃO E SILVA. *Sur une construction axiomatique de la théorie des distributions.* Revista da Faculdade de Ciências de Lisboa (1954-55).

[4] L. SCHWARTZ. *Théorie des distributions, I, II.* PARIS (1950-51).

[5] S. L. SOBOLEV. *Méthode nouvelle à résoudre le probléme de Cauchy pour les équations linéaires hyperboliques normales.* Mat. Sbornik 1 (43), 39-72, (1936).

CHAPTER VII

DISTRIBUTIONS OF SEVERAL VARIABLES; FUNDAMENTAL CONCEPTS

7.1. Intervals in $/R^n$ space

Let n be any integer >1. Given two points, $\boldsymbol{a}=(a_1,\ldots, a_n)$ and $\boldsymbol{b}=(b_1,\ldots, b_n)$ in the $/R^n$ space, we shall write $\boldsymbol{a}<\boldsymbol{b}$, iff $a_j< b_j$ for $j=1,\ldots, n$, and $\boldsymbol{a}\leq\boldsymbol{b}$ iff $a_j\leq b_j$ for $j=1,\ldots, n$. Then the bounded intervals $]\boldsymbol{a}, \boldsymbol{b}[$, $[\boldsymbol{a}, \boldsymbol{b}]$, $]\boldsymbol{a}, \boldsymbol{b}]$, $[\boldsymbol{a}, \boldsymbol{b}[$ with the extremities $\boldsymbol{a}, \boldsymbol{b}$ are to be defined as in the case of one single dimension. For example, $[\boldsymbol{a}, \boldsymbol{b}]$ is the set of all points $\boldsymbol{x}$ of $/R^n$ such that $\boldsymbol{a}\leq\boldsymbol{x}\leq\boldsymbol{b}$ (a rectangle in $n=2$, a parallelepiped if $n=3$, etc). In turn, the set of all points $\boldsymbol{x}$ of $/R^n$ such that $\boldsymbol{a}<\boldsymbol{x}$ is the open interval $]\boldsymbol{a},+\infty_n[$, *unbounded on the right*. In any case, an interval I in $/R^n$ is the Cartesian product of n intervals in $/R$. For example, if $I=[\boldsymbol{a}, \boldsymbol{b}[$, then $I=I_1\times I_2\times\cdots\times I_n$, with $I_1=[a_1, b_1[,\ldots, I_n=[a_n, b_n[$.

In order to make the reciprocal of this statement also true, we shall call every Cartesian product of intervals $I_1,\ldots, I_n$ in $/R$ an interval I in $/R^n$. Then the interval is said to be degenerate, iff at least one of the intervals $I_1,\ldots, I_n$ is.

7.2. Distributions on an interval I in $/R^n$

Let I be any interval in $/R^n$, hence the Cartesian product of n intervals $I_1, \ldots, I_n$ in $/R$, and consider the space $C(I)$ (in short C) of all complex valued functions $f(\boldsymbol{x}) = f(x_1, \ldots, x_n)$, which are defined and continuous on I. As in the case of one single variable, $C(I)$ is a complex vector space (and even a complex commutative algebra), relatively to the usual algebraic operations. For each $k = 1, \ldots, n$ we shall denote by D_k the partial derivation operator with respect to x_k, that is

$D_k = \dfrac{\partial}{\partial x_k}$. Then, for each system $\boldsymbol{r} = (r_1, \ldots, r_n)$ of n integers $r_k \geq 0$, we put:

$$\boldsymbol{D}^{\boldsymbol{r}} = D_1^{r_1} \cdots D_n^{r_n}$$

and denote by $C^{\boldsymbol{r}}(I)$ the set of all functions f such that $\boldsymbol{D}^{\boldsymbol{k}} f$ (for $\boldsymbol{k} \leq \boldsymbol{r}$) exists and is continuous on I in the ordinary sense, *independently of the order in which the differentiation are performed.*

On the other hand, considering for each k a fixed point c_k in I_k, arbitrarily chosen, we shall put

$$\mathfrak{J}_k f(\boldsymbol{x}) = \int_{c_k}^{x_k} f(x_1, \ldots, \xi_k, \ldots, x_n)\, d\xi_k, \quad \forall f \in C, \ k = 1, \ldots, n.$$

The **integration operator** $\mathfrak{J}_k$ defined in this way, is obviously a linear mapping of the space C into itself. More generally, for every system $\boldsymbol{r} = (r_1, \ldots, r_n)$ of n integers, $r_k \geq 0$, we shall denote by $\mathfrak{J}^{\boldsymbol{r}}$ the operator $\mathfrak{J}_1^{r_1} \cdots \mathfrak{J}_n^{r_n}$. Obviously, $\mathfrak{J}_k$ is a right inverse of D_k, i.e., $D_k \mathfrak{J}_k f = f$, for any $f \in C$. More generally, for every system $\boldsymbol{r} = (r_1, \ldots, r_n)$ of non-negative integers, we have $\boldsymbol{D}^{\boldsymbol{r}} \mathfrak{J}^{\boldsymbol{r}} f = f$, $\forall f \in C$.

As each D_k *is not defined on the whole space* C (as a mapping of C into C), there arises the problem of enlarging the set C, in or-der that the operators $D_1, \ldots, D_n$ may be extended as mappings of the enlarged set into itself according to some natural conditions, which we are going to state precisely in the form of axioms. The new

set will be denoted by $\mathscr{D}(I)$ and its elements will be called **distributions** on I. The set $\mathscr{D}(I)$, provided with the n basic operators $D_1, \ldots, D_n$, is just defined, up to an isomorphism, by the following system of axioms:

AXIOM 1. *If $f \in C(I)$, then $f \in \mathscr{D}(I)$.*

AXIOM 2. *To each $f \in \mathscr{D}(I)$ and each $k=1, \ldots, n$ there corresponds an element $D_k f$* of *$\mathscr{D}(I)$ (the derivative of f with respect to x_k), in such a way that: (i) if f is a function having a derivative f'_{x_k}, with respect to x_k, in ordinary sense and continuous on I, then $D_k f$ coincides with f'_{x_k}; (ii) the operators $D_1, \ldots, D_k$ are mutually interchangeable, that is: $D_j D_k f = D_k D_j f$, for all j, $k=1, \ldots, n$, and all $f \in \mathscr{D}(I)$.*

DEFINITION. If $\boldsymbol{r}$ is any system $(r_1, \ldots, r_n)$ of n non-negative integers, then $\boldsymbol{D}^{\boldsymbol{r}} = D_1^{r_1} \cdots D_n^{r_n}$.

AXIOM 3. *For every $f \in \mathscr{D}(I)$ there exists a system $\boldsymbol{r}$ of n integers ≥ 0 and a function $F \in C(I)$ such that $f = \boldsymbol{D}^{\boldsymbol{r}} F$.*

AXIOM 4. *If $\boldsymbol{r}$ is a system of n integers $r_k \geq 0$ and $F, G \in C(I)$, then we have $\boldsymbol{D}^{\boldsymbol{r}} F = \boldsymbol{D}^{\boldsymbol{r}} G$ if and only if $F-G$ is of the form $F-G=\Theta_1 + \cdots + \Theta_n$, where Θ_k is a polynomial in x_k of degree $< r_k$ whose coefficients are continuous functions on I independent of x_k (for $k=1, \ldots, n$).*

More explicity, each Θ_k considered in this axiom is of the form:

$$\Theta_k = \sum_{\nu=0}^{r_k-1} a_{k\nu} x_k^{\nu}$$

where the coefficients $a_{k\nu}$ are continuous functions on I independent of x_k. We shall denote by $\mathscr{P}_{kr_k}$ the set of all functions Θ_k of this form (for $k=1, \ldots, n$) and by $\mathscr{P}_{\boldsymbol{r}}$ the set of all functions Θ of the form

$\Theta = \Theta_1 + \ldots + \Theta_n$ with $\Theta_k \in \mathcal{P}_{kr_k}$ (which we call **pseudo-polynomials** of degree $< \boldsymbol{r}$). Thus

$$\mathcal{P}_{\boldsymbol{r}} = \mathcal{P}_{1r_1} + \cdots + \mathcal{P}_{nr_n}.$$

In turn, the set of all systems $\boldsymbol{r}$ of n integers ≥ 0 will be denoted by $I\!N_0^n$.

As for the case of one single variable, it can be proved, in a similar way, that this axiomatic system is both consistent and categorical. The only essential difference arises in the proof of consistency, about the definition of the equivalent relation. We are going to see precisely what this difference consists of.

Axiom 3 says that every distribution on I is determined by a couple $(\boldsymbol{r}, F)$ where $\boldsymbol{r} \in I\!N_0^n$ and $F \in C(I)$. On the other hand, axiom 4 leads to define a relation $\sim$ in the set of all such couples in the following way: $(\boldsymbol{r}, F) \sim (\boldsymbol{s}, G)$ iff there exists a system $\boldsymbol{m}$ of integers such that $\boldsymbol{m} \geq \boldsymbol{r}, \boldsymbol{s}$ and

7.2.1. $$\mathfrak{I}^{\boldsymbol{m}-\boldsymbol{r}} F - \mathfrak{I}^{\boldsymbol{m}-\boldsymbol{s}} G \in \mathcal{P}_{\boldsymbol{m}}.$$

The difficulty arises just when it is necessary to prove that, if there exists at least one $\boldsymbol{m} \geq \boldsymbol{r}, \boldsymbol{s}$ satisfying 7.2.1., then every other system $\boldsymbol{h}$ such that $\boldsymbol{h} \geq \boldsymbol{r}, \boldsymbol{s}$ satisfies the corresponding condition. Now this can be proved with the aid of two lemmas:

LEMMA 1. *If $\Theta \in \mathcal{P}_{\boldsymbol{r}}$ then $\mathfrak{I}^{\boldsymbol{p}} \Theta \in \mathcal{P}_{\boldsymbol{r}+\boldsymbol{p}}$ for every $\boldsymbol{p} \in I\!N_0^n$.*

LEMMA 2. If $\Theta \in \mathcal{P}_{\boldsymbol{r}}$ and, in addition, $\Theta \in C^{\boldsymbol{p}}$ then $\boldsymbol{D}^{\boldsymbol{p}} \Theta \in \mathcal{P}_{\boldsymbol{r}-\boldsymbol{p}}$ for $\boldsymbol{p} \leq \boldsymbol{r}$.

In fact, suppose that these two lemmas are true and denote by $\boldsymbol{\mu}$ the least system of integers such that $\boldsymbol{\mu} \geq \boldsymbol{r}, \boldsymbol{s}$, that is, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ with $\mu_i = sup(r_i, s_i)$, $i = 1, \ldots, n$. Then, if $\boldsymbol{m}$ is any system $\geq \boldsymbol{r}, \boldsymbol{s}$, satisfying 7.2.1., we obtain, by applying $\boldsymbol{D}^{\boldsymbol{m}-\boldsymbol{\mu}}$ to both members of 7.2.1. and taking lemma 2 into account:

$$\mathfrak{I}^{\boldsymbol{\mu}-\boldsymbol{r}} F - \mathfrak{I}^{\boldsymbol{\mu}-\boldsymbol{s}} G \in \mathcal{P}_{\boldsymbol{\mu}}.$$

The remaining part of the proof is analogous to the one given for the case $n=1$. So, it is easily proved that the relation $\sim$ just defined is an equivalence relation, and the class of all couples equivalent to $(\boldsymbol{r}, F)$ is denoted by $[\boldsymbol{r}, F]$ etc. It is, however, convenient to observe that the derivation operators can now be defined in general by putting:

$$\boldsymbol{D}^{\boldsymbol{p}}[\boldsymbol{r}, F]=[\boldsymbol{r}+\boldsymbol{p}, F]$$

for every system $\boldsymbol{p}\in \mathbb{N}_0^n$; in particular, $D_1=D^{(1,0,\dots,0)},\dots, D_n=D^{(0,0,\dots,1)}$. The preceding definition shows immediately that these operators are interchangeable.

PROOF OF LEMMA 1. It is almost immediate. It will be sufficient to remember that $\mathfrak{J}^{\boldsymbol{p}}$ equals the product $\mathfrak{J}^{p_1}\cdots\mathfrak{J}^{p_n}$ *regardless of the order,* and that, if Θ_k is a polynomial of degree $<r_k$ in x_k, whose coefficients are continuous functions on I independent of x_k, then $\mathfrak{J}_j\Theta_k$ is again a polynomial in x_k, with coefficients of the same type and of degree $<r_k+1$ or $<r_k$, according to $j=k$ or $j\neq k$. ♦

PROOF OF LEMMA 2. It can be reduced to the following proposition: *if $\Theta\in\mathcal{P}_{\boldsymbol{r}}$ and, in addition, $\Theta\in C^{\boldsymbol{p}}$, then Θ can be represented in the form $\Theta=w_1+\cdots+w_n$, where $w_k\in\mathcal{P}_{kr_k}$, and, in addition, $w_k\in C^{\boldsymbol{p}}$.*

In fact, this implies that $\boldsymbol{D}^{\boldsymbol{p}}w_k\in\mathcal{P}_{k,\,r_k-p_k}$, hence $\boldsymbol{D}^{\boldsymbol{p}}\,\Theta\in\mathcal{P}_{\boldsymbol{r}-\boldsymbol{p}}$, by an argument similar to the one used for lemma 1.

To prove the preceding proposition, remember that I is the Cartesian product of n intervals $I_1,\dots, I_n$ in $\mathbb{R}$. Let $c_{k1},\dots, c_{kr_k}$ be r_k points chosen arbitrary in I_k for $k=1,\dots, n$. Then, to each function $f\in C$ and each $k=1,\dots, n$, corresponds one, and only one, function $f_k\in\mathcal{P}_{kr_k}$, such that:

7.2.2. $\qquad f_k(x)=f(x) \;\text{ for }\; x_k=c_{k\nu}\,,\;\; \nu=1,\dots, r_k\,.$

To see this it is sufficient to apply the *Lagrange interpolation formula*:

7.2.3. $$f_k(x)=\sum_{\nu=1}^{r_k-1} f_{k\nu}(x)\frac{\varphi_\nu(x_k)}{\varphi_\nu(c_{k\nu})}$$

where

7.2.4. $$f_{k\nu}(x)=f(x_1,\dots,x_{k-1},c_{k\nu},x_{k+1},\dots,x_n)$$

and

$$\varphi_\nu(x_k)=(x_k-c_{k,1})\cdots(x_k-c_{k,\nu-1})(x_k-c_{k,\nu+1})\cdots(x_k-c_{k,r_k}).$$

Thus $\varphi_\nu(c_{k\mu})=0$ for $\nu\neq\mu$, which along with 7.2.3. and 7.2.4. implies 7.2.2.

Let us denote by π_k the mapping $f\to f_k$ defined in this way for $k=1,\dots,n$. It is readily seen that π_k is a *projection of C onto $\mathcal{P}_{kr_k}$*, that is a linear mapping of C onto $\mathcal{P}_{kr_k}$, such that $\pi_k f=f$, for every $f\in\mathcal{P}_{kr_k}$.

Suppose now that Θ is a function $\in\mathcal{P}_{\boldsymbol{r}}$ having a continuous derivative $\boldsymbol{D}^{\boldsymbol{p}}\Theta$ on I in ordinary sense ($\boldsymbol{p}\leq\boldsymbol{r}$). Then Θ is of the form

$\Theta=\sum_1^n\Theta_k$ with $\Theta_k\in\mathcal{P}_{kr_k}$. Put $w_1=\pi_1\Theta$; since $\pi_1\Theta_1=\Theta_1$, we have

$$\Theta-w_1=(1-\pi_1)\Theta_2+\cdots+(1-\pi_1)\Theta_n$$

and since $\pi_1\left(\sum_\nu a_{k\nu}\,x_k^\nu\right)=\sum_\nu\pi_1(a_{k\nu}(\boldsymbol{x}))x_k^\nu$, it is easily seen that

$(1-\pi_1)\Theta_k\in\mathcal{P}_{kr_k}$ for $k=2,\dots,n$. Put in general

$$w_k=\pi_k(\Theta-w_1-\cdots-w_{k-1})\text{ for }k=2,\dots,n.$$

Then it is easily seen by repeated application of the same argument that $\Theta=\sum_{1}^{n} w_k$ with $w_k \in \mathcal{P}_{kr_k}$. Finally, observe that w_k is a polynomial in x_k, which is obtained by repeated application of Lagrange's formula 7.2.3.; therefore, its coefficients are linear combinations of functions, which derive from $\Theta(\boldsymbol{x})$ by replacing one or more variables $x_1, \ldots, x_n$ by constants. Since $\boldsymbol{D}^{\boldsymbol{p}}\Theta$ exists in ordinary sense and is continuous on I, it follows that the same property holds for $\boldsymbol{D}^{\boldsymbol{p}} w_k (k=1, \ldots, n)$. ♦

7.3. Vector operations and other fundamental concepts

Let f and g be any two distributions on an interval I in $\mathbb{R}^n$, $f=\boldsymbol{D}^{\boldsymbol{r}}F$ and $g=\boldsymbol{D}^{\boldsymbol{s}}G$, where $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{N}_0^n$ and $F, G \in C(I)$. As in the case of one variable, we shall put, by definition:

$$f+g=\boldsymbol{D}^{\boldsymbol{m}}(\mathfrak{I}^{\boldsymbol{m}-\boldsymbol{r}}F+\mathfrak{I}^{\boldsymbol{m}-\boldsymbol{s}}G)$$

where $\boldsymbol{m}$ is any system of n integers such that $\boldsymbol{m} \geq \boldsymbol{r}, \boldsymbol{s}$. On the other hand, we shall put, by definition:

$$\lambda f=\boldsymbol{D}^{\boldsymbol{r}}(\lambda F), \quad \forall \lambda \in \mathbb{C}.$$

It is easily seen, as in the case of one variable, that *the set $\mathscr{D}(I)$ of all distributions on I becomes a complex vector space with the preceding two definitions.* Moreover, it is obvious that the derivation operators $\boldsymbol{D}^{\boldsymbol{p}}$ are linear mappings of this space into itself.

Translation operators can also be defined as in the case of one variable. If $f=\boldsymbol{D}^{\boldsymbol{p}}F$, with $F \in C(I)$, and $\boldsymbol{h} \in \mathbb{R}^n$, then $\zeta_h f=\boldsymbol{D}^{\boldsymbol{p}}(\zeta_h F)$, where

$$(\zeta_h F)(\boldsymbol{x})=F(\boldsymbol{x}-\boldsymbol{h})=F(x_1-h_1, \ldots, x_n-h_n).$$

For every $\boldsymbol{r} \in I\!N_0^n$, we shall denote by $C_{\boldsymbol{r}}(I)$ – in short $C_{\boldsymbol{r}}$ – the set of all distributions f on I of the form $f = \boldsymbol{D}^{\boldsymbol{r}} F$, with $F \in C(I)$.

7.4. Restriction operators. Global distributions

The restriction operators, for distributions on intervals in $I\!R^n$, may be defined and denoted exactly as in the case of one variable and they have similar properties. In particular, if I is any interval in $I\!R^n$, we can identify every distribution f on I with its restriction to the interior of I, so that $\mathscr{D}(I) \subset \mathscr{D}(\mathring{I})$.

Besides, the collecting principle can be extended to distributions on intervals in $I\!R^n$ by an argument similar to the one used in the case of one variable, but it is a little more complicated; now the projections π_k considered in 7.2. should be used for each variable x_k separately in order to "collect" to each other the given distributions.

7.4.1. DEFINITION. If Ω is a (non-empty) open set in $I\!R^n$, a **global distribution** on Ω is any system $f = (f_I)$ that may be defined by assigning to each compact interval $I \subset \Omega$ one distribution f_I on I, in such a way that, if J is a compact subin-terval of I, then $f_J = \rho_J f_I$.

We shall denote by $\overline{\mathscr{D}}(\Omega)$ the set of all global distributions on Ω and, as in the case of one variable, we shall put by definition:

$$(f_I) + (g_I) = (f_I + g_I), \quad \lambda(f_I) = (\lambda f_I), \quad \boldsymbol{D}^{\boldsymbol{r}}(f_I) = (\boldsymbol{D}^{\boldsymbol{r}} f_I).$$

Then $\overline{\mathscr{D}}(\Omega)$ becomes a complex vector space and $\boldsymbol{D}^{\boldsymbol{r}}$ a linear mapping of $\overline{\mathscr{D}}(\Omega)$ into itself. In particular, every function $f \in C(\Omega)$ may be identified with the global distribution (f_I), where f_I is the restriction of f to each compact interval $I \subset \Omega$, so that $C(\Omega) \subset \overline{\mathscr{D}}(\Omega)$.

7.4.2. DEFINITION. A global distribution f on Ω is said to be of **finite rank**, if and only if there exists $r \in \mathbb{N}_0^n$ and $f \in C(\Omega)$ such that $f = D^r F$; otherwise, f is said to be of **infinite rank**.

In particular, if Ω is an interval, it is easily seen, by the collecting principle, that every distribution f on Ω can be identified with a global distribution of finite rank on Ω. So, in the general case, the global distributions of finite rank on Ω will be called **distributions** on Ω, and the set of all these objects will be denoted by $\mathscr{D}(\Omega)$. This set, which is obviously a vector subspace of $\overline{\mathscr{D}}(\Omega)$, could also be defined directly by a system of axioms, as in the case of intervals. (It should be observed that, *contrary to the case of* $\mathbb{R}$, *the components of an open set* Ω *in* $\mathbb{R}^n$ *are not, in general, intervals.*)

In the preceding definitions, we could consider, more generally, as the domain of a distribution, any set $\boldsymbol{\Delta}$ such that

$$\Omega \subset \boldsymbol{\Delta} \subset \overline{\Omega}$$

where Ω is any (non-empty) open set in $\mathbb{R}^n$. But as in the case of intervals, it is easily seen that every distribution on $\boldsymbol{\Delta}$ can be identified with a distribution on Ω, so that $\mathscr{D}(\boldsymbol{\Delta}) \subset \mathscr{D}(\Omega)$.

If $f, g \in \mathscr{D}(\Omega)$ and $\mathcal{O}$ is an open set contained in Ω, we write $f = g$ on $\mathcal{O}$, if and only if the restrictions of f and g to each interval $I \subset \mathcal{O}$ coincide. We say that f is **null** on $\mathcal{O}$, if and only if f equals the null function on $\mathcal{O}$. From the collecting principle follows that the *union of all open sets where a global distribution f is null is again a set where f is null.* That being so:

7.4.4. DEFINITION. If f is a global distribution on an open set Ω in $\mathbb{R}^n$ and if Ω_0 is the greatest open set where f is null, then the set $\Omega \setminus \Omega_0$ is called the **carrier** of f.

7.5. Locally summable functions as distributions

A function f is said to be **locally summable** on an open set Ω in IR^n if and only if f is summable on each compact interval $I \subset \Omega$. The integral of f over I may be denoted by $\int_I f$. If the extremities of I are $\boldsymbol{a} = (a_1, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, \ldots, b_n)$ with $\boldsymbol{a} \leq \boldsymbol{b}$, then we may also denote the integral by the notation

$$\int_{\boldsymbol{a}}^{\boldsymbol{b}} f(x)dx$$

or more explicity,

$$\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

If the condition $\boldsymbol{a} \leq \boldsymbol{b}$ is not satisfied, we shall put, by definition

$$\int_{\boldsymbol{a}}^{\boldsymbol{b}} f(x)dx = sgn \prod_k (b_k - a_k) \cdot \int_{\boldsymbol{\alpha}}^{\boldsymbol{\beta}} f(\boldsymbol{x})d\boldsymbol{x},$$

where $\boldsymbol{\alpha} = inf(\boldsymbol{a}, \boldsymbol{b})$ and $\boldsymbol{\beta} = sup(\boldsymbol{a}, \boldsymbol{b})$.

7.5.1. DEFINITION. If f is a locally summable function on an interval I in IR^n, any function F such that

$$F(\boldsymbol{x}) = \int_{\boldsymbol{c}}^{\boldsymbol{x}} f(\xi)d\xi, \quad \forall \boldsymbol{x} \in I$$

where $\boldsymbol{c}$ is an arbitrary fixed point of I, is said to be an **integral function** of f on I.

It can be proved that, *if F is an integral function of f on I, then $F \in C(I)$ and $f(\boldsymbol{x}) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} F(\boldsymbol{x})$ almost everywhere in ordinary sense. Moreover, if F_1 and F_2 are two integral functions of f, then*

$F_1 - F_2$ is a function Θ of the form $\Theta = \sum_1^n \Theta_k$, where Θ_k is a continuous function on I independent of x_k, for $k = 1, \ldots, n$.

As in the case of one variable, two locally summable functions f and g on I have the same integral function, if and only if $f(\boldsymbol{x}) = g(\boldsymbol{x})$ almost everywhere on I. In this case, f and g are said to be **equivalent** on I, and the vector space of the corresponding equivalent classes $[f]$ is defined as in the case of one variable.

Besides we shall denote by $\widetilde{f}$ (standardized f) the function defined by the formula:

$$\widetilde{f}(\boldsymbol{x}) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} \int_c^{\boldsymbol{x}} f(\xi)\, d\xi$$

only at the points $\boldsymbol{x}$ for which the written derivatives exist in ordinary sense, independently of the order, and leading to the same value.

From now on, when we speak of locally summable functions, it will be in general understood that they are standard functions, and we shall replace any equivalence class $[f]$ by the corresponding standard function $\widetilde{f}$. The vector space of all locally summable functions on I will be denoted by $\mathring{L}(I)$. That being so, it is easily proved, as in the case of one variable, that

7.5.4. *By assigning to each locally summable function f on I the distribution $f^* = D_1 \cdots D_n F$, where F is any integral function of f, there is defined a one-to-one linear mapping of $\mathring{L}(I)$ into $\mathscr{D}(I)$, such that:*

(i) if $f \in C(I)$, then $f = f^*$;

(ii) if f is absolutely continuous with respect to x_k on I_k, for almost every system of values of the remaining variables, then, to the derivative f'_{x_k} in functional sense, corresponds the derivative $D_k f^*$ in distributional sense.

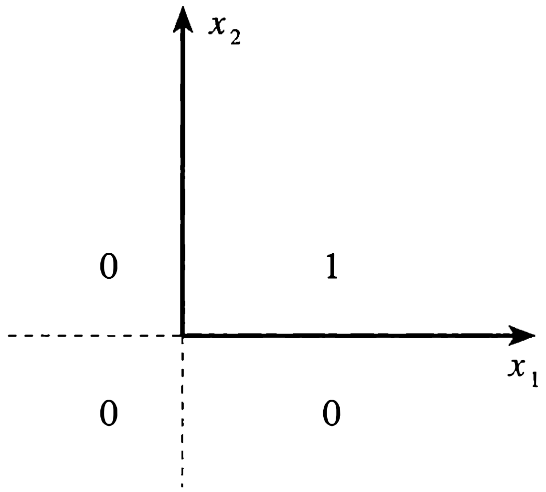
That being so, it is natural to identify each function $f \in \mathring{L}(I)$ with the corresponding distribution $f^* \in \mathscr{D}(I)$.

An important example of a (non-standard) locally summable function is the *Heaviside function* on $/R^n$ (which we shall denote by $H^{[n]}$) defined as follows:

$$H^{[n]}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } x_k \geq 0 \text{ for } \textit{all } k=1,\dots,n \\ 0 & \text{if } x_k < 0 \text{ for } \textit{some } k=1,\dots,n. \end{cases}$$

It is obvious that $H^{[n]}(\boldsymbol{x}) = H(x_1)\cdots H(x_n)$, $\forall \boldsymbol{x} = (x_1,\dots,x_n) \in /R^n$.

The *standardized Heaviside function* $\widetilde{H}^{[n]}$ is equal to $H^{[n]}$ at any continuity point of $H^{[n]}$ and is not defined at any discontinuity point of $H^{[n]}$. For example, if $n=2$, $\widetilde{H}^{[n]}$ is not defined *only* on the semi-axis $x_2=0$, $x_1 \geq 0$ and $x_1=0$, $x_2 \geq 0$.



We shall put in general $\overline{D} = D_1 \cdots D_n$.

The Dirac distribution on $/R^n$, which is denoted by $\delta^{[n]}$, can be defined as the pure mixed derivative of $\widetilde{H}^{[n]}$, that is, $\delta^{[n]} = \overline{D}\widetilde{H}^{[n]}$.

In general, given a locally summable function f, even if f is not a standard function, we may denote also by $\boldsymbol{D}^{\boldsymbol{r}}f$, where $\boldsymbol{r}$ is any system of integers, the distribution $\boldsymbol{D}^{\boldsymbol{r}}\widetilde{f}$. For example, we may write $\delta^{[n]} = \overline{D}H^{[n]}$.

Remarks about notation:

I) We shall often denote simply by H the Heaviside function on $/R^n$, whenever no mistake seems possible. In particular, no misun-

derstanding may arise, if the independent variables are written; for example, the meaning of expressions such as $H(x_1, \ldots, x_n)$, $H(x_3)$, etc., becomes quite clear. It should also be observed that, in practice, variables appear generally without subscripts, but this gives no trouble; for example, there will be no doubt about the meaning of expressions such as $H(x, y)$, $H(t)$, etc., or formulas such as $H(x, t) = H(x)H(t)$, $f_t(x, t) = f(x, t)H(t)$, etc., when x, y, t are real variables and f a function on $/R^2$.

Observe that the *Dirac distribution* at a point $\boldsymbol{a}$ of $/R^n$ is to be defined as in the case $n=1$:

$$\delta_{(\boldsymbol{a})} = \delta(\hat{\boldsymbol{x}} - \boldsymbol{a}) = D_1 \cdots D_n H(\hat{\boldsymbol{x}} - \boldsymbol{a}).$$

II) It must be observed that the preceding conventions about dummy variables cannot be extended, without some modifications, to distributions on $/R^n$, Now, we shall adopt the following conventions:

a) If f is a distribution on a subset of $/R^n$ and $\boldsymbol{x}, \boldsymbol{y}, \ldots$ are variables on $/R^n$, then $f(\hat{\boldsymbol{x}}) = f(\hat{\boldsymbol{y}}) = \cdots = f$.

b) If f is a distribution of one single variable, then $f(\hat{x}_1), \ldots, f(\hat{x}_n)$ denote *distinct* distributions on subsets of $/R^n$.

For example, the symbols $\hat{x}_1, \ldots, \hat{x}_n$ denote n distinct functions on $/R^n$ – the **coordinate functions**. In turn, $H(\hat{x}_1), \ldots, H(\hat{x}_n)$ denote n distinct locally summable functions on $/R^n$, whose product is $H^{[n]}$, and so forth.

7.6. Measures as distributions

Let Ω be an open set in $/R^n$. The concept of a measure μ on Ω can be defined exactly as we did for the case $n=1$. *For the sake of simplicity we shall restrict us here to the case where Ω is an open interval I.*

7.6.1. DEFINITION. Let μ be a measure on I and $\boldsymbol{c} = (c_1, \ldots, c_n)$ a point of I. If we put:

$$J_k = \begin{cases} [c_k, x_k] & \text{for } c_k \le x_k \\ ]x_k, c_k[& \text{for } x_k < c_k \end{cases} \qquad (k=1, 2, \ldots, n)$$

then the function F defined by

$F(\boldsymbol{x}) = sgn \prod_k (x_k - c_k) \cdot \mu(J_1 \times J_2 \times \cdots \times J_n)$ for all $\boldsymbol{x} \in I$ is called the **integral function** of μ from $\boldsymbol{c}$.

In order to see how to derive μ from F, it is convenient to consider, for every system $\boldsymbol{r} = (r_1, \ldots, r_n)$ of integers $r_k \ge 0$ and every vector $\boldsymbol{h} = (h_1, \ldots, h_n) \in /R^n$, the operator

7.6.2. $$\overline{\boldsymbol{\Delta}}_{\boldsymbol{h}} = \boldsymbol{\Delta}_{1h_1} \cdots \boldsymbol{\Delta}_{nh_n}$$

where $\boldsymbol{\Delta}_{ih_i}$ is the difference operator defined by

$$\boldsymbol{\Delta}_{ih_i} f(\boldsymbol{x}) = f(x_1, \ldots, x_i + h_i, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n).$$

For example, for $n=2$

$$\begin{aligned} \overline{\boldsymbol{\Delta}}_{\boldsymbol{h}} f(\boldsymbol{x}) &= \boldsymbol{\Delta}_{1h_1} \boldsymbol{\Delta}_{2h_2} f(x_1, x_2) \\ &= \boldsymbol{\Delta}_{1h_1} [f(x_1, x_2 + h_2) - f(x_1, x_2)] \\ &= f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - f(x_1, x_2 + h_2) + f(x_1, x_2). \end{aligned}$$

Now, it is easily seen that *if F is an integral function of the measure μ on I (in $/R^n$), then, for every pair of points $\boldsymbol{a}$, $\boldsymbol{b}$ of I, such that $\boldsymbol{a} < \boldsymbol{b}$, we have:*

7.6.3. $$\mu]\boldsymbol{a}, \boldsymbol{b}] = \overline{\boldsymbol{\Delta}}_{\boldsymbol{h}} F(\boldsymbol{a}), \quad \text{with } \boldsymbol{h} = \boldsymbol{b} - \boldsymbol{a}.$$

Moreover,

$$\mu[\boldsymbol{a}, \boldsymbol{b}] = \lim_{x \to a^-} \mu]\boldsymbol{x}, \boldsymbol{b}]$$

$$\mu]\boldsymbol{a}, \boldsymbol{b}[= \lim_{x \to b^-} \mu]\boldsymbol{a}, \boldsymbol{x}]$$

and analogously for the other types of bounded intervals J, such that $\bar{J}\subset I$ (in particular for degenerate intervals).

Thus the measure μ can be determined entirely from its integral function F.

It can also be seen that *F is continuous on the right at every point $\boldsymbol{a}$ of I, i.e., $F(\boldsymbol{a})=F(\boldsymbol{a}^{+})$.*

7.6.4. DEFINITION. By a **primitive** of a measure μ on I we shall understand any function F on I, continuous on the right, satisfying 7.6.3.

Obviously, every integral function of μ is a primitive of μ, *but not conversely.*

Let us put, for every interval $J=]\boldsymbol{a}, \boldsymbol{b}]$ with $\boldsymbol{a}, \boldsymbol{b}\in I$ $(\boldsymbol{a}<\boldsymbol{b})$:

$$\bar{\boldsymbol{\Delta}} F(J)=\bar{\boldsymbol{\Delta}}_{\boldsymbol{h}} F(\boldsymbol{a}) \quad \text{with} \quad \boldsymbol{h}=\boldsymbol{b}-\boldsymbol{a}.$$

Then, a function F on I is said to be of **bounded variation**, if and only if to each interval $J=]\boldsymbol{a}, \boldsymbol{b}]$ with $\boldsymbol{a}, \boldsymbol{b}\in I$, $(\boldsymbol{a}<\boldsymbol{b})$, there corresponds a number $M(J)$, such that, for every partition of each interval $J_k=]a_k, b_k]$ in a finite number of left open intervals $J_{k1}, \ldots, J_{kp_k}$ $(k=1, 2, \ldots, n)$, we have

$$\sum_{\nu_1=1}^{p_1} \cdots \sum_{\nu_n=1}^{p_n} |\bar{\boldsymbol{\Delta}} F(J_{1\nu_1}\times \cdots \times J_{n\nu_n})| \leq M(J).$$

That being so, it is easily seen that:

7.6.5. *A function F on I is a primitive of some measure μ on I, if and only if F is of bounded variation on I and continuous on the right at every point $\boldsymbol{a}$ of I. Moreover, two such functions F_1 and F_2 are primitives of the same measure if and only if F_1-F_2 if of the form $\Theta_1+\cdots+\Theta_n$ where Θ_k is a function of bounded variation on I, independent of x_k, $k=1, 2, \ldots, n$.*

In the set $\mathfrak{M}(I)$ of all measures on I, there is defined the structure of a complex vector space, as in the case $n=1$. On the other hand, every function $f \in \mathring{L}(I)$ can be identified with the measure μ_f, defined by $\mu_f(J) = \int_J f$.

Finally, observe that every function F of bounded variation on I is locally summable on I and uniquely determined by the corresponding standard function. Thus, applying 7.6.5. and taking into account axiom 4 in 7.2., we arrive at the following conclusion:

7.6.6. *By assigning to each measure μ on I the distribution $f^* = \overline{D}F$, where F is any primitive of μ, there is defined a one-to-one linear mapping of $\mathfrak{M}(I)$ into $\mathscr{D}(I)$ such that, if μ is a locally summable function on I, then $\mu = f^*$.*

That being so, it is natural to put in the general case $\mu = f^*$, so that $\mathfrak{M}(I)$ becomes a vector subspace of $\mathscr{D}(I)$. This result holds, if we consider instead of an open interval I, any open (non-empty) set Ω in $/R^n$.

7.7. Concepts of multiplication; tensor products, concrete examples.

Let I be any interval in $/R^n$. The product of a continuous function f on I with a measure μ on I can be defined as in the case of one variable. For example, we have, for every continuous function on $/R^n$ and every point $\boldsymbol{a}$ of $/R^n$:

$$f(\hat{\boldsymbol{x}})\delta(\hat{\boldsymbol{x}} - \boldsymbol{a}) = f(\boldsymbol{a})\delta(\hat{\boldsymbol{x}} - \boldsymbol{a}).$$

If $\boldsymbol{r}$ is a system of n integers $r_k \geq 0$, $\mathfrak{M}_r(I)$ – or simply $\mathfrak{M}_r$ – denotes the set of all distributions $f = D^r F$, where $F \in \mathfrak{M}(I)$. Obviously, $\mathfrak{M}_r$ is a vector subspace of $\mathscr{D}$ and C^r a subalgebra of C.

Now we can define the product fg of a function $f \in C^r$ and a distribution $g \in \mathfrak{M}_r$, so as to satisfy the two conditions:

i) If $g \in \mathfrak{M}$, then fg is the product of the function f by the measure g in previous sense.

ii) If $D_k f \in C^r$ and $g \in \mathfrak{M}_r$, then

$$D_k(fg) = D_k f \cdot g + f \cdot D_k g \quad \text{for } k = 1, \dots, n.$$

Then if $f \in C^r$ and $g = D^r G$ with $G \in \mathfrak{M}$, the product fg is uniquely defined by the following formula (cf. chapter IV, 4.1.1.):

$$fg = \sum_{k_1=0}^{r_1} \cdots \sum_{k_n=0}^{r_n} (-1)^{\|r-k\|} \binom{r_1}{k_1} \cdots \binom{r_n}{k_n} D^k(f^{(r-k)} G)$$

where $\|\boldsymbol{r} - \boldsymbol{k}\| = (r_1 - k_1) + \cdots + (r_n - k_n)$.

For example, for $f \in C^r(/R^n)$ and $\boldsymbol{a} \in /R^n$:

$$f\delta^{(r)}(\hat{\boldsymbol{x}} - \boldsymbol{a}) = \sum_{k_1=0}^{r_1} \cdots \sum_{k_n=0}^{r_n} (-1)^{\|r-k\|} \binom{r_1}{k_1} \cdots \binom{r_n}{k_n} f^{(r-k)}(\boldsymbol{a}) \delta^{(k)}(\hat{\boldsymbol{x}} - \boldsymbol{a}).$$

As in the case of one variable, it is easily proved that $\mathfrak{M}_r$ *becomes a module on the algebra* C^r.

Besides, we have several different possibilities of extending this concept of product, as in the case of one variable, and even new possibilities. For example:

Let p be, not a system of integers, but an integer ≥ 0. Then we shall denote by $C^p(I)$ or simply C^p the set of all functions f having continuous derivatives $f^{(\boldsymbol{k})}$ on I, in ordinary sense, *of all orders* $\leq p$, that is, such that $\|\boldsymbol{k}\| = |k_1| + \cdots + |k_n| \leq p$. On the other hand, we shall denote by $\mathfrak{M}_p(I)$ or simply by $\mathfrak{M}_p$ the set of all distributions g on I, which can be expressed as *sums* $\sum_1^m g_\nu$ *of a finite (arbitrary) number of distributions* g_ν, *belonging to* $\mathfrak{M}_r$ *with* $\boldsymbol{r} = (r_1, \dots, r_n)$ *and* $|r_1| + \cdots + |r_n| \leq p$.

Now, if we require the distributive law to be maintained, it can be shown that, if $f \in C^p$ and $g \in \mathfrak{M}_p$, the product fg is uniquely defined by:

$$fg = \sum_1^m f g_\nu$$

where fg_ν, is given by the previous general formula. *Thus* $\mathfrak{M}_p$ *becomes a module on the algebra* C^p.

Observe that, in particular, *the space* $\mathscr{D}$ *of all distributions is a module over* C^∞, *the space of infinitely differentiable functions (on I).*

Another new possibility arises from the concept of "tensor product". Let I and J be two intervals respectively in $/R^m$ and $/R^n$ spaces ($m, n \geq 0$). If f and g are two continuous functions on I and J respectively, then the expression $f(\boldsymbol{x})g(\boldsymbol{y}) = f(x_1, \dots, x_m)g(y_1, \dots, y_n)$ defines obviously a continuous function on the interval $I \times J \subset /R^{m+n}$.

Let now f and g be two distributions on I and J respectively, $f = D^r F$ and $g = D^s G$, with $F \in C(I)$, $G \in C(J)$. Then it is readily seen that the expression

7.7.1. $$D^{r \oplus s}[F(\hat{\boldsymbol{x}})G(\hat{\boldsymbol{y}})]$$

denotes a distribution on $I \times J$, uniquely determined by f and g. That being so

7.7.2. DEFINITION. The distribution 7.7.1. will be called the **tensor product** (or **direct product**) of f by g and denoted by $f \otimes g$ or by $f(\hat{\boldsymbol{x}})g(\hat{\boldsymbol{y}})$.

It is readily seen that this tensor product is bilinear and associative, but, of course, not commutative. Furthermore, it can obviously be extended to any finite system of distributions.

For example: $H^{[m]} \otimes H^{[n]} = H^{[m+n]}$, $\delta^{[m]} \otimes \delta^{[n]} = \delta^{[m+n]}$, $\delta^{[3]} = \delta \otimes \delta \otimes \delta$, etc. In a less rigorous, but more convenient notation, we may write also, for example (cf. 1.5.): $\delta(x, y, z) = \delta(x)\delta(y)\delta(z)$, $D_t \delta(x, t) = \delta(x)\delta'(t)$, etc.

Many concrete situations lead to considering tensor products of distributions, as we have already seen in 1.5. For example, let $f(x, y)$ be a locally summable functions on $/R^2$, then $f(x, y)\delta(z)$ will be a distribution whose carrier is contained in the x, y-plane. This may be the case of an electric charge distribution of *surface density* $f(x, y)$ on this plane.

Analogously $f(x, y)\delta'(t)$ may represent an electric doublet on the x, y-plane, and so forth.

Similar situations may arise relating to curves, surfaces or, more generally, manifolds in $/R^n$-spaces.

7.8. Change of variables. Concrete examples; δ-distributions of a hypersurface

Let α be a distribution on an open set Ω in $/R^n$ and $\boldsymbol{r}$ a system of n integers $r_k \geq 0$. Then the symbol αD^r will denote the operator defined by the formula

$$(\alpha D^r)f = \alpha(D^r f)$$

for all distributions f on Ω such that α is multipliable by $D^r f$. In particular, if $\alpha \in C^\infty(\Omega)$, the domain of αD^r will be $\mathscr{D}(\Omega)$.

By a **linear differential operator of finite order** we shall understand any operator A which can be represented as the sum of a finite number of operators of the form αD^r; then, the order of A is the greatest value of $\|\boldsymbol{r}\|$ occurring actually in all terms of the sum.

That being so, let us consider any two integers $m, n \geq 1$ and a mapping $\boldsymbol{h}$ of an open set $\Omega^* \subset /R^n$ into an open set $\Omega \subset /R^m$. Then $\boldsymbol{h}$ is defined by a system of m real-valued functions $h_1, \ldots, h_m$ on Ω^*:

$$x_i = h_i(t_1, \ldots, t_n) \quad \text{for} \quad i = 1, \ldots, m,$$

which may be written $\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{t})$.

Let f be now a *complex-valued* function on Ω and suppose $f \in C^1(\Omega)$, $h_i \in C^1(\Omega^*)$ for $i=1, \ldots, m$. Then

$$D_{t_j} f(\boldsymbol{h}(\boldsymbol{t})) = \sum_{i=1}^{m} f'_{x_i}(\boldsymbol{h}(\boldsymbol{t})) \frac{\partial h_i}{\partial t_j}$$

or else, putting $h_{ij} = \dfrac{\partial h_i}{\partial t_j}$

7.8.1. $$D_{t_j}(f \circ \boldsymbol{h}) = \sum_{i=1}^{m} h_{ij}(D_{x_i} f \circ \boldsymbol{h}), \quad j=1, \ldots, n.$$

a) *Let us consider at first the case when $m=n$, and suppose that the Jacobean of $\boldsymbol{h}$ with respect to $\boldsymbol{t}$ (i.e. the determinant $|h_{ij}|$) is different from zero on Ω^*:*

$$J\begin{pmatrix} h_1 & \cdots & h_n \\ t_1 & \cdots & t_n \end{pmatrix} \neq 0 \text{ for all } \boldsymbol{t} \in \Omega^*.$$

Then 7.8.1. can be solved with respect to the functions $D_{x_i} f \circ \boldsymbol{h}$:

$$(D_{x_i} f) \circ \boldsymbol{h} = \sum_{j=1}^{n} \alpha_{ij} D_{t_j}(f \circ \boldsymbol{h}), \quad i=1, \ldots, n$$

where $\alpha_{ij} \in C^1(\Omega^*)$ for $i, j=1, \ldots, n$ and $[\alpha_{ij}]$ is the inverse of the matrix $[h_{ij}]$. This result may be expressed by writing:

7.8.2. $$D_{x_i} = \sum_{j=1}^{n} \alpha_{ij} D_{t_j}, \quad i=1, \ldots, n.$$

From now on the change of variables for distributions of n variables may be defined essentially as in the case $n=1$.

7.8.3. DEFINITION. If $f = D^{\boldsymbol{r}} F$, with $F \in C(\Omega)$ and $\boldsymbol{r} = (r_1, \ldots, r_n)$ and if $\alpha_{ij} \in C^{\boldsymbol{r}}(\Omega^*)$ for all i, j then

$$f \circ \boldsymbol{h} = D_{\boldsymbol{x}}^{r}(F \circ \boldsymbol{h}) \in C_r(\Omega^*)$$

with

$$D_{\boldsymbol{x}}^{r} = \left(\sum_{j=1}^{n} \alpha_{1j} D_{t_j}\right)^{r_1} \cdots \left(\sum_{j=1}^{n} \alpha_{nj} D_{t_j}\right)^{r_n}.$$

Uniqueness and other properties of the composition $f \circ \boldsymbol{h}$ may be proved as in the case of one variable.

b) *Consider now the case $m<n$ and suppose that the characteristic of the matrix $[h_{ij}]$ is equal to m for all $\boldsymbol{t} \in \Omega^*$.* Then, for every $\boldsymbol{t}^0 \in \Omega^*$ we could solve 7.8.1. with respect to $(D_{\boldsymbol{x}} f) \circ \boldsymbol{h}$ *in some neighborhood of $\boldsymbol{t}^0$*. But, in order to obtain a global solution (on Ω), it is convenient to "normalize" the system 7.8.1., i.e. to consider the "normal" system deduced from the first:

7.8.4. $$\sum_{j=1}^{n} h_{\nu j} D_{t_j}(f \circ \boldsymbol{h}) = \sum_{i=1}^{m} \widetilde{h}_{\nu i}(D_{x_i} f \circ \boldsymbol{h}), \quad \nu = 1, \dots, m$$

where $$\widetilde{h}_{\nu i} = \sum_{j=1}^{n} h_{\nu j} h_{ij}, \quad \nu = 1, \dots, m.$$

Then the determinant $|\widetilde{h}_{\nu i}|$ is different from zero for all $\boldsymbol{t} \in \Omega^*$ and the system 7.8.4. can be solved with respect to $D_{x_i} f \circ \boldsymbol{h}$, for $i = 1, \dots, m$

$$(D_{x_i} f) \circ \boldsymbol{h} = \sum_{j=1}^{n} \alpha_{ij} D_{t_j}(f \circ \boldsymbol{h})$$

or in short

$$D_{x_i} = \sum_{j=1}^{n} \alpha_{ij} D_{t_j}, \quad i = 1, \dots, m,$$

where the coefficients α_{ij} are C^1 functions on Ω^* *uniquely determined by the given functions h_i.*

From now on the change of variables for distributions may be defined as in the previous case.

Suppose in particular $m=1$. Let us consider the change of variables defined by a function $u=\boldsymbol{h}(x_1,\dots,x_n)$ mapping an open set Ω^* in $I\!R^n$ into Ω in $I\!R$ (now the new variables are $x_1,\dots,x_n$ instead of $t_1,\dots,t_n$). Assume $\boldsymbol{h}\in C^1(\Omega^*)$ and

$$(h'_{x_1})^2+\cdots+(h'_{x_n})^2\neq 0 \quad \text{on} \quad \Omega^*.$$

Then, every function $f(u)$ of the real variable u, such that $f\in C^1(\Omega)$, is transformed into a function $f(\boldsymbol{h}(x_1,\dots,x_n))=(f\circ\boldsymbol{h})(\boldsymbol{x})$ such that

7.8.5. $$D_{x_j}(f\circ\boldsymbol{h})=h'_{x_j}(f'\circ\boldsymbol{h})\,,\quad j=1,\dots,n,$$

where $f'=D_uf$. From this follows:

$$\sum_{j=1}^{n} h'_{x_j}D_{x_j}(f\circ\boldsymbol{h})=(f'\circ\boldsymbol{h})\sum_{j=1}^{n}(h'_{x_j})^2;$$

hence putting

$$\alpha_j=\frac{h'_{x_j}}{(h'_{x_1})^2+\cdots+(h'_{x_n})^2}\,,\quad j=1,\dots,n,$$

we obtain

$$f'\circ\boldsymbol{h}=\sum_{1}^{n}\alpha_jD_{x_j}(f\circ\boldsymbol{h})\,,$$

that is

7.8.6. $$D_u=\alpha_1D_{x_1}+\cdots+\alpha_nD_{x_n}\,.$$

Observe now that, for each $u\in\boldsymbol{h}(\Omega^*)$, the equation $\boldsymbol{h}(\boldsymbol{x})=u$ represents a hypersurface Σ_u in $I\!R^{n+1}$, and that $\alpha_1,\dots,\alpha_n$ are the components of the vector

$$\frac{1}{|grad\, \boldsymbol{h}|}\boldsymbol{n} \text{ with } \boldsymbol{n}=\frac{1}{|grad\, \boldsymbol{h}|}grad\, \boldsymbol{h}$$

which is normal to Σ_u at each point $\boldsymbol{x}$. So 7.8.6. can be written simply

7.8.7.
$$D_u=\frac{1}{|grad\, \boldsymbol{h}|}\frac{\partial}{\partial \boldsymbol{n}}$$

where $\partial/\partial\boldsymbol{n}$ denotes *normal derivation* with respect to the hypersurface Σ_u, i.e. the differentiation along the unitary vector $\boldsymbol{n}$ (more precisely, along *the vector field* $\boldsymbol{n}$).

c) Suppose finally $m>n$. Then $\boldsymbol{h}$ maps Ω^* onto a manifold V of dimension $\leq n$ contained in Ω and, given a distribution f on Ω, the composition $f\circ\boldsymbol{h}$ exists, if and only if there exists the restriction f_v of f to V, as well as $f_v\circ\boldsymbol{h}$; then $f\circ\boldsymbol{h}=f_v\circ\boldsymbol{h}$. We shall speak later about this new concept of restriction.

Examples: Consider the distribution δ on $/R$ and a C^1 mapping f of an open set Ω^* in $/R^n$ into $/R$ such that $|grad\, f|\neq 0$ on Ω^*. Then it is easily seen that $\delta\circ f$ exists and is given by

$$H'(f(\boldsymbol{x}))=\frac{1}{|grad\, f|}\frac{\partial}{\partial\boldsymbol{n}}H(f(\boldsymbol{x}))$$

where $\partial/\partial\boldsymbol{n}$ denotes the derivation along the vector field $\boldsymbol{n}=|grad\, f|^{-1}$ $grad\, f$. Observe that $H(f(\boldsymbol{x}))$ equals 0 or 1 according as $f(\boldsymbol{x})<0$ or $f(\boldsymbol{x})>0$; so, denoting by Σ the hypersurface $f(\boldsymbol{x})=0$, we could say that $H(f(\boldsymbol{x}))$ equals 0 on the left of Σ and 1 on the right of Σ (Σ is supposed to be oriented by means of the normal $\boldsymbol{n}$). On the other hand, since $\sum_1^n(f'_{x_i})^2\neq 0$ on Ω^*, there exists, for every $\boldsymbol{x}^0\in\Sigma$, a bounded open interval I in $/R^n$, containing $\boldsymbol{x}^0\in\Sigma$, such that the equation $f(\boldsymbol{x})=0$ can be solved in I with respect to one of the variables $x_1,\ldots, x_n$, say x_1:

$$x_1 = \varphi(x_2, \ldots, x_n)$$

and
$$\varphi'_{x_j} = -f'_{x_j}/f'_{x_1} \quad \text{for } j \neq 1.$$

Now we have (supposing, as we can, $f'_{x_1} > 0$ on I)

$$\frac{\partial}{\partial \boldsymbol{n}} = \frac{|grad\, f|}{f'_{x_1}} D_{x_1} = \sqrt{1 + \sum_{k=2}^{n} (\varphi'_{x_k})^2}\, D_{x_1}.$$

Let $J = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be an interval such that $\bar{J} \subset I$ and put

$$\psi(\boldsymbol{x}) = \int_{a_2}^{x_2} \cdots \int_{a_n}^{x_n} \sqrt{1 + \sum_{k=2}^{n} [\varphi'_{x_k}(\xi_2, \ldots, \xi_n)]^2}\, H(f(x_1, \xi_2, \ldots, \xi_n))\, d\xi_2 \cdots d\xi_n.$$

It is readily seen that, in the neighborhood I of $\boldsymbol{x}^0$, we have

$$\frac{\partial}{\partial \boldsymbol{n}} H(f(\boldsymbol{x})) = \bar{D}\psi(\boldsymbol{x})$$

where $\bar{D} = D_{x_1} D_{x_2} \cdots D_{x_n}$. Applying this argument to each $\boldsymbol{x}^0 \in \Sigma$, it follows that

7.8.8. $\dfrac{\partial}{\partial \boldsymbol{n}} H(f(\boldsymbol{x}))$ *is the measure on Ω^* assigning to each bounded interval J, such that $\bar{J} \subset \Omega^*$, the "area" of $\Sigma \cap J$.*

We shall denote by δ_Σ this measure (*δ-distribution of the oriented hypersurface Σ*) and by H_Σ the function $H(f(\boldsymbol{x}))$ (*Heaviside function of Σ*). Hence we have

7.8.9.
$$\delta_\Sigma = \frac{\partial}{\partial \boldsymbol{n}} H_\Sigma \quad \text{and} \quad grad\, H_\Sigma = \delta_\Sigma \cdot \boldsymbol{n}.$$

It can also be shown that, if $f \in C^2$, then

$$\Delta H_\Sigma = \delta'_\Sigma = \left(\frac{\partial}{\partial \boldsymbol{n}}\right)^2 H_\Sigma \quad (\text{where } \Delta = \text{div grad}).$$

These considerations, except the preceding result, can be extended to the case where Σ is any oriented piecewise smooth manifold in $\mathbb{R}^n$. It should be observed that, by considerations of such type, *all the classical vector and tensor analysis can be rebuilt for distributions with proofs which are in general more natural and more simple than the classical ones.*

As an example of the δ-distributions of a hypersurface, consider the distribution $\delta(|\boldsymbol{x}|-\rho)$, where $\boldsymbol{x}\in\mathbb{R}^3$, $|\boldsymbol{x}|=\sqrt{x_1^2+x_2^2+x_3^2}$ and $\rho>0$.

It is easily seen that this distribution is the δ of the sphere $|\boldsymbol{x}|=\rho$. A concrete example may be a distribution of electric charge with surface density $1/4\pi$ on the sphere, supposed to be a conductor in electrostatic equilibrium. Then the charge distribution $\delta(|\boldsymbol{x}|-\rho)$ creates the electric field $\boldsymbol{u}$ defined by $\boldsymbol{u}=\rho^2\dfrac{\boldsymbol{x}}{|\boldsymbol{x}|^3}H(|\boldsymbol{x}|-\rho)$ which derives from the electric potential

$$V=\begin{cases}\dfrac{1}{4\pi}\dfrac{1}{\rho} & \text{for } |\boldsymbol{x}|\leq\rho\\[2ex] \dfrac{1}{4\pi}\dfrac{1}{|\boldsymbol{x}|} & \text{for } |\boldsymbol{x}|>\rho\,.\end{cases}$$

It is easily seen then that $\Delta V=-4\pi\delta(|\boldsymbol{x}|-\rho)$.

Consider now the distribution $\delta(\boldsymbol{x}^2-v^2t^2)$, with $\boldsymbol{x}\in\mathbb{R}^3$, t real $\neq 0$ and v constant >0. We have now

$$\delta(\boldsymbol{x}^2-v^2t^2)=\frac{1}{4|\boldsymbol{x}|}[\delta(|\boldsymbol{x}|-vt)+\delta(|\boldsymbol{x}|+vt)],$$

where $\delta(|\boldsymbol{x}|-vt)=\dfrac{\partial}{\partial\boldsymbol{n}}H(|\boldsymbol{x}|-vt)$ for each $t\neq 0$. This distribution is defined only for $t\neq 0$ and its carrier is just the *wave cone* $\boldsymbol{x}^2-v^2t^2=0$,

minus the origin. In turn, the carrier of $\delta(\boldsymbol{x}^2 - v^2 t^2 + \rho)$, with $\rho > 0$, is an hyperboloid of two leaves in the space $/R_x^3 \times /R_t$, etc.

7.9. Topological vector space of distributions of several variables

Let I be a compact interval in $/R^n$, $\boldsymbol{p} = (p, p, \ldots, p) \in /N_0^n$, and consider the vector space $C(I)$ provided with the usual norm $\|f\| = max|f(\boldsymbol{x})|$. If we denote by $C_{\boldsymbol{p}}(I)$ the vector space of all distributions f of the form $f = \bar{D}^{\boldsymbol{p}} F = D_1^p \cdots D_n^p F$, where $F \in C(I)$, it is natural to consider $C_{\boldsymbol{p}}(I)$ provided with the semi-norm corresponding to the ball $\bar{D}^{\boldsymbol{p}} U$, where $U = \{f : f \in C(I), \|f\| = 1\}$. Now the kernel of $\bar{D}^{\boldsymbol{p}}$ is the set $G_{\boldsymbol{p}} = \sum_{k=1}^{n} G_{k,\boldsymbol{p}}$ of all pseudo-polynomials of degree $< (p, \ldots, p)$, and it can be proved as in the case of $n=1$, by means of Lagrange's interpolation formula, that $G_{\boldsymbol{p}}$ is *closed* in $C(I)$. Hence $C_{\boldsymbol{p}}(I)$ is a normed space. On the other hand:

$$\mathscr{D}(I) = \bigcup_{p=0}^{\infty} C_{\boldsymbol{p}}(I),$$

and it is easily shown, as in the case $n=1$, that the injection $C_{\boldsymbol{p}} \to C_{\boldsymbol{p}+1}$ is compact for all $\boldsymbol{p}$. Hence $\mathscr{D}(I)$, *considered as the inductive limit of the normed spaces* $C_{\boldsymbol{p}}(I)$, *is a* (*LN**) *space.*

In particular, the convergence of sequences can be defined directly as follows:

7.9.1. *A sequence of distributions* $f_k \in \mathscr{D}(I)$ *converges to* $g \in \mathscr{D}(I)$, *if and only if there exists an integer* $p \geq 0$, *a sequence of functions* $F_k \in C(I)$, *and a function* $G \in C(I)$ *such that:* $f_k = \bar{D}^{\boldsymbol{p}} F_k$ *for all* k, $g = \bar{D}^{\boldsymbol{p}} G$ *and* $\|F_k - G\| \to 0$.

It is still readily seen that convergence in the mean on I implies convergence in distributional sense.

Let now Ω be an open set in $I\!R^n$. The vector space $\overline{\mathscr{D}}(\Omega)$ is provided with the topology of the projective limit of the (LN^*) spaces $\mathscr{D}(I)$, where I is any compact interval in Ω, by means of the linear mappings ρ_I. This means that a filter σ converges to 0 in $\overline{\mathscr{D}}(\Omega)$, if and only if $\rho_I\sigma$ converges to 0 in $\mathscr{D}(I)$ for every compact interval $I\subset\Omega$.

In any of these distributional spaces, the following property is obviously true:

$$f_k \to g \Rightarrow D^r f_k \to D^r g\,,\quad \forall \boldsymbol{r}\in I\!N_0^n.$$

Examples: 1 – Put

$$\delta_k^{[n]}(\boldsymbol{x}) = \delta_k(x_1)\cdots\delta_k(x_n)$$

where $\delta_k = H_k'$ (cf. 6.8.1.). Then it can be seen, as in the case $n=1$, that

$$\lim_{k\to\infty} D^r\delta_k^{[n]} = D^r\delta^{[n]}\quad \forall \boldsymbol{r}\in I\!N_0^n.$$

2 – Considering primitives of the measures $\delta(\boldsymbol{x})$ and $\delta\left(|\boldsymbol{x}|-\dfrac{1}{k}\right)$, for $k=1,2,\ldots$ it is easily seen that

$$\lim_{k\to\infty} \delta\left(|\boldsymbol{x}|-\frac{1}{k}\right) = \delta(\boldsymbol{x}).$$

3 – Let $U_k(\boldsymbol{x})$ be equal to $\dfrac{1}{k}$ $\left(\text{resp. } \dfrac{1}{|\boldsymbol{x}|}\right)$ for $|\boldsymbol{x}|\le \dfrac{1}{k}$ $\left(\text{resp. } |\boldsymbol{x}|>\dfrac{1}{k}\right)$ $(k=1,2,\ldots,\ \boldsymbol{x}\in I\!R^3)$. Then

$$\int_I \left[U_k(\boldsymbol{x}) - \frac{1}{|\boldsymbol{x}|}\right] d\boldsymbol{x} \to 0$$

for every bounded interval I, so that $U_k \to \dfrac{1}{|\boldsymbol{x}|}$ in distributional sense. Hence

$$\Delta U_k \to \Delta\left(\frac{1}{|\boldsymbol{x}|}\right)$$

and, therefore (cf. 2):

$$\Delta \frac{1}{|\boldsymbol{x}|} = -4\pi\delta(\boldsymbol{x}) \quad \text{on } /R^3.$$

CHAPTER VIII

PARTIAL INTEGRALS AND MULTIPLE INTEGRALS. CONVOLUTION.

8.1. Partial limits for distributions of two variables.

Let I and J be two intervals in $/R$, and suppose that J is unbounded on the right. Given two functions $f(x, y)$ and $g(x)$ respectively on $I \times J$ and I, $f(x, y)$ is said to **converge uniformly** on I to $g(x)$ as $y \to +\infty$, if and only if for every $\varepsilon > 0$, there exists a $\eta \in J$ (independent of x), such that: $|f(x, y) - g(x)| < \varepsilon$ for all $y > \eta$ and $x \in I$.

On the other hand, if α is any real, we write $f(x, y) \in o(y^\alpha)$ **uniformly** on I as $y \to +\infty$, if and only if (iff) $\dfrac{f(x, y)}{y^\alpha} \to 0$ uniformly on I as $y \to +\infty$. Put for every $f \in C(I \times J)$:

$$\mathfrak{I}_x f(x, y) \equiv \int_{x_0}^{x} f(\xi, y) d\xi \,, \quad \mathfrak{I}_y f(x, y) \equiv \int_{y_0}^{y} f(x, \eta) d\eta \,,$$

where x_0, (respectively y_0) is a fixed point, arbitrarily chosen in I (respectively J). The following lemma is easily proved (cf. 6.1.3.):

8.1.1. LEMMA. *If $\alpha > -1$ and $f(x, y) \in o(y^\alpha)$ uniformly on I as* $y \to +\infty$, then

$$\mathfrak{I}_x f \in o(y^\alpha) \quad \text{and} \quad \mathfrak{I}_y f \in o(y^{\alpha+1})$$

uniformly on I as $y \to +\infty$.

This lemma leads to the following:

8.1.2. DEFINITION. If $f \in \mathscr{D}(I \times J)$ and $\alpha > -1$, we write $f(x, y) \in o(y^\alpha)$ on I as $y \to +\infty$, iff there exist $m, n \in IN_0$ and $F \in C(I \times J)$ such that:

(1) $f(x, y) = D_x^m D_y^n F(x, y)$;

(2) $F(x, y) \in o(y^{\alpha+n})$ uniformly on each compact interval $I^* \subseteq I$ as $y \to +\infty$.(7)

Applying this definition and the lemma, the following properties are easily shown:

8.1.3. *If $f(x, y) \in o(y^\alpha)$ and $g(x, y) \in o(y^\alpha)$ on I as $y \to +\infty$, then, for all $\lambda, \mu \in \mathbb{C}$:*

$$\lambda f + \mu g \in o(y^\alpha) \text{ on } I \text{ as } y \to +\infty.$$

8.1.4. *If $f(x, y) \in o(y^\alpha)$ on I as $y \to +\infty$, then $D_x f(x, y) \in o(y^\alpha)$ on I as $y \to +\infty$.*

8.1.5. *If $f(x, y) \in o(y^\alpha)$ on I as $y \to +\infty$ and $\varphi(x)$ is multipliable by $f(x, y)$,* then *$\varphi(x) f(x, y) \in o(y^\alpha)$ on I as $y \to +\infty$.*

Consider now $g \in \mathscr{D}(I)$ and $f \in \mathscr{D}(I \times J)$; then:

8.1.6. DEFINITION. We say that $f(x, y)$ **converges** on I to $g(x)$ as $y \to +\infty$, if and only if $f(x, y) - g(x) \in o(1)$ on I as $y \to +\infty$. The distribution $f(x, y)$ is said to be **convergent** on I as $y \to +\infty$, iff there exists a distribution g on I satisfying the preceding condition.

---

(7) A more general condition could be required instead of (2), but this definition is quite sufficient for applications.

The uniqueness property as well as the linearity property of convergence are in this case immediate consequences of 8.1.3. Then we can write:

$$g(x) = \lim_{y \to +\infty} f(x, y) \quad \text{or} \quad g(x) = f(x, +\infty) \quad \text{on} \quad I,$$

to express that $f(x, y) \to g(x)$ on I as $y \to +\infty$.

On the other hand, the following important property, *which does not hold in classical analysis,* is an immediate consequence of 8.1.4.:

8.1.7. DIFFERENTIATION PROPERTY. *If* $f(x, y) \to g(x)$ *on* I *as* $y \to +\infty$, *then* $D_x f(x, y) \to D_x g(x)$ *on* I *as* $y \to +\infty$, that is:

$$D_x \lim_{y \to +\infty} f(x, y) = \lim_{y \to +\infty} D_x f(x, y) \textit{ on } I.$$

It turn, from 8.1.5., follows

8.1.7'. MULTIPLICATION PROPERTY. *If* $f(x, y) \to g(x)$ *on* I *as* $y \to +\infty$ *and* $\varphi(x)$ *is multipliable by* $f(x, y)$, *then*:

$$\lim_{y \to +\infty} [\varphi(x) f(x, y)] = \varphi(x) \lim_{y \to +\infty} f(x, y) \textit{ on } I.$$

Moreover, applying 8.1.7. and the linearity property, it is easily shown:

8.1.8. SUBSTITUTION PROPERTY. *If* $f(x, y) \to g(x)$ *on* I *as* $y \to +\infty$, *and if* h *is a mapping of an interval* I^* *into* I, *such that* $f(h(t), y)$ *exists* (cf. 6.8.), *then* $g(h(t))$ *exists too and* $f(h(t), y) \to$ $\to g(h(t))$ *on* I^* *as* $y \to +\infty$.

This substitution rule concerns the *parameter* x. Substitution rules concerning the *converging variable* y can be easily found as generalization of the criteria given in 6.6.

The "O" symbol is extended to distributions $f(x, y)$ on $I \times J$, with respect to y, in the following way:

8.1.9. DEFINITION. If $\alpha > -1$, we write $f(x, y) \in O(y^\alpha)$ on I as $y \to +\infty$, if there exist $m, n \in /N_0$ and $F \in C(I \times J)$ such that:

(i) $f(x, y) = D_x^m D_y^n F(x, y)$;

(ii) for every compact interval $I^* \subseteq I$, there exists a number M such that

$$\frac{F(x, y)}{(1+|y|)^{\alpha+n}} \leq M \text{ on } I^* \times J.^{(8)}$$

More generally, if $\varphi \in C^\infty(J)$, we write $f(x, y) \in O(\varphi(y))$ on I as $y \to +\infty$, if and only if there exists a real y_0 and a distribution $f_0(x, y) \in O(1)$ on I as $y \to +\infty$ such that $f(x, y) = \varphi(y) f_0(x, y)$, for $y > y_0$ and $x \in I$.

Besides the linearity property, it is easily shown:

8.1.10. DIFFERENTIATION PROPERTY. *If α is any real and $f(x, y) \in O(y^\alpha)$ on I as $y \to +\infty$, then $D_x f(x, y) \in O(y^\alpha)$ and $D_y f(x, y) \in O(y^{\alpha-1})$ on I as $y \to +\infty$.*

Obviously all preceding considerations extend to the case when J is an interval unbounded on the left, and $y \to -\infty$.

8.2. Partial integrals for distributions of two variables.

Let I and J be any two intervals in $/R$, $f(x, y)$ a distribution on $I \times J$. A distribution $\varphi(x, y)$ such that $D_y \varphi(x, y) = f(x, y)$ will be called a (**partial**) **primitive** of $f(x, y)$ with respect to y. On the other hand, a distribution $u(x, y)$ on $I \times J$ is said to be **independent** of y, if

---

(8) The choice of $(1+|y|)^{\alpha+n}$ instead of $y^{\alpha+n}$ is only to make the quotient continuous on $I^* \times J$.

and only if it reduces to a distribution g of the variable x only, i.e., iff it is of the form $u(x, y) = D_x^m G(x)$ with $m \in I\!N_0$, $G \in C(I)$.

8.2.1. LEMMA. *A distribution u on $I \times J$ is independent of y, iff $D_y u = 0$.*

PROOF. It is readily seen that, if u is independent of y, then $D_y u = 0$. Suppose now conversely that $D_y u = 0$ and assume $u = D_x^m D_y^n U$ with $m, n \in I\!N_0$ and $U \in C(I)$. Then $D_y u = D_x^m D_y^{n+1} U = 0$ and therefore (cf. 7.2. axiom 4) U must be of the form

$$U(x, y) = \sum_0^{m-1} x^i a_i(y) + \sum_0^{n} y^j b_j(x), \text{ with } a_i \in C(J) \text{ and } b_j \in C(I).$$

Hence $u(x, y) = D_x^m D_y^n U(x, y) = n! D_x^m b_n(x)$. ♦

That being so, it is easily proved, as in the case of one variable:

8.2.2. THEOREM. *Every distribution f on $I \times J$ has infinitely many primitives with respect to y and two such primitives differ by a distribution independent of y.*

We are now able to define, in a natural way, the concept of **partial** (or **parametric**) **integral** of a distribution $f(x, y)$. It will be sufficient to consider integrals on $I\!R$. Let I be any interval in $I\!R$ and $f \in \mathscr{D}(I \times I\!R)$; then:

8.2.3. DEFINITION. The integral $\int_{I\!R} f(x, y)dy$ is said to be **convergent** on I, if and only if there exists a primitive φ of f with respect to y wich is convergent on I as $y \to +\infty$ and as $y \to -\infty$. Then, we write $\int_{I\!R} f(x, y)dy = \varphi(x, +\infty) - \varphi(x, -\infty)$ on I.

From 8.2.2., follows at once the uniqueness of the partial integral. From the properties of partial limits we can deduce the *linearity property for partial integrals*, as well as the following properties:

8.2.4. DIFFERENTIATION PROPERTY. *If* $\int_{IR} f(x, y)\,dy$ *is convergent on I, so is* $\int_{IR} f_x'(x, y)\,dy$ *and*

$$D_x \int_{IR} f(x, y)\,dy = \int_{IR} D_x f(x, y)\,dy \text{ on } I.$$

8.2.5. SUBSTITUTION PROPERTY. *If* $\int_{IR} f(x, y)\,dy = g(x)$ *on* I *and if h is any continuous mapping of an interval* I^* *into I such that* $f(h(t), y)$ *exists, then*

$$\int_{IR} f(h(t), y)\,dy = g(h(t)) \text{ on } I^*.$$

As for substitutions concerning the integration variable y, the criteria established in 6.6. can be easily extended to partial integrals. In particular, we have, for all $h \in IR$:

8.2.6. $$\int_{IR} f(x, y+h)\,dy = \int_{IR} f(x, y)\,dy.$$

8.2.7. $$\int_{IR} f(x, hy)\,dy = \frac{1}{|h|} \int_{IR} f(x, y)\,dy.$$

Criterium 6.3.6. can also be extended to partial integrals:

8.2.8. THEOREM. *If for any compact interval* $I^* \subseteq I$, *there exists a compact interval K such that* $f(x, y) = 0$ *on* $I^* \times (IR - K)$, *then the integral* $\int_{IR} f(x, y)\,dy$ *is convergent on I.*

PROOF. Suppose $f = D_x^m D_y^n F$, with $F \in C(I \times IR)$. The hypothesis implies that, in a set $\{x \in I^*, y < -y_0\}$, $F(x, y)$ reduces to a pseudo polynomial P of degree $<(m, n)$, which we can assume to be zero,

otherwise we could subtract P from F (remember that P is uniquely defined by the value of $F(x, y)$ for m values of x in I^* and n values of y in $\mathbb{R}$). Then there exists a primitive of f with respect to y, say φ, which is zero for $x \in I^*$, $y < -y_0$ and reduces to a function ψ independent of y for $x \in I^*$, $y > y_0$. Now, it is easily seen that $\varphi \to \psi$ on I as $y \to +\infty$ and $\varphi \to 0$ on I as $y \to -\infty$, so that

$$\int_{\mathbb{R}} f(x, y)\, dy = \psi(x). \quad \blacklozenge$$

Finally, the following extensions of 6.5.1. and 6.5.2. are easily proved:

8.2.9. THEOREM. *If $\int_{\mathbb{R}} f(x, y)\, dy$ is convergent on I, then $f \in O(y^{-1})$ on I as $y \to \infty$. On the other hand, if there exists $\alpha < -1$ such that $f \in O(y^{\alpha})$ on I as $y \to \infty$, then $\int_{\mathbb{R}} f(x, y)\, dy$ is convergent on I.*

Remarks. 1 – If $f(x, y)$ is a function, then for the convergence of $\int_{\mathbb{R}} f(x, y)\, dy$ on *I in distributional sense*, it is not sufficient (nor necessary) that the integral be convergent *for each* $x \in I$. Obviously, a sufficient condition is that the integral be uniformly convergent on each compact subinterval contained in I. More generally, it can be proved that, if f *is summable on each set* $I^* \times \mathbb{R}$, *where* I^* *is a compact interval contained in I, then* $\int_{\mathbb{R}} f(x, y)\, dy$ *is convergent on I in distributional sense.*

2 – The differentiation property can be associated with the linearity property in a more general property. Let $p(D)$ be a *derivation polynomial,* that is an operator of the form $p(D) = \sum_{1}^{n} a_k D^k$, with $a_1, \ldots, a_n \in \mathbb{C}$. Then we have

$$p(D_x)\int_{/R} f(x, y)\,dy = \int_{/R} p(D_x) f(x, y)\,dy \text{ on } I,$$

whenever the first integral is convergent on I.

Example. The preceding remarks offer a simple justification of formula in 6.3.5. - 2. Observe that:

8.2.10. $$\int_{/R} \frac{e^{ixy}}{1+y^2}\,dy = \pi e^{-|x|} \textit{ for each } x \in /R.$$

This can be easily found by the *method of residues*. Besides, as x, y are real variables, we have $|e^{ixy}| = 1$ and

$$\left|\frac{e^{ixy}}{1+y^2}\right| = \frac{1}{1+y^2} \quad \text{for all } x, y \in /R.$$

Thus the integral 8.2.10. is dominated, for all $x \in /R$, by the integral

$\int_{/R} (1+y^2)^{-1} dy$, which is obviously convergent. Hence, according to

the Weierstrass test, the first integral is *uniformly convergent on /R and therefore convergent on /R in distributional sense*. Consequently,

$$(1-D_x^2)\int_{/R} \frac{e^{ixy}}{1+y^2}\,dy = \int_{/R} \frac{(1-D_x^2)e^{ixy}}{1+y^2}\,dy = \int_{/R} e^{ixy}\,dy \quad \text{on} \quad /R.$$

On the other hand,

$$D_x^2 e^{-|x|} = -D_x(e^{-|x|} sgn\ x) = e^{-|x|} - 2e^{-|x|}\,\delta(x),$$

so that

$$(1-D_x^2)e^{-|x|} = 2\delta(x).$$

Hence, from 8.2.10. follows:

8.2.11. $$\int_{/R} e^{ixy}dy = 2\pi\delta(x) \text{ on } /R.$$

8.3. Multiple integrals (on $/R^n$).

Let f be a distribution on $/R^n$ and λ any complex number. We say that $f(\boldsymbol{x})$ **converges** to λ as $\boldsymbol{x} \to +\infty_n$ if and only if there exist $\boldsymbol{r} \in /N_0^n$ and $F \in C(/R^n)$ such that $f = D^{\boldsymbol{r}}F$ and

$$\frac{F(\boldsymbol{x})}{x_1^{r_1} \cdots x_n^{r_n}} \to \frac{\lambda}{r_1! \cdots r_n!} \quad \text{as } x_1 \to +\infty, \ldots, x_n \to +\infty.$$

Then, we write $\lambda = \lim\limits_{x \to +\infty_n} f(\boldsymbol{x})$ or $\lambda = f(+\infty_n)$.

The uniqueness of the limit, as well as the linearity property can be proved by an argument similar to the one used in the case $n=1$. The concept of convergence as $\boldsymbol{x} \to -\infty$ is analogously defined.

On the other hand, every distribution φ such that $\bar{D}\varphi = f$ (where $\bar{D} = D_1 \cdots D_n$) will be called a **pure mixed primitive** of f. It is easily seen that:

8.3.1. THEOREM. *Every $f \in \mathscr{D}'(/R^n)$ has infinitely many pure mixed primitives and two such primitives differ necessarily by a distribution of the form $\sum_1^n u_j$ where u_j is a distribution independent of x_j (that is, of the form $D^{\boldsymbol{r}}\boldsymbol{u}$ where $\boldsymbol{u}$ is a continuous function on $/R^n$ independent of x_j).*

That being so, we shall write by definition.

8.3.2. $$\int_{\boldsymbol{x}}^{\boldsymbol{x}'} f(\boldsymbol{\xi})\,d\boldsymbol{\xi} = \bar{\boldsymbol{\Delta}}_{\boldsymbol{x}'-\boldsymbol{x}}\, \varphi(\boldsymbol{x}),$$

where φ is any pure mixed primitive of f and $\overline{\boldsymbol{\Delta}}_{\boldsymbol{h}}$ is the mixed difference operator $\Delta_{1h_1}\cdots\Delta_{nh_n}$.

From 8.3.1. follows that formula 8.3.2. defines actually a distribution $\phi(\boldsymbol{x},\boldsymbol{x}')$ on $/R^{2n}$ independent of the choice of the pure mixed primitive φ. To see this it is sufficient to observe that $\Delta_{jh_j}u_j=0$ for every distribution u_j independent of x_j.

8.3.3. DEFINITION. A distribution f is said to be **integrable** on $/R^n$, iff $\int_{\boldsymbol{x}}^{\boldsymbol{x}'} f(\boldsymbol{\xi})\,d\boldsymbol{\xi}$ is convergent as $(\boldsymbol{x},\boldsymbol{x}')\to(-\infty_n,+\infty_n)$. Then we write:

8.3.4.
$$\int_{/R^n} f(\boldsymbol{x})\,d\boldsymbol{x} = \lim_{\substack{\boldsymbol{x}\to-\infty_n\\ \boldsymbol{x}'\to+\infty_n}} \int_{\boldsymbol{x}}^{\boldsymbol{x}'} f(\boldsymbol{\xi})\,d\boldsymbol{\xi}.$$

For example, if $n=2$

$$\int_{/R^2} f(x_1,x_2)\,dx_1\,dx_2=\varphi(+\infty,+\infty)-\varphi(+\infty,-\infty)-\varphi(-\infty,+\infty)+\varphi(-\infty,-\infty)$$

where φ is a primitive of f with respect to $\boldsymbol{x}$.

The integral of f on $/R^n$ can also be denoted by $\int_{/R^n} f$ or simply by $\int f$. Uniqueness and linearity properties are immediate consequences of the corresponding properties for limits. In order to obtain further criteria it is convenient to introduce a suitable definition of bounded distributions.

8.3.5. DEFINITION. A distribution f is said to be **bounded on** $/R^n$, if and only if there exists $\boldsymbol{r}\in /N_0^n$ and $F\in C(/R^n)$ such that:

(i) $f=D^{\boldsymbol{r}}F$;

(ii) for every regular matrix A of order n, the function $x_1^{-r_1}\cdots x_n^{-r_n}\,F(A\boldsymbol{x})$ is bounded on $/R^n$.

The linearity property of boundedness is easily proved.

8.3.6. DEFINITION. Given $f\in\mathscr{D}(/R^n)$ and $\varphi\in C^\infty(/R^n)$, we write $f\in O(\varphi)$ as $|\boldsymbol{x}|\to\infty$ or simply $f\in O(\varphi)$, if and only if there exists a distribution f_0 bounded on $/R^n$ and a real $\varepsilon>0$, such that $f=\varphi f_0$, for $|\boldsymbol{x}|>\varepsilon$.

That being so, the following generalization of 6.5.1. is easily obtained:

8.3.7. THEOREM. *If there exists $\alpha<-n$ such that $f\in O(|\boldsymbol{x}|^\alpha)$, then f is integrable on $/R^n$.*

On the other hand:

8.3.8. THEOREM. *Suppose $f\in O(|\boldsymbol{x}|^\alpha)$ with $\alpha<-n$ and let $\boldsymbol{h}$ be a C^∞ one-to-one mapping of $/R^n$ onto itself such that*

(i) the Jacobian matrix $[D_ih_j]$ of $\boldsymbol{h}$ is regular on $/R^n$ and converges to a regular matrix as $|\boldsymbol{t}|\to\infty$,

(ii) $D^{\boldsymbol{r}}D_ih_j\in o(\boldsymbol{t}^{\boldsymbol{r}})$, for all $\boldsymbol{r}\in/N_0^n$, $i,j=1,\dots,n$ (9).

Then the classical substitution rule applies:

$$\int_{/R^n} f(\boldsymbol{x})\,d\boldsymbol{x}=\int_{/R^n} f(\boldsymbol{h}(\boldsymbol{t}))\left|J\begin{pmatrix}\boldsymbol{h}\\ \boldsymbol{t}\end{pmatrix}\right|d\boldsymbol{t}.$$

We shall outline the proof only in the case when $\boldsymbol{h}$ is a non-degenerate *affine mapping,* that is a mapping of the form $\boldsymbol{h}(\boldsymbol{t})=\boldsymbol{c}+\boldsymbol{M}(\boldsymbol{t})$, where $\boldsymbol{c}$ is any vector in $/R^n$ and M is a regular matrix of order n. This case may be taken as a model for the general case since $\boldsymbol{h}$ behaves *asymptotically* just as an affine mapping according to *(i).*

---

(9) As far as functions are concerned is understood that the stated conditions are to be taken in ordinary sense.

Put $\varphi(\boldsymbol{x})=(1+x_1^2+\cdots+x_2^n)^{1/2}$ and suppose $f\in O(|\boldsymbol{x}|^\alpha)$, with $\alpha<-n$. Then, it is readily seen that $f\in O(\varphi^\alpha)$, i.e. there exists $\boldsymbol{r}\in /N_0^n$ and $F\in C$, such that $f=\varphi^\alpha D^{\boldsymbol{r}}F$, with $x_1^{-r_1}\cdots x_n^{-r_n}\,F(A\boldsymbol{x})$ bounded on $/R^n$ for every regular matrix A of order n. In such conditions it is easily found:

$$\int f(\boldsymbol{x})\,d\boldsymbol{x}=(-1)^{\|r\|}\int \varphi^{(r)}(\boldsymbol{x})\,F(\boldsymbol{x})\,d\boldsymbol{x}\,,\ \text{where}\ \|\boldsymbol{r}\|=r_1+\cdots+r_n.$$

Now:

$$\int f(\boldsymbol{x})\,d\boldsymbol{x}=(-1)^{\|r\|}\int \varphi^{(r)}(\boldsymbol{x})F(\boldsymbol{x})d\boldsymbol{x}=\int (-1)^{\|r\|}\varphi^{(r)}(\boldsymbol{h}(\boldsymbol{t}))F(\boldsymbol{h}(\boldsymbol{t}))\,|det\,M|\,d\boldsymbol{t}$$

and it can be seen, without difficulty, that the last integral is just equal to

$$(-1)^{\|r\|}\int f(\boldsymbol{h}(\boldsymbol{t}))\,|det\,M|\,d\boldsymbol{t}.$$

8.4. Partial and multiple integrals.

Let us consider a distribution $f(\boldsymbol{x},\boldsymbol{y})$ on $/R^{m+n}$, with $\boldsymbol{x}\in /R^m$ and $\boldsymbol{y}\in /R^n$ (m, $n=1, 2, \ldots$). The concept of partial integral $\int_{/R^n} f(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}$ can be easily defined as a generalization of preceding concepts of partial and multiple integral, with similar properties. But there is a new property:

8.4.1. THEOREM. *If $f(\boldsymbol{x},\boldsymbol{y})$ is integrable on $/R^{m+n}$ and in addition the integral $\int_{/R^n} f(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}$ is convergent on $/R^m$, then*

$$\int_{/R^{m+n}} f(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{x}\,d\boldsymbol{y}=\int_{/R^m}\left(\int_{/R^n} f(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{y}\right)d\boldsymbol{x}\,.$$

This is a consequence of a property for limits that we can state as follows:

8.4.2. THEOREM. *If $f(\boldsymbol{x}, \boldsymbol{y})$ is convergent as $(\boldsymbol{x}, \boldsymbol{y}) \to (+\infty_m, +\infty_n)$ and if in addition $f(\boldsymbol{x}, \boldsymbol{y})$ is convergent on $\mathbb{R}^m$ as $\boldsymbol{y} \to +\infty_n$, then*

$$\lim_{\substack{x \to +\infty_m \\ y \to +\infty_n}} f(\boldsymbol{x}, \boldsymbol{y}) = \lim_{x \to +\infty_m} \left(\lim_{y \to +\infty_n} f(\boldsymbol{x}, \boldsymbol{y}) \right).$$

PROOF. It is sufficient to prove this rule in the case $m = n = 1$. Suppose that the hypothesis holds. Then there exist four integers r, s, t, u, two functions $F_1, F_2 \in C(\mathbb{R}^2)$, a function $G \in C(\mathbb{R})$ and a number λ, such that $f = D_x^r D_y^s F_1 = D_x^t D_y^u F_2$ and

(i) $\dfrac{F_1(x, y)}{x^r y^s} \to \dfrac{\lambda}{r!\, s!}$ as $(x, y) \to (+\infty, +\infty)$;

(ii) $\dfrac{F_2(x, y)}{y^u} \to \dfrac{G(x)}{u!}$, uniformly on each compact set in $\mathbb{R}^m$ as $y \to +\infty$;

We can assume that $t = r$, $u = s$. Take $\varepsilon > 0$, then according to (i) there exist a, $b > 0$ such that:

8.4.3. $$\left| \frac{F_1(x, y)}{x^r y^s} - \frac{\lambda}{r!\, s!} \right| < \varepsilon \text{ for } x > a,\ y > b.$$

Take now r additional points $x_j > a$, s additional points $y_k > b$ and consider two pseudo-polynomials $\mathcal{P}_1(x, y)$, $\mathcal{P}_2(x, y)$ of degree (m, n) such that $F_1 - \mathcal{P}_1$ and $F_2 - \mathcal{P}_2$ vanish on the lines $x = x_j$, $y = y_k$. Then if we put $F_0 = F_1 - \mathcal{P}_1$, we have $f = D_x^r D_y^s F_0$, $F_0 = F_2 - \mathcal{P}_2$ and it is easily seen that (i), (ii) are again satisfied with F_0 in the place of F_1 and F_2 ($t = r$, $u = s$), since the coefficients of the pseudo-polynomials are obtained as linear combinations of the values of $F_1(x, y)$ and $F_2(x, y)$ on the lines

$x=x_j$, $y=y_k$. Hence from 8.4.3. follows, with F_0 in the place of F_1, and taking the limit as $y \to +\infty$:

$$\left|\frac{G(x)}{x^r}-\frac{\lambda}{r!}\right| \leq s!\varepsilon \text{ for } x>a.$$

The number ε being arbitrary, this implies that $\dfrac{G(x)}{x^r} \to \dfrac{\lambda}{r!}$ as $x \to +\infty$, which means that $\lambda = \lim\limits_{x\to+\infty} \lim\limits_{y\to+\infty} f(x, y)$. ♦

More generally:

8.4.4. *If $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, with $\boldsymbol{x} \in /R^m$, $\boldsymbol{y} \in /R^n$, $\boldsymbol{z} \in /R^p$ is convergent on $/R^{m+n}$ as $\boldsymbol{z} \to +\infty_p$ and if $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is convergent on $/R^m$ as $(\boldsymbol{y}, \boldsymbol{z}) \to (+\infty_n, +\infty_p)$, then*

$$\lim_{\substack{y\to+\infty_n \\ z\to+\infty_p}} f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \lim_{y\to+\infty_n}\left(\lim_{z\to+\infty_p} f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\right)$$

8.5. Convolution of two distributions on */R*.

Consider two distributions $f=D^mF$ and $g=D^nG$, where F, $G \in C(/R)$. Then we have:

$$f(x-t)=D_x^mF(x-t)=(-1)^mD_t^mF(x-t)$$

so that, for every $k=0, 1, \ldots$

$$D_t^kf(x-t)=(-1)^kD_x^kf(x-t).$$

This suggests to write by definition

$$f(x-t)\,g(t)=f(x-t)\,D_t^nG(t)=\sum_{k=0}^{n}\binom{n}{k}D_t^{n-k}(G(t)D_x^kf(x-t))$$

with

$$G(t)D_x^k f(x-t) = D_x^{m+k}(F(x-t)\,G(t)),$$

that is

8.5.1. $$f(x-t)\,g(t) = \sum_{k=0}^{n} \binom{n}{k} D_x^{m+k} D_t^{n-k}(F(x-t)\,G(t)).$$

It is easily seen that the "product" $f(x-t)\,g(t)$ does not depend on the representation of the distributions f and g. We can prove it as we have done for the product of a C^n function with a C_n distribution in 4.1. The analogy between these two situations comes from the following proposition, which can be proved without difficulty, but which is not essential for the following subject: *The mapping* $t \to f(x-t)$ *of* $\mathbb{R}$ *into the space* $\mathscr{D}(\mathbb{R})$ *is infinitely differentiable.*

Consider now the expression $f(x-t)\,g(t-y)$. We have two possible interpretations:

$$f(x-t)\,g(t-y) = \sum_{k=0}^{n} \binom{n}{k} D_x^{m+k} D_t^{n-k}(F(x-t)\,G(t-y))$$

8.5.2.

$$f(x-t)\,g(t-y) = \sum_{k=0}^{m} \binom{m}{k} (-1)^k D_t^{m-k} D_y^{n+k}(F(x-t)\,G(t-y)).$$

Remembering that the functions F and G can be approached by two sequences $\{F_n\}$ and $\{G_n\}$ of C^∞ functions converging uniformly on each compact interval, it is readily seen that:

8.5.3. *The right members of the formulas* 8.5.2. *represent the same distribution.*

A direct proof of this proposition does not seem to be easy.

8.5.4. DEFINITION. If the integral $\int_{-\infty}^{+\infty} f(x-t)g(t)dt$ is convergent on $/R$, the distribution

$$h(x) = \int_{/R} f(x-t)g(t)dt$$

is called the **convolution** of f and g and denoted by $f*g$.

From this definition, taking into account the linearity property of the partial integral, as well as 8.5.1., follows immediately that the *convolution is bilinear,* that is, we have:

8.5.5. $$(\alpha f_1+\beta f_2)*g=\alpha(f_1*g)+\beta(f_2*g), \ \forall \alpha,\beta\in\mathbb{C},$$

whenever f_1*g and f_2*g exist, and analogously for the right side. Moreover

8.5.6. COMMUTATIVE LAW: *If $f*g$ exists, $g*f$ exists too, and $f*g=g*f$.*

PROOF. Suppose that $f*g$ exists and put $h=f*g$, that is

$h(x)=\int_{/R} f(x-t)g(t)dt$. Then for each $y\in /R$, we have:

$$h(x-y)=\int_{/R} f(x-y-t)g(t)dt$$

and it is obvious that the last integral is still convergent with respect to (x, y) on $/R^2$. On the other hand, *for each* $y\in /R$, we may perform on this integral the substitution $t=u-y$, which gives:

$$h(x-y)=\int_{/R} f(x-u)g(u-y)du\,.$$

Now, taking 8.5.3. into account, it can be seen that the last integral is also convergent with respect to y *for each* $x\in /R$. In particular, for $x=0$, we have:

$$h(-y) = \int_{/R} f(-u)\, g(u-y)\, du\,.$$

Hence by the substitutions $y = -x$, $u = -t$:

$$h(x) = \int_{/R} g(x-t)\, f(t)\, dt,$$

that is, $h = g * f$. ♦

In the general case the convolution is not associative. But the following criterion can be used in several cases:

8.5.7. *If* $\int_{/R^2} f(x-y)\, g(y-t)\, h(t)\, dy\, dt$, *where* f, g, $h \in \mathscr{D}$, *is convergent on* $/R$, *then*

$$(f * g) * h = f * (g * h) = \int_{/R^2} f(x-y)\, g(y-t)\, h(t)\, dy\, dt.$$

This is an immediate consequence of 8.4.4.

In turn, from the differentiation and substitution properties for partial integrals and from 8.5.6., follows immediately, taking definition 8.5.4. into account:

8.5.8. DIFFERENTIATION PROPERTY. *If* $f * g$ *exists, then* $D(f * g)$ exists too, and

$$D(f * g) = (Df) * g = f * (Dg).$$

8.5.9. TRANSLATION PROPERTY. *If* $f * g$ *exists, then for every* $h \in /R$

$$\tau_h(f * g) = (\tau_h f) * g = f * (\tau_h g).$$

On the other hand:

8.5.10. *If $f*g$ and $f*(\hat{x}g)$ exists, then $(\hat{x}f)*g$ exists too and*

$$\hat{x}(f*g)=(\hat{x}f)*g+f*(\hat{x}g).$$

PROOF. It is sufficient to observe that $(\hat{x}f)*g$ is given by

$$\int_{/R}(x-t)f(x-t)g(t)dt=x\int_{/R}f(x-t)g(t)dt-\int_{/R}f(x-t)tg(t)dt. \ \blacklozenge$$

This important property shows that multiplication by x, with respect to convolution, behaves like a derivation operator.

Finally, we can analogously prove that

8.5.11. *If $f*g$ exists, then*

$$e^{ax}(f*g)=(e^{ax}f)*(e^{ax}g) \quad \forall a\in\mathbb{C}.$$

8.6. Convolution of distributions whose carrier is bounded on the left and (or) on the right.

We shall denote by $\mathscr{D}_*(/R)$ or simply $\mathscr{D}_*$ the vector space of all distributions on $/R$ with bounded carrier.

8.6.1. THEOREM. *The convolution $f*g$ exists whenever $f\in\mathscr{D}_*$ and $g\in\mathscr{D}$. Besides,*

(i) $f*(g*h)=(f*g)*h$, whenever $f,\ g\in\mathscr{D}_*,\ h\in\mathscr{D}$;

(ii) $\delta*f=f$, for every $f\in\mathscr{D}$.

PROOF. a) Suppose $f\in\mathscr{D}_*,\ g\in\mathscr{D}$. Then there exists a bounded interval J such that $g(x-t)f(t)$ vanishes for $(x,\ t)\notin /R\times J$. Hence

$\int_{/R}g(x-t)f(t)\,dt$ is convergent on $/R$ and gives $f*g$.

b) Suppose $f,\ g\in\mathscr{D}_*,\ h\in\mathscr{D}$. Then by an argument similar to the preceding it is shown that the integral

$$\int_{/R^2} f(x-y)\,g(y-t)\,h(t)\,dy\,dt$$

is convergent on $/R$, and this according to 8.5.7. implies *(i)*.

c) Consider $f=D^nF$, where $F\in C(/R)$, and put $F_1=FH$, $F_2=F-F_1$. Now:

$$H*F_1=\int_0^\infty H(x-t)F_1(t)\,dt=\int_0^x F_1(t)\,dt.$$

Hence $\delta*D^nF_1=D^{n+1}(H*F_1)=D^nF_1$. It is seen analogously that $\delta*D^nF_2=D^nF_2$, so that $\delta*f=f$. ♦

This theorem along with 8.5.5. can be expressed by saying:

8.6.2. *The space $\mathscr{D}_*$ is an algebra under convolution and $\mathscr{D}$ is a module over that algebra, having δ as unit element.*

Property *(ii)* in 8.6.1. can be expressed explicity by the important formula

$$f(x)=\int_{/R}\delta(x-t)\,f(t)\,dt \quad \text{(DIRAC'S FORMULA).}$$

We shall denote by $\mathscr{D}_+$ (respectively $\mathscr{D}_-$) the vector space of all distributions vanishing on the left (resp. on the right) of 0 and by $\widetilde{\mathscr{D}}_+$ (resp. $\widetilde{\mathscr{D}}_-$) the space of all distributions whose carrier is bounded on the left (resp. on the right) of 0.

8.6.3. THEOREM. *The space $\widetilde{\mathscr{D}}_+$ (resp. $\widetilde{\mathscr{D}}_-$) is an algebra under convolution and $\mathscr{D}_+$ (resp. $\mathscr{D}_-$) is a subalgebra of $\widetilde{\mathscr{D}}_+$ (resp. $\widetilde{\mathscr{D}}_-$).*

In fact, if $f, g\in\widetilde{\mathscr{D}}_+$, there exists a real c such that f and g vanish for $x<c$. Then $f(x-t)g(t)$ vanish for $t<c$ and $t>x-c$. Hence

$\int_{/R} f(x-t)\,g(t)\,dt$ is convergent on $/R$ and vanishes for $x<2c$. The

remaining parts of the theorem are easily proved. ♦

8.7. Convolution and order of growth, tempered distributions and rapidly decreasing distributions (on /*R*).

Several criteria can be found, connecting convolution with order of growth of distributions. One of these criteria is the following:

8.7.1. THEOREM. *Let α and β be two real numbers satisfying one of the following conditions*

(i) $\alpha+\beta<-1$ *and* $\alpha\geq 0$;

(ii) $\alpha+\beta<-3$ *and* $\beta\leq\alpha<0$.

On the other hand, let f and g be two continuous functions on /R such that $f\in O(x^{\alpha})$ and $g\in O(x^{\beta})$[10]. *Then* $f*g$ *exists and* $f*g\in O(x^{\alpha})$.

PROOF. a) Suppose $\alpha+\beta<-1$ with $\alpha\geq 0$. Then as $f\in O(x^{\alpha})$, there exists a number M such that $|f(x)|\leq M(1+|x|)^{\alpha}$ for all $x\in /R$. Hence

$$|f(x-t)|\leq M(1+|x-t|)^{\alpha}\leq M(1+|x|)^{\alpha}(1+|t|)^{\alpha} \quad \forall x,\, t\in /R,$$

since $\alpha\geq 0$.

So the integral

$\int_{/R} f(x-t)g(t)dt$ is dominated by $M(1+|x|)^{\alpha}\int_{/R}(1+|t|)^{\alpha}|g(t)|dt.$

Since $g\in O(x^{\beta})$ and $\alpha+\beta<-1$, the last integral exists. Hence the first integral is uniformly convergent on each compact interval in /*R* and its absolute value is $\leq MK(1+|x|)^{\alpha}$ where $K=\int_{/R}(1+|t|)^{\alpha}|g(t)|dt$. Consequentely, $f*g\in O(x^{\alpha})$.

b) Suppose now $\alpha+\beta<-3$, with $\beta\leq\alpha<0$, and consider the integer n such that $0\leq\alpha+n<1$. Then it is easily seen that $\alpha+\beta+n<-1$ so that $x^k f*x^{n-k}g$ exists and is $O(x^{\alpha+k})$ for $k=0, 1,\dots, n$ according to the previous conclusion. Hence (cf. 8.5.10):

---

(10) It is understood: "in ordinary sense as $x\to\infty$".

$$x^n(f*g)=\sum_{k=0}^{n}\binom{n}{k}(x^k f*x^{n-k}g)\in O(x^{\alpha+n})$$

so that $f*g\in O(x^\alpha)$. ♦

8.7.2. COROLLARY. *Let α be a real <-2, A_α the set of all continuous functions f on /R such that $f\in O(x^\alpha)$ as $x\to\infty$ and B_α the set of all continuous functions g on R such that there exists a real number $\beta>0$ (depending on g) satisfying the conditions $\alpha+\beta<-1$ and $g\in O(x^\beta)$. Then A_α is an algebra under convolution and B_α is a module over that algebra.*

PROOF. Applying to the theorem (changing the roles of α and β), it is readily seen that $f*g$ exists and belongs to B_α whenever $f\in A_\alpha$ and $g\in B_\alpha$; and that $f*g\in A_\alpha$ whenever f, $g\in A_\alpha$. So we have only to prove the associative law: $f*(g*h)=(f*g)*h,\ f,\ g\in A_\alpha,\ h\in B_\alpha$. But this can be easily seen applying 8.5.7. as we did for 8.6.1. ♦

8.7.3. Remark: The preceding theorem and corollary can be extended to locally summable functions according to the following criterium (FUBINI-TONELLI THEOREM): *If f, $g\in L(/R)$*, then

$\int_{/R} f(x-t)g(t)dt$ *is convergent almost everywhere in /R and defines a function $h\in L(/R)$*. It can still be stated that the preceding integral is convergent in the mean on */R*, so that $f*g$ exists in the distributional sense. Applying 8.5.11. and taking the Fubini-Tonelli theorem into account, it is a simple matter to obtain the following generalization of 8.7.1.:

8.7.4. THEOREM. *Let α, β be two real numbers satisfying the conditions (i) or (ii) of* 8.7.1., *α', β' two real numbers such that $\alpha'+\beta'\leq 0$ and f, g two locally summable functions such that $f\in O(x^\alpha e^{\alpha'|x|})$* and *$g\in O(x^\beta e^{\beta'|x|})$. Then $f*g$ exists and $f*g\in O(x^\alpha e^{\gamma|x|})$*, where $\gamma=max(\alpha',\beta')$.

For the proof it is convenient to consider f and g in the form $f=f_1+f_2$, $g=g_1+g_2$, with f_1, $g_1 \in C_+$, f_2, $g_2 \in C_-$, remembering that $f_1 * g_1 \in C_+$, $f_2 * g_2 \in C_-$.

From 8.7.4. is easily deduced a corresponding generalization of 8.7.2.

Now, applying the differentiation property, we can derive from the preceding criteria corresponding rules for distributions. For example, let us denote by $\widetilde{A}_\alpha$ for every $\alpha<-2$, the set of all distributions of the form $f=\sum_{k=0}^{p} D^{n_k} F_k$, where p, $n_1, \ldots, n_p$ are arbitrary integers and F_k locally summable functions such that $F_k \in O(x^\alpha)$, and by $\widetilde{B}_\alpha$ the set of all distributions of the form $g=\sum_{k=0}^{q} D^{r_k} G_k$ where q, $r_1, \ldots, r_q$ are arbitrary integers and G_k locally summable functions such that $G_k \in O(x^\beta)$ with $\alpha+\beta<-3$ and $\alpha \leq \beta$ (β depending on g). Then it is easily seen that $\widetilde{A}_\alpha$ is an algebra under convolution and $\widetilde{B}_\alpha$ a module over $\widetilde{A}_\alpha$.

8.7.5. DEFINITION. A distribution f on $/R$ is said to be **tempered** (**slowly increasing** or of **polynomial type**) if there exists a real α such that $f \in O(x^\alpha)$ (in distributional sense).

An equivalent definition to this is the following: f *is tempered if and only if there exist two integers* n, k *and a function* $F \in C(/R)$ *such that* $f=D^n F$ *and* $F \in O(x^k)$ *in ordinary sense.*

We shall denote by $\breve{\mathscr{D}}(/R)$ or simply by $\breve{\mathscr{D}}$ the set of all tempered distributions. It is readily seen that $\breve{\mathscr{D}}$ *is a vector space closed under* D.

8.7.6. DEFINITION. A distribution f on $/R$ is said to be **rapidly decreasing** if and only if for every $\alpha<0$, f can be represented in the

form $f=\sum_{k=1}^{p} D^{n_k}F_k$, where p, $n_1,\ldots, n_p$ are arbitrary integers ($n_k \geq 0$, $p \geq 1$) and F_k continuous functions such that $F_k \in O(x^\alpha)$ in ordinary sense.

We shall denote by $\widehat{\mathscr{D}}$ the set of all rapidly decreasing distributions on $/R$. From preceding results it is easily deduced:

8.7.7. COROLLARY. *$\widehat{\mathscr{D}}$ is an algebra under convolution and $\breve{\mathscr{D}}$ a module over $\widehat{\mathscr{D}}$.*

A similar result can be obtained concerning the space $\check{\mathscr{D}}$ of all distributions of **exponential** type (that is, of the form $f=D^nF$, where F is a continuous function on $/R$ such that $F \in O(e^{\alpha|x|})$ for *some* real α) and the space $\hat{\mathscr{D}}$ of all **exponentially decreasing** distributions (that is, of the form $f=D^nF$, where F is a continuous function such that $F \in O(e^{\alpha|x|})$ for *all* real number α).

Observe that $\mathscr{D}_* \subset \hat{\mathscr{D}} \subset \widehat{\mathscr{D}} \subset \breve{\mathscr{D}} \subset \check{\mathscr{D}} \subset \mathscr{D}$.

8.7.8. Convolution in $/R^n$. The concept of convolution of distributions on $/R$ is readily extended to the case of distributions on $/R^n$, and all preceding properties of convolutions can be generalized to this case: only we are now concerned with derivation operators, translation operators, etc., corresponding to the different variables.

Theorem 8.6.1. is readily extended to distributions of several variables. As for theorem 8.6.3. it gives place to new possibilities in the case of n variables.

Let $\boldsymbol{\Gamma}$ be any convex cone in $/R^n$ whose vertex is at the origin and not reducing to a half space. We shall denote by $\widetilde{\mathscr{D}}_{\Gamma}$ the set of all distributions on $/R^n$ vanishing outside some cone $\boldsymbol{a}+\boldsymbol{\Gamma}$ with $\boldsymbol{a} \in /R^n$. Then it is easily seen that *$\widetilde{\mathscr{D}}_{\Gamma}$ is an algebra under convolution and $\mathscr{D}_{\Gamma}$ a subalbegra of $\widetilde{\mathscr{D}}_{\Gamma}$*; besides, there exists a maximal subspace of $\mathscr{D}$ distinct from $\mathscr{D}_{\Gamma}$ which is a module over $\widetilde{\mathscr{D}}_{\Gamma}$.

Finally, the criteria given in 8.7. can also be extended to the case of n variables and combined between them and the preceding ones, according to the different variables.

CHAPTER IX

FOURIER TRANSFORMATION.

9.1. Fourier transformation for tempered distributions on $/R$.

Let f be any distribution on $/R$. If the integral $\int_{/R} e^{ixy} f(y)\,dy$ is convergent on $/R$, then the distribution

9.1.1. $$g(x) = \int_{/R} e^{ixy} f(y)\,dy$$

is called the **Fourier transform** of f and we write $g(x) = \mathfrak{F}_{x|y}\, f(y)$, or simply $g = \mathfrak{F} f$. Frequently the Fourier transform of f is also denoted by $\hat{f}$. For simplicity we shall omit the subscript $/R$ in the integral sign when no confusion can arise.

As an example, we have seen that $\int \frac{e^{ixy}}{1+y^2}\,dy = \pi e^{-|x|}$ in distributional sense. Hence,

$$\mathfrak{F}_{x|y} \frac{1}{1+y^2} = \pi e^{-|x|}.$$

From this we deduce that

$$\int e^{ixy}dy = 2\pi\delta(x),$$

and so

9.1.2. $$\mathfrak{F}1 = 2\pi\delta.$$

On the other hand, it is readily seen that

9.1.3. $$\mathfrak{F}\delta = 1.$$

We now establish some fundamental properties of the Fourier transform.

9.1.4. *If* $\mathfrak{F}f$ *and* $\mathfrak{F}g$ *exist, then* $\mathfrak{F}(\lambda f + \mu g)$ *exists for all* $\lambda, \mu \in \mathbb{C}$ and $\mathfrak{F}(\lambda f + \mu g) = \lambda(\mathfrak{F}f) + \mu(\mathfrak{F}g)$.

PROOF. This is an immediate consequence of the linearity property for integrals. ♦

9.1.5. *If* $\mathfrak{F}f$ *exists, then* $\mathfrak{F}(Df)$ *exists and* $\mathfrak{F}(Df) = -i\hat{x}(\mathfrak{F}f)$.

PROOF. $e^{ixy}f'(y) = D_y(e^{ixy}f(y)) - ixe^{ixy}f(y)$. If $\mathfrak{F}f$ exists, i.e., if $e^{ixy}f(y)$ is integrable on $/R$, then $e^{ixy}f(y) \to 0$ on $/R$ as $y \to \infty$, and therefore

$$\int e^{ixy}f'(y)\,dy = -ix\int e^{ixy}f(y)\,dy. \; ♦$$

9.1.6. *If* $\mathfrak{F}f$ *exists, then* $\mathfrak{F}(\hat{y}f)$ *exists and* $\mathfrak{F}(\hat{y}f) = -iD(\mathfrak{F}f)$.

PROOF. If $\mathfrak{F}f$ exists, then by the differentiation property

$$-iD_x\int e^{ixy}f(y)\,dy = \int e^{ixy}yf(y)\,dy. \; ♦$$

Combining 9.1.4., 9.1.5. and 9.1.6., gives:

9.1.7. *If P is any polynomial, then*

$$\mathfrak{F}(P(D)f) = P(-ix)(\mathfrak{F}f)$$
$$\mathfrak{F}(P(y)f) = P(-iD)(\mathfrak{F}f).$$

We now establish some existence criteria for Fourier transforms.

9.1.8. *If f is summable on $/R$, then $\mathfrak{F}f$ exists and is a bounded continuous function.*

PROOF. Suppose $f \in L(/R)$. Since $|e^{ixy}f(y)| = |f(y)|$ for all $x, y \in /R$, the integral $\int |e^{ixy}f(y)|\,dy$ is dominated by the integral $\int |f(y)|\,dy$ which is convergent and independent of x. Hence, $\int e^{ixy}f(y)\,dy$ is uniformly convergent on $/R$, and thus it is convergent in the distributional sense and represents a continuous function $g(x)$ on $/R$. Finally,

$$|g(x)| \leq \int |f(y)|\,dy \quad \text{for all} \quad x \in /R. \; \blacklozenge$$

We shall denote by C_b the space of all bounded continuous functions on $/R$. Recall that $\breve{\mathscr{D}}$ denotes the space of all tempered distributions on $/R$. From 9.1.7. and 9.1.8. follows:

9.1.9. *If $f \in \breve{\mathscr{D}}$ then $\mathfrak{F}f$ exists and $\mathfrak{F}f \in \breve{\mathscr{D}}$.*

PROOF. Suppose $f \in \breve{\mathscr{D}}$. There are $m, p \in /N_0$ and $F \in C(/R)$ such that $f = D^m F$ and $F \in O(x^p)$ in the ordinary sense as $x \to \infty$.

Set $\Phi = \dfrac{F}{(1+i\hat{x})^{p+2}}$. Then $f = D^m((1+i\hat{x})^{p+2}\Phi)$, $\Phi \in C(/R)$, and $\Phi \in O(x^{-2})$.

Therefore, by 9.1.8., $\mathfrak{F}\Phi$ exists and $\mathfrak{F}\Phi \in C_b \subset \breve{\mathscr{D}}$. Hence, by 9.1.7., $\mathfrak{F}f$ also exists and $\mathfrak{F}f = (-i\hat{x})^m(1+D)^{p+2}(\mathfrak{F}\Phi) \in \breve{\mathscr{D}}$. ♦

We next propose to study the problems of the inversion of $\mathfrak{F}$. We observe that $\mathfrak{F}$ transformed 1 into $2\pi\delta$, δ into 1, D into multiplication by $-i\hat{x}$, and multiplication by $\hat{x}$ into $-iD$. Hence, if $\mathfrak{F}^{-1}$ exists, it must transform δ into $1/2\pi$, 1 into δ, etc. Thus we might expect that $\mathfrak{F}^{-1}$ is given by the formula:

9.1.10.
$$f(y) = \frac{1}{2\pi}\int e^{-ixy} g(x)\,dx.$$

We shall temporarily denote by $\mathfrak{F}^*$ the transformation $g \to f$ defined by 9.1.10. It is readily seen that $\mathfrak{F}^*$ has the required properties and that $\mathfrak{F}^*f$ exists for all $f \in \breve{\mathscr{D}}$. Moreover,

9.1.11. *If $f \in \breve{\mathscr{D}}$ and $g = \mathfrak{F}f$, then $f = \mathfrak{F}^*g$; conversely, if $g \in \breve{\mathscr{D}}$* and *$f = \mathfrak{F}^*g$, then $g = \mathfrak{F}f$.*

PROOF. Suppose $f \in \breve{\mathscr{D}}$ and set $g = \mathfrak{F}f$, $h = \mathfrak{F}^*g$. Then

$$h(y) = \frac{1}{2\pi}\int e^{-ixy}\left(\int e^{ixy'} f(y')\,dy'\right)dx,$$

and, if we may interchange the order of summation, we find

$$h(y) = \frac{1}{2\pi}\int\left(\int e^{ix(y'-y)}dx\right) f(y')\,dy'.$$

But

$$\int e^{ix(y'-y)}dx = 2\pi\delta(y'-y) = 2\pi\delta(y-y')$$

and by Dirac's formula

$$h(y) = \int \delta(y-y') f(y')\,dy' = f(y).$$

It is shown analogously that if $g \in \breve{\mathscr{D}}$ and $f = \mathfrak{F}^* g$, then $g = \mathfrak{F} f$. We need only justify the interchange of summations, and by 8.4.4. it is sufficient to show the convergence of the double integral

9.1.12. $$\iint e^{ix(y'-y)} f(y')\,dx\,dy'.$$

The integral

9.1.13. $$\iint \frac{e^{ix(y'-y)}}{1+x^2} f(y')\,dx\,dy'$$

with $f \in L$, is uniformly convergent on $/R$ since for all $x,\ y,\ y' \in /R$,

$$\left|\frac{e^{-ixy}}{1+x^2}\right| = \frac{1}{1+x^2} \text{ and } |e^{ixy'} f(y')| = |f(y')|,$$

and the functions $(1+x^2)^{-1}$ and $f(y')$ are summable on $/R$. Hence, by applying the operator $1-D^2$ to 9.1.13., we see that 9.1.12. is convergent for $f \in L$. The result for $f \in \breve{\mathscr{D}}$ now follows by an argument similar to the proof of 9.1.9. if we observe that 9.1.10. represents $\mathfrak{F}^* \mathfrak{F}$ and that $\mathfrak{F}^* \mathfrak{F}(Df) = Df$, $\mathfrak{F}^* \mathfrak{F}(\hat{x} f) = \hat{x} f$, for all $f \in \breve{\mathscr{D}}$. ♦

Thus we have proved that $\mathfrak{F}^* = \mathfrak{F}^{-1}$ for $\mathfrak{F}$ restricted to $\breve{\mathscr{D}}$. We ask if tempered distributions are the only distributions having a Fourier transform in the previous sense.

The answer is affirmative:

9.1.14. *If the integral* $\int e^{ixy} f(y)\,dy$ *is convergent on* $/R$, *then* $f \in \breve{\mathscr{D}}$.

PROOF. Suppose $\int e^{ixy} f(y)\,dy$ is convergent on $/R$. Then $e^{ixy} f(y)$ is of the form $(1+iy)^{-1} D_x^m D_y^n F(x, y)$ where $F(x, y) \in O(y^n)$ uniformly on each bounded interval as $y \to \infty$. Hence,

$$f(y) = (1+iy)^{-1}(e^{-ixy} D_x^m D_y^n F(x, y))$$

and, since the right member is independent of x, it follows that $f \in O(y^{2n-1})$ and so $f \in \breve{\mathscr{D}}$. ♦

The preceding results may be summarized as follows:

9.1.15. THEOREM. *$\mathfrak{F}$ is a one-to-one linear mapping of the space $\breve{\mathscr{D}}$ onto itself, changing D into multiplication by $-i\hat{x}$, multiplication by x into $-iD$,* 1 *into* $2\pi\delta$, *and δ into* 1. *$\mathfrak{F}^{-1}$ is given by* 9.1.10.

9.2. Fourier transformation and convolution.

The following theorem is well-known:

9.2.1. THEOREM. *If f and g are summable functions on $/R$, then $\mathfrak{F}$ transforms the convolution $f*g$ into the usual product of the continuous functions $\mathfrak{F}f$ and $\mathfrak{F}g$. That is,*

$$\mathfrak{F}(f*g) = (\mathfrak{F}f)(\mathfrak{F}g).$$

PROOF. By the theorem of Fubini-Tonelli (8.7.3.): if $f,\ g \in L$, then $f*g$ exists and $f*g \in L$. Let $\hat{f} = \mathfrak{F}f$, $\hat{g} = \mathfrak{F}g$. Then by 9.1.8., $\hat{f},\ \hat{g} \in C_b$ and

$$\hat{f}(x)\hat{g}(x) = \int_{/R} e^{ixu} f(u)\,du \int_{/R} e^{ixv} g(v)\,dv = \int_{/R^2} e^{ix(u+v)} f(u)g(v)\,du\,dv.$$

Now let $u+v=y$, $v=t$. Then $u=y-t$, the Jacobian of the transformation is 1 and the transformation maps $/R^2$ onto $/R^2$. Therefore,

$$\hat{f}(x)\hat{g}(x) = \int_{/R^2} e^{ixy} f(y-t)g(t)\,dy\,dt = \int_{/R} e^{ixy}\left(\int_{/R} f(y-t)g(t)\,dt\right)dy,$$

and so $\hat{f}\hat{g} = \mathfrak{F}(f*g)$. ♦

9.2.2. COROLLARY. *Let f, g be distributions on $/R$ of the form $f=D^mF$, $g=D^nG$, where F and G are locally summable functions satisfying the condition that there exists an integer p such that $(1+i\hat{x})^pF$ and $(1+i\hat{x})^{-p}G$ are summable on $/R$. Then $\mathfrak{F}(f*g)=$ $=(\mathfrak{F}f)(\mathfrak{F}g)$.*

This is a consequence of theorem 9.2.1. and properties 8.5.8. and 8.5.10. The corollary can obviously be extended to distributions which can be expressed as finite sums of the preceding forms. Recalling the definition of the space $\widehat{\mathscr{D}}$ of all rapidly decreasing distributions, it is easily deduced from 9.2.2.:

9.2.3. COROLLARY. *If $f\in\widehat{\mathscr{D}}$, $g\in\breve{\mathscr{D}}$, then $\mathfrak{F}(f*g)=(\mathfrak{F}f)(\mathfrak{F}g)$.*

In order to characterize the Fourier transforms of the rapidly decreasing distributions, we shall first establish two general criteria:

9.2.4. THEOREM. *If f is a distribution of the form D^nF, where F is a locally summable function on $/R$ and $F\in O(\hat{x}^{-r})$ for r, an integer ≥ 2, and if $\phi=\mathfrak{F}f$, then ϕ is a C^{r-2} function and $\phi^{(k)}\in O(\hat{x}^n)$ for $k=0$, $1,\ldots,r-2$.*

PROOF. Suppose the hypothesis is satisfied and put $\varphi=\mathfrak{F}F$. Then $\phi=(-i\hat{x})^n\varphi$ and since $\hat{x}^kF\in O(\hat{x}^{-2})$ for $k=0,1,\ldots,r-2$, it follows that $D^k\varphi\in C_b$ for $k=0,1,\ldots,r-2$ by 9.1.6. and 9.1.8. Hence $\phi\in C^{r-2}$ and $\phi^{(k)}\in O(\hat{x}^n)$ for $k=0,1,\ldots,r-2$. ♦

9.2.5. THEOREM. *If ϕ is a C^r function such that $\phi^{(r)}\in O(\hat{x}^{n-r})$ for $n, r\in/N_0$, and if $f=\mathfrak{F}\phi$, then f is of the form $f=(1+D)^{n+2}F$ for $F\in C$ such that $F\in O(\hat{x}^{-r})$.*

PROOF. Suppose the hypothesis is satisfied and put $\varphi=(1+i\hat{x})^{-n-2}\phi$, $F=\mathfrak{F}\varphi$. Then $f=(1+D)^{n+2}F$. On the other hand, $\phi\in O(\hat{x}^{n-k})$ for $k=0,1,\ldots,r$, and this implies $\phi^{(r)}\in O(\hat{x}^{-2})$. Hence $\hat{x}^rF\in C_b$ and so $F\in O(\hat{x}^{-r})$. ♦

9.2.6. DEFINITION. A **tempered** C^∞ function on $/R$ is a function

$\phi \in C^\infty(IR)$ satisfying the condition that for every $r=0, 1, \ldots,$ there exists an integer n such that $\phi^{(r)} \in O(x^n)$ in the ordinary sense as $x \to \infty$. We denote by $\mathfrak{M}$ the set of all tempered C^∞ functions on IR.

It is easily seen that $\mathfrak{M}$ is a vector subspace of $\breve{\mathscr{D}} \cap C^\infty$; but observe that $\mathfrak{M} \neq \breve{\mathscr{D}} \cap C^\infty$. From 9.2.4. and 9.2.5. we have:

9.2.7. COROLLARY. *The Fourier transformation $\mathfrak{F}$ maps the convolution algebra $\widehat{\mathscr{D}}$ onto the multiplication algebra $\mathfrak{M}$.*

PROOF. a) Suppose $f \in \widehat{\mathscr{D}}$. This implies that for every $r=0, 1, 2,$

$\ldots$, f can be represented in the form $f=\sum_{k=1}^{m} D^{r_k} F_k$ where $F_k \in O(\hat{x})^{-r-2}$

for $k=1, 2, \ldots, m$. Then if $\phi = \mathfrak{F} f$, it is easily seen from 9.2.4. that $\phi \in C^r$ and $\phi^{(r)} \in O(\hat{x}^\mu)$ where $\mu = max(r_1, \ldots, r_k, \ldots, r_m)$. Hence, $\phi \in \mathfrak{M}$.

b) Suppose $\phi \in \mathfrak{M}$. Then for every $r=0, 1, 2, \ldots,$ there exists n such that $\phi^{(r)} \in O(\hat{x}^n)$. Thus if we put $f = \mathfrak{F}^* \phi$, we conclude from 9.2.5. (which obviously extends to $\mathfrak{F}^*$), that f is of the form $(1+D)^{n+2}F$, where F is a continuous function such that $F \in O(\hat{x}^{-r})$. Hence, $f \in \widehat{\mathscr{D}}$. ♦

9.3. The Fourier transformation as a continuous mapping.

It can be seen that the Fourier transformation is not continuous with respect to the topology of $\mathscr{D}$ restricted to $\breve{\mathscr{D}}$. However, we can define a stronger topology on $\breve{\mathscr{D}}$ which will make $\mathfrak{F}$ as well as D continuous, and extends the usual topologies on function subspaces of $\breve{\mathscr{D}}$. In the space C_b of all bounded continuous functions on IR a norm is usually defined by $\|f\| = \sup_{x \in IR} |f(x)|$. Then,

9.3.1. LEMMA. *$\mathfrak{F}$ defines a continuous mapping of the normed space L into the normed space C_b.*

PROOF. It is sufficient to observe that if $f\in L$, then

$$\|\mathfrak{F}f\|\leq\int|f|=\|f\|_L \text{ (cf. proof of 9.1.8.). } \blacklozenge$$

We shall try to define the strongest topology on $\breve{\mathscr{D}}$ making both $\mathfrak{F}$ and D continuous and inducing a topology on C_b (resp. L) weaker than the norm topology of C_b (resp. L). If such a topology exists, then $\mathfrak{F}^{-1}$ and the mapping $f\to\hat{x}f$ will also be continuous.

These considerations lead us to the following definition of convergence for sequences:

9.3.2. DEFINITION. A sequence of distributions $\{f_n\}\subset\breve{\mathscr{D}}$ converges in the **tempered sense** to a distribution $g\in\breve{\mathscr{D}}$ if there exist an integer p, a sequence of functions $\{F_n\}\subset C_b$ and a function $G\in C_b$ such that

(i) $f_n=D^pF_n$ for all n;
(ii) $g=D^pG$;
(iii) $(1+x^2)^{-p}(F_n-G)$ converges to 0 uniformly on $/R$ as $n\to\infty$.

It is now a simple exercise to verify that this concept of convergence satisfies all of the preceding conditions.

In order to define in $\breve{\mathscr{D}}$ the strongest topology satisfying the same condition, we shall denote by C_b^{-r} for $k=0, 1, 2, \ldots$, the space of all distributions of the form $f=D^r(1+\hat{x}^2)^rF$ with $F\in C_b$, and we shall consider C_b^{-r} provided with the image topology of C_b by means of the mapping $F\to D^r(1+\hat{x}^2)^rF$ of C_b onto C_b^{-r}. Then it is easily seen (as in the case of distributions on a compact interval) that C_b^{-r} is a normed space and the injection $C_b^{-r}\to C_b^{-r-1}$ is compact for $r=0, 1, 2, \ldots$. On the other hand $\breve{D}=\bigcup_{r=0}^{\infty}C_b^{-r}$ so that $\breve{\mathscr{D}}$ with the inductive limit topology of the normed spaces C_b^{-r} is an (LN^*)-space. Then it can be seen that this topology is the strongest one satisfying all preceding conditions and such that the concept of convergence for sequences agrees with that defined directly in 9.3.2.

It can also be proved that the substitution $x=t/(t^2-1)$ defines a one-to-one continuous linear mapping of the locally convex space $\breve{\mathscr{D}}$ into the locally convex space $\mathscr{D}[-1, 1]$.

9.4. Fourier transformation and scalar product.

Sometimes the Fourier transformation is defined by the formula

9.4.1.
$$g(x)=\frac{1}{\sqrt{2\pi}}\int_{IR} e^{ixy} f(y)\,dy,$$

instead of 9.1.1. So far as no misunderstanding may arise, we shall still write in this case $g=\mathfrak{F}f$. The fundamental properties of Fourier transformations that we have previously proved are not altered by this change of form. But we now have of course,

$$\mathfrak{F}\delta=\frac{1}{\sqrt{2\pi}}, \quad \mathfrak{F}1=\sqrt{2\pi}\,\delta.$$

The advantage of this new form is that it preserves in many cases the hermitic scalar product of two distributions.

9.4.2. DEFINITION. A **rapidly decreasing** C^∞ function on IR is a function $\phi\in C^\infty$ such that $\phi^{(n)}\in O(\hat{x}^{-r})$ for all n, $r=0, 1, 2, \ldots$.

We shall denote by $\mathbb{S}$ the set of all rapidly decreasing functions. $\mathbb{S}$ is a proper vector subspace of $\widehat{\mathscr{D}}\cap\mathfrak{M}$. For example, $exp(-\hat{x}^2)\in\mathbb{S}$. It is easily seen that $\langle f, \phi\rangle$ exists on IR whenever $f\in\breve{\mathscr{D}}$ and $\phi\in\mathbb{S}$. Moreover, if we consider the topology on $\mathbb{S}$ defined by the sequence of norms

$$\|\phi\|_n=\sup_{x\in IR}\,(|\phi(x)|, (1+x^2)|\phi'(x)|, \ldots, (1+x^2)^n|\phi^{(n)}(x)|),$$

it can be proved that $\breve{\mathscr{D}}$ is isomorphic to $\mathbb{S}'$. (In the theory of Schwartz,

the space $\breve{\mathscr{D}}$ is defined to be $\mathbb{S}'$). On the other hand, it is easily seen by applying 9.2.4. and 9.2.5., that $\mathfrak{F}$ maps the space $\mathbb{S}$ onto itself. Thus $\mathbb{S}$ is at the same time a multiplication algebra and a convolution algebra.

9.4.3. THEOREM. *If* $f \in \breve{\mathscr{D}}$ *and* $\phi \in \mathbb{S}$, *then* $\langle f, \phi \rangle = \langle \mathfrak{F}f, \mathfrak{F}\phi \rangle$.

PROOF. Set $g = \mathfrak{F}f$, $\psi = \mathfrak{F}\phi$. Then,

$$\langle f, \phi \rangle = \int_{IR} f(x)\overline{\phi(x)}dx = \int_{IR}\left(\frac{1}{\sqrt{2\pi}}\int_{IR} e^{-ixy} g(y)\,dy\right)\overline{\phi(x)}dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_{IR^2} e^{-ixy} g(y)\,\overline{\phi(x)}dxdy = \int_{IR} g(y)\left(\frac{1}{\sqrt{2\pi}}\int_{IR} e^{-ixy}\overline{\phi(x)}dx\right)dy$$

$$= \int_{IR} g(y)\,\overline{\psi(y)}dy,$$

since the double integral exists and $exp(-ixy) = \overline{exp(ixy)}$. ♦

In the theory of Schwartz, this theorem is true by definition since $\mathfrak{F}$ is defined as the transpose of $\mathfrak{F}$ restricted to $\mathbb{S}$:

$$\langle \mathfrak{F}f, \phi \rangle = \langle f, \mathfrak{F}\phi \rangle, \text{ for all } f \in \breve{\mathscr{D}}, \quad \phi \in \mathbb{S}.$$

Let us now consider the Hilbert space of all square summable functions on IR, which we shall denote by $\mathcal{H}$. We shall put

$$\|f\|_2 = \sqrt{\langle f, f \rangle} \text{ for all } f \in \mathcal{H}.$$

Convergence in this norm is called **convergence in the square mean**. It is well known that every $f \in \mathcal{H}$ can be approached in the square mean by a sequence of functions $\{\phi_n\} \subset C^{\infty}_{*}(IR)$, so that in particular $\mathbb{S}$ is dense in $\mathcal{H}$. On the other hand, it follows from 9.4.3. that if $\phi \in \mathbb{S}$, then $\|\phi\|_2 = \|\mathfrak{F}\phi\|_2$. Consequentely, the Fourier transformation restricted to $\mathbb{S}$ can be extended to a linear isometry of the space

$\mathcal{H}$ onto itself. We shall provisionally denote this mapping by $\widetilde{\mathfrak{F}}$. In particular, if f is a locally summable function with bounded carrier, then clearly $f \in \mathcal{H}$ and $\widetilde{\mathfrak{F}}f = \mathfrak{F}f$. Thus, in general, $\widetilde{\mathfrak{F}}f$ is given by the limit, as $a \to -\infty$ and $b \to +\infty$, in the square mean of

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{ixy} f(y)\, dy.$$

9.4.4. LEMMA. *If* $f,\ g \in \mathcal{H}$ *then* $\int_{-n}^{n} f(x-y) g(y)\, dy$ *converges uniformly on each compact subset of* $/R$, *as* $n \to +\infty$, *to a continuous function* h *such that* $\mathfrak{F}h = (\widetilde{\mathfrak{F}}f)(\widetilde{\mathfrak{F}}g)$.

PROOF. Set $f_n(x) = f(x)(H(x+2n) - (H(x-2n))$, $g_n(x) = g(x)(H(x+n) - H(x-n)$. Then $f_n,\ g_n \in L \cap \mathcal{H}$ for all n and

$\int_{-n}^{n} f(x-y) g(y)\, dy = (f_n * g_n)(x)$ for $|x| < n$. Hence, if we put $\hat{f}_n = \mathfrak{F}f_n$,

$\hat{g}_n = \mathfrak{F}g_n$, $\tilde{f} = \widetilde{\mathfrak{F}}f$ and $\tilde{g} = \widetilde{\mathfrak{F}}f$, we have $\hat{f}_n \hat{g}_n = \mathfrak{F}(f_n * g_n) \in L$ for all n, and since $\hat{f}_n \to \tilde{f}$, $\hat{g}_n \to \tilde{g}$ in the square mean, then $\hat{f}_n \hat{g}_n \to \tilde{f}\tilde{g}$ in the square mean, and therefore $f_n * g_n \to h = \widetilde{\mathfrak{F}}^{-1}(\tilde{f}\tilde{g})$ uniformly on $/R$.

Consequently, $\int_{-n}^{n} f(x-y) g(y)\, dy$ converges uniformly on each compact subset of $/R$ to the function h, which is obviously continuous. ♦

9.4.5. THEOREM. *Every function* $f \in \mathcal{H}$ *is a tempered distribution and* $\widetilde{\mathfrak{F}}f = \mathfrak{F}f$ *for every* $f \in \mathcal{H}$. *Moreover, convergence in the square mean implies convergence in the distributional sense and if* $f, g \in \mathcal{H}$, *then* $f * g$ *exists in the distributional sense and is a continuous function such that* $\mathfrak{F}(f * g) = (\mathfrak{F}f)(\mathfrak{F}g)$.

PROOF. Set $\tilde{f}=\mathfrak{F}f$, $\tilde{f}_0=(1+i\hat{x})^{-1}\tilde{f}$ and $f_0=\mathfrak{F}^{-1}\tilde{f}_0$, (observe that $\tilde{f}_0\in\mathcal{H}\cap L$). Since $\mathfrak{F}^{-1}(1+i\hat{x})=\delta-\delta'$, $(\delta-\delta')*f_0=(1-D)f_0$, it follows from the lemma that $f=(1-D)f_0$ and hence $f\in\breve{\mathscr{D}}$, since $f_0\in C_b$. The remainder of the theorem follows from the preceding results.

9.4.6. COROLLARY. If $f, g\in\mathcal{H}$, then $\langle f, g\rangle=\langle\mathfrak{F}f, \mathfrak{F}g\rangle$.

PROOF. It is sufficient to observe that $\mathfrak{F}$ is an isometric linear mapping of the hermitic space $\mathcal{H}$ onto itself. ♦

9.5. Fourier transformations on $/R^n$.

The Fourier transformation on $/R^n$ may be defined by

$$g(\boldsymbol{x})=\int_{/R^n} e^{i\boldsymbol{xy}}f(\boldsymbol{y})\,d\boldsymbol{y}$$

where f is a distribution on $/R^n$, and $\boldsymbol{xy}=\sum_{k=1}^{n} x_k y_k$. If the integral is convergent on $/R^n$, we write $g=\mathfrak{F}f$.

A distribution f on $/R^n$ is said to be **tempered** if and only if there exist two systems $\boldsymbol{p}$, $\boldsymbol{r}\in /N_0^n$ and a function $F\in C(/R)$ such that $f=D^{\boldsymbol{p}}F$ and $F\in O(x_1^{r_1}\dots x_n^{r_n})$ in the ordinary sense. We write $f\in\breve{\mathscr{D}}(/R^n)$ or simply $f\in\breve{\mathscr{D}}$.

All preceding properties of the Fourier transformation can be extended to the present case with the obvious modifications concerning the existence of n derivation operators and n coordinate functions $\hat{x}_1,\dots,\hat{x}_n$. Thus,

$$\mathfrak{F}(D_k f)=(-i\hat{x}_k)(\mathfrak{F}f),$$

$$\mathfrak{F}(\hat{x}_k f)=(-iD_k)(\mathfrak{F}f),$$

for all $f \in \breve{\mathscr{D}}$ and $k=1,\ldots,n$. Moreover, in the inversion formula the coefficient $\frac{1}{2\pi}$ must be replaced by $\frac{1}{(2\pi)^n}$.

Observe that if $f \in \breve{\mathscr{D}}(/R^n)$ we can define

$$g_k(\mathbf{x}) = \int_{/R} e^{ix_k y_k} f(\mathbf{x})\, dx_k,$$

the Fourier transform of f with respect to x_k. (It is easily proved that this partial integral is convergent on $\prod_{j \neq k} /R_{x_j}$). Then we write $g_k = \mathfrak{F}_k f$, and it is easily seen that

$$\mathfrak{F} = \mathfrak{F}_1 \cdots \mathfrak{F}_n .$$

For the existence of $\mathfrak{F}_k f$ it is not necessary that $f \in \breve{\mathscr{D}}(/R^n)$. It is sufficient that there exist an integer p, such that

$$f \in O(x_k^p) \ \text{ on } \prod_{j \neq k} /R_{x_j} \ \text{ as } x_k \to \infty .$$

SUPPLEMENTARY BIBLIOGRAPHY

[1] A. ANDRADE GUIMARÃES. *Sur une façon de définir sans dualité l'espace des distributions tempérées sur la droite et la transformation de Fourier.*
Portugaliæ Mathematica, 18 (1959), 125-153.

[2] J. SEBASTIÃO E SILVA. *Les fonctions analytiques comme ultra-distributions dans le calcul opérationnel.*
Math. Annalen, vol. 136, pp 58-96, (1958).

[3] J. SEBASTIÃO E SILVA. *Sur le calcul symbolique des opérateurs differentiels à coefficients variables.*
Rend. Accad. Lincei (8), 27 (1959), 42, 119-122.

[4] J. SEBASTIÃO E SILVA. *Sur la définition et la structure des distributions vectorielles.*
Porth. Math., 19 (1960), 1-80.

[5] J. SEBASTIÃO E SILVA. *Sur le calcul symbolique des opérateurs permutables à spectre vide ou non-borné.*
Annali di Matematica Pura ed Applicata, (4), (58) (1962), 219-275.

[6] I. MARINESCU. *Espaces vectoriels pseudo-topologiques et théorie des distributions.*
Deutscher Verlag der Wissenschaften, Berlin, 1963.

ÍNDICE GERAL

Esta 1.ª edição de TEXTOS DIDÁCTICOS – Vol. III, de José Sebastião e Silva, foi composta, impressa e brochada para a *Fundação Calouste Gulbenkian,* nas oficinas da ORGAL *Impressores* – Porto. A tiragem é de 2000 exemplares.

Julho de 2002

Depósito Legal N.º 148 805/00

ISBN 972-31-0971-9