JOSÉ SEBASTIÃO E SILVA

TEXTOS DIDÁCTICOS

Volume III

SERVIÇO DE EDUCAÇÃO E BOLSAS
FUNDAÇÃO CALOUSTE GULBENKIAN | LISBOA

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 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

$$\mathfrak{S}_k f(\boldsymbol{x}) = \int_{c_k}^{x_k} f(x_1, \ldots, \xi_k, \ldots, x_n)\, d\xi_k, \quad \forall f \in C, \ k = 1, \ldots, n.$$

The **integration operator** $\mathfrak{S}_k$ 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 $\mathfrak{S}^{\boldsymbol{r}}$ the operator $\mathfrak{S}_1^{r_1} \cdots \mathfrak{S}_n^{r_n}$. Obviously, $\mathfrak{S}_k$ is a right inverse of D_k, i.e., $D_k \mathfrak{S}_k f = f$, for any $f \in C$. More generally, for every system $\boldsymbol{r} = (r_1, \ldots, r_n)$ of non-negative integers, we have $\boldsymbol{D}^{\boldsymbol{r}} \mathfrak{S}^{\boldsymbol{r}} f = f$, $\forall f \in C$.

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 $/N_0^n$.

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 $\boldsymbol{r}\in/N_0^n$ 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. $$\mathfrak{I}^{m-r}F-\mathfrak{I}^{m-s}G\in\mathcal{P}_m.$$

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* $\boldsymbol{p}\in/N_0^n$.

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:

$$\mathfrak{I}^{\mu-r}F-\mathfrak{I}^{\mu-s}G\in\mathcal{P}_{\boldsymbol{\mu}}.$$

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 $\boldsymbol{p}\in/N_0^n$; in particular, $D_1=D^{(1,0,\ldots,0)},\ldots, D_n=D^{(0,0,\ldots,1)}$. 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 $I_1,\ldots, I_n$ in $/R$. Let $c_{k1},\ldots, c_{kr_k}$ be r_k points chosen arbitrary in I_k for $k=1,\ldots, n$. Then, to each function $f\in C$ and each $k=1,\ldots, n$, corresponds one, and only one, function $f_k\in\mathcal{P}_{kr_k}$, such that:

7.2.2. $\qquad f_k(x)=f(x)$ for $x_k=c_{kv}$, $v=1,\ldots, r_k$.

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. $$f_{k\nu}(x)=f(x_1,\ldots,x_{k-1},c_{k\nu},x_{k+1},\ldots,x_n)$$

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 $k=1,\ldots,n$. 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 $\in\mathcal{P}_r$ having a continuous derivative $\boldsymbol{D}^p\Theta$ 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 $k=2,\ldots,n$. Put in general

$$w_k=\pi_k(\Theta-w_1-\cdots-w_{k-1})\text{ for }k=2,\ldots,n.$$

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 $/R^n$, $f = \boldsymbol{D}^{\boldsymbol{r}} F$ and $g = \boldsymbol{D}^{\boldsymbol{s}} G$, where $\boldsymbol{r}, \boldsymbol{s} \in /N_0^n$ and $F, G \in C(I)$. As in the case of one variable, we shall put, by definition:

$$f + g = \boldsymbol{D}^{\boldsymbol{m}}(\mathfrak{I}^{m-r} F + \mathfrak{I}^{m-s} G)$$

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 $\boldsymbol{h} \in /R^n$, 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 $\boldsymbol{r} \in /N_0^n$, 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 $/R^n$, 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 $/R^n$, 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 $/R^n$ 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 $/R^n$, 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 $\boldsymbol{r} \in /N_0^n$ and $f \in C(\Omega)$ such that $f = \boldsymbol{D}^{\boldsymbol{r}} F$; 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 /R, the components of an open set Ω in $/R^n$ 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 $/R^n$. 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 $/R^n$ 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 R^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 R^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* $f(\boldsymbol{x}) = \dfrac{\partial}{\partial x_1} \cdots \dfrac{\partial}{\partial x_n} F(\boldsymbol{x})$ *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:

$$\widetilde{f}(\boldsymbol{x}) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} \int_c^x f(\xi)\, d\xi$$

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.

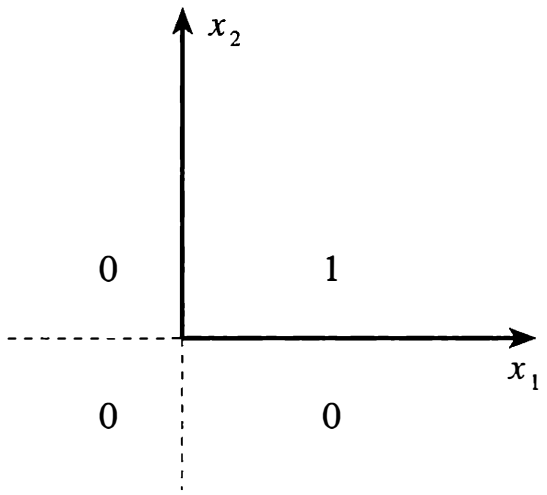
That being so, it is natural to identify each function $f \in \mathring{L}(I)$ with the corresponding distribution $f^* \in \mathscr{D}(I)$.

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:

$$H^{[n]}(\boldsymbol{x}) = \begin{cases} 1 \text{ if } x_k \geq 0 \text{ for } \textit{all } k = 1, \ldots, n \\ 0 \text{ if } x_k < 0 \text{ for } \textit{some } k = 1, \ldots, n. \end{cases}$$

It is obvious that $H^{[n]}(\boldsymbol{x}) = H(x_1) \cdots H(x_n)$, $\forall \boldsymbol{x} = (x_1, \ldots, x_n) \in /R^n$.

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 $n = 1$:

$$\delta_{(a)} = \delta(\hat{\boldsymbol{x}} - \boldsymbol{a}) = D_1 \cdots D_n H(\hat{\boldsymbol{x}} - \boldsymbol{a}).$$

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 $\boldsymbol{x}$, $\boldsymbol{y}$, ... 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:

$$J_k = \begin{cases} [c_k, x_k] & \text{for } c_k \le x_k \\ ]x_k, c_k[& \text{for } x_k < c_k \end{cases} \quad (k = 1, 2, \ldots, n)$$

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 $\boldsymbol{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$, 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 $\mathbb{R}^n$), 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 $\boldsymbol{a}, \boldsymbol{b} \in I$ $(\boldsymbol{a} < \boldsymbol{b})$:

$$\overline{\boldsymbol{\Delta}} F(J) = \overline{\boldsymbol{\Delta}}_{\boldsymbol{h}} F(\boldsymbol{a}) \quad \text{with} \quad \boldsymbol{h} = \boldsymbol{b} - \boldsymbol{a}.$$

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

$$\sum_{\nu_1=1}^{p_1} \cdots \sum_{\nu_n=1}^{p_n} \left| \overline{\boldsymbol{\Delta}} F(J_{1\nu_1} \times \cdots \times J_{n\nu_n}) \right| \leq M(J).$$

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 $r_k\geq 0$, $\mathfrak{M}_{\boldsymbol{r}}(I)$ – or simply $\mathfrak{M}_{\boldsymbol{r}}$ – denotes the set of all distributions $f=D^{\boldsymbol{r}}F$, where $F\in\mathfrak{M}(I)$. Obviously, $\mathfrak{M}_{\boldsymbol{r}}$ is a vector subspace of $\mathscr{D}$ and $C^{\boldsymbol{r}}$ 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

$$D_k(fg) = D_k f \cdot g + f \cdot D_k g \quad \text{for} \quad k = 1, \ldots, n.$$

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 $f^{(k)}$ 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* $\boldsymbol{r} = (r_1, \ldots, r_n)$ *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:

$$fg = \sum_{\nu}^{m} fg_\nu$$

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 $f(\boldsymbol{x})g(\boldsymbol{y}) = f(x_1, \ldots, x_m)g(y_1, \ldots, y_n)$ 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^*:*

$$J\begin{pmatrix} h_1 & \cdots & h_n \\ t_1 & \cdots & t_n \end{pmatrix} \neq 0 \quad \text{for all} \quad \boldsymbol{t}\in\Omega^*.$$

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 $f=D^r F$, with $F\in C(\Omega)$ and $\boldsymbol{r}=(r_1,\ldots,r_n)$ and if $\alpha_{ij}\in C^r(\Omega^*)$ for all i, j then

$$f \circ \boldsymbol{h} = D_x^r (F \circ \boldsymbol{h}) \in C_r(\Omega^*)$$

with

$$D_x^r = \left(\sum_{j=1}^{n} \alpha_{1j} D_{t_j} \right)^{r_1} \cdots \left(\sum_{j=1}^{n} \alpha_{nj} D_{t_j} \right)^{r_n}.$$

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 $(D_x f) \circ \boldsymbol{h}$ *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. $$\sum_{j=1}^{n} h_{\nu j} D_{t_j} (f \circ \boldsymbol{h}) = \sum_{i=1}^{m} \tilde{h}_{\nu i} (D_{x_i} f \circ \boldsymbol{h}), \quad \nu = 1, \ldots, m$$

where $$\tilde{h}_{\nu i} = \sum_{j=1}^{n} h_{\nu j} h_{ij}, \quad \nu = 1, \ldots, m.$$

Then the determinant $|\tilde{h}_{\nu i}|$ 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 $i = 1, \ldots, m$

$$(D_{x_i} f) \circ \boldsymbol{h} = \sum_{j=1}^{n} \alpha_{ij} D_{t_j} (f \circ \boldsymbol{h})$$

or in short

$$D_{x_i} = \sum_{j=1}^{n} \alpha_{ij} D_{t_j}, \quad i = 1, \ldots, m,$$

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 $u=\boldsymbol{h}(x_1,\ldots,x_n)$ mapping an open set Ω^* in R^n into Ω in R (now the new variables are $x_1,\ldots,x_n$ instead of $t_1,\ldots,t_n$). 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 $f(\boldsymbol{h}(x_1,\ldots,x_n))=(f\circ\boldsymbol{h})(\boldsymbol{x})$ such that

7.8.5. $$D_{x_j}(f\circ\boldsymbol{h})=h'_{x_j}(f'\circ\boldsymbol{h}), \quad j=1,\ldots,n,$$

where $f'=D_u f$. 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

$$\alpha_j=\frac{h'_{x_j}}{(h'_{x_1})^2+\cdots+(h'_{x_n})^2}, \quad j=1,\ldots,n,$$

we obtain

$$f'\circ\boldsymbol{h}=\sum_{1}^{n}\alpha_j D_{x_j}(f\circ\boldsymbol{h}),$$

that is

7.8.6. $$D_u=\alpha_1 D_{x_1}+\cdots+\alpha_n D_{x_n}.$$

Observe now that, for each $u\in\boldsymbol{h}(\Omega^*)$, the equation $\boldsymbol{h}(\boldsymbol{x})=u$ represents a hypersurface Σ_u in R^{n+1}, and that $\alpha_1,\ldots,\alpha_n$ are the components of the vector

$$\frac{1}{|grad\, \boldsymbol{h}|}\boldsymbol{n} \quad \text{with} \quad \boldsymbol{n}=\frac{1}{|grad\, \boldsymbol{h}|}grad\, \boldsymbol{h}$$

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 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 $\boldsymbol{x}\in R^3$, $|\boldsymbol{x}|=\sqrt{x_1^2+x_2^2+x_3^2}$ 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

$$V=\begin{cases}\dfrac{1}{4\pi}\dfrac{1}{\rho} & \text{for } |\boldsymbol{x}|\le\rho\\[2ex] \dfrac{1}{4\pi}\dfrac{1}{|\boldsymbol{x}|} & \text{for } |\boldsymbol{x}|>\rho\,.\end{cases}$$

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 $\boldsymbol{x}\in R^3$, 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 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:

$$f_k \to g \Rightarrow D^r f_k \to D^r g, \quad \forall r \in N_0^n.$$

Examples: 1 – Put

$$\delta_k^{[n]}(x) = \delta_k(x_1) \cdots \delta_k(x_n)$$

where $\delta_k = H_k'$ (cf. 6.8.1.). Then it can be seen, as in the case $n=1$, that

$$\lim_{k\to\infty} D^r \delta_k^{[n]} = D^r \delta^{[n]} \quad \forall r \in N_0^n.$$

2 – Considering primitives of the measures $\delta(x)$ and $\delta\left(|x| - \frac{1}{k}\right)$, for $k = 1, 2, \ldots$ it is easily seen that

$$\lim_{k\to\infty} \delta\left(|x| - \frac{1}{k}\right) = \delta(x).$$

3 – Let $U_k(x)$ be equal to $\frac{1}{k}\left(\text{resp. } \frac{1}{|x|}\right)$ for $|x| \le \frac{1}{k}\left(\text{resp. } |x| > \frac{1}{k}\right)$ ($k = 1, 2, \ldots, x \in R^3$). Then

$$\int_I \left[U_k(x) - \frac{1}{|x|}\right] dx \to 0$$

for every bounded interval I, so that $U_k \to \frac{1}{|x|}$ in distributional sense. Hence

$$\Delta U_k \to \Delta\left(\frac{1}{|x|}\right)$$

and, therefore (cf. 2):

$$\Delta \frac{1}{|x|} = -4\pi\delta(x) \quad \text{on } /R^3.$$